UNIVERSIDAD DE COSTA RICA
SISTEMA DE ESTUDIOS DE POSGRADO

MODELO ESPACIAL BAYESIANO PARA LA INCIDENCIA DE DENGUE EN LA ISLA PRINCIPAL DE PUERTO RICO PARA EL AÑO 2014

Tesis sometida a la consideración de la Comisión del Programa de Estudios de Posgrado en Matemática para optar al grado y título de Maestría Académica en Matemática Aplicada.

GREIVIN HERNÁNDEZ GONZÁLEZ

Ciudad Universitaria Rodrigo Facio, Costa Rica

2017

Agradecimientos

A mi familia, que ha estado de una u otra forma en este proceso de aprendizaje que emprendí desde mi niñez, a mi madre Mercedes por ser siempre una luz en mi vida, darme una dirección y aún en su ausencia terrenal recordarme cuanto deseaba verme en estas instancias, a mi hermano Andrés por convertirse en un ejemplo de lucha, esfuerzo y superación, por mostrarme con sus actos que si se quiere algo y se trabaja por ello se puede obtener, a mi tía Francisca (Paquita) por ser una fuente de apoyo, por convertir su casa en mi hogar, por ser una segunda madre y alguien a quien le debo gran parte de mis logros académicos, a mi padre Carlos por todas las veces que me dió palabras de aliento, paz y tranquilidad, a mi hermano Alexánder por creer en mi.

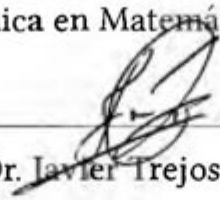
A una persona muy especial, quien siempre ha creído en mi, a quien admiro por ser una mujer luchadora, trabajadora y que a pesar de los malos momentos se ha mantenido cerca de una u otra forma para darme sus palabras de apoyo, por siempre tener sus mejores deseos, simplemente gracias Caro.

A mi director de tesis, Luis Barboza, por su gran ayuda, su infinita paciencia y por las muchas enseñanzas en todo este proyecto, a mis lectores Maikol Solís por sus buenos consejos y aportes, y Alexánder Ramírez por darme el apoyo para sacar adelante esta maestría y por su insistencia en la culminación de la misma.

Al Profesor Fabio, quien tuvo la gentileza de ayudarme a conseguir gran parte de los datos necesarios para llevar a cabo el presente estudio, al profesor Javier Trejos por sus muchas observaciones en el proceso de redacción, forma y escritura del documento, por tomar de su valioso tiempo y leer página por página mi tesis y amablemente darme sus correcciones.

Finalmente, pero no por ello menos importante, le agradezco a mis amigos, mis otros hermanos, Hugo, Rónald, Miguel y Ólger, por todos esos cafés, los buenos momentos compartidos, por ser una influencia tan positiva en mi vida académica y personal y por tomarse su tiempo para leer y darme algunas observaciones de mi trabajo.

"Esta tesis fue aceptada por la Comisión del Programa de Estudios de Posgrado en Matemática de la Universidad de Costa Rica, como requisito parcial para optar al grado y título de Maestría Académica en Matemática Aplicada"



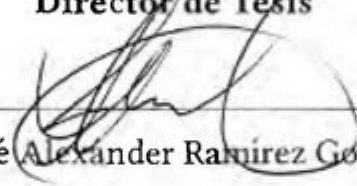
Dr. Javier Trejos Zelaya

Representante del Decano

Sistema de Estudios de Posgrado

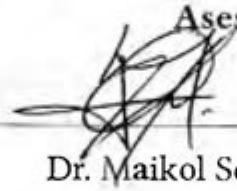
Dr. Luis Alberto Barboza Chinchilla

Director de Tesis



Dr. José Alexánder Ramírez González

Asesor



Dr. Maikol Solís Chacón

Asesor

Dr. Pedro Méndez Hernández

Director

Programa de Posgrado en Matemática

Greivin Hernández González

Candidato

Índice general

Agradecimientos ii

Resumen vi

Lista de tablas vii

Lista de figuras ix

1. Marco Teórico 1

1.1. Inferencia Bayesiana . . . 1

1.2. Interpolación Espacial Óptima . . . 2

1.2.1. Datos de tipo espacial . . . 3

1.2.2. El supuesto de estacionariedad . . . 5

1.2.3. Variograma y semivariograma . . . 6

1.2.4. Algunas familias de funciones de covarianza . . . 9

1.2.5. Ajustando un modelo de variograma . . . 12

1.2.6. Kriging: el método y sus ecuaciones . . . 12

1.3. Modelos lineales generalizados mixtos (glmm) . . . 20

1.4. Los Modelos para Datos de Área: CAR y SAR . . . 21

1.4.1. Herramientas de exploración para datos de área . . . 21

1.4.2. Modelo Condicional Autorregresivo (CAR) . . . 23

1.4.3. Modelo Simultáneo Autorregresivo (SAR) . . . 29

1.5. Cadenas de Markov de Monte Carlo . . . 31

1.5.1. Simulación de Monte Carlo . . . 31

1.5.2. Cadenas de Markov . . . 32

1.6.1. Muestreo de Gibbs y el Algoritmo Metropolis-Hastings . . . 33

1.6.2. Convergencia del Algoritmo Metropolis-Hastings . . . 36

1.10. Comparación y evaluación de modelos . . . 38

1.10.1. Criterio de información de la devianza (DIC) . . . 38

1.10.2. Criterio de información de Watanabe-Akaike (WAIC) . . . 40
1.10.3. Diagnóstico de convergencia de Geweke . . . 41

2. Sobre los Datos **43**
2.1. Origen de los datos . . . 43
2.2. Indice de Vegetación Mejorado . . . 45
2.3. Interpolación espacial de datos climáticos . . . 48
2.3.1. Análisis exploratorio de datos, caso no espacial . . . 49
2.3.2. Análisis exploratorio de datos, caso espacial . . . 51

3. Modelo Condicional Autoregresivo para datos de dengue en Puerto Rico **61**
3.1. Datos y análisis exploratorio . . . 61
3.2. Modelos no espaciales, primer ajuste . . . 65
3.3. Modelos no espaciales, ajuste final . . . 67
3.4. Modelos espaciales, elección del modelo y resultados . . . 69
3.4.1. Resultados del modelo Besag-York-Mollie . . . 73
3.4.2. Resultados del modelo Lee-Mitchell . . . 78
3.5. Conclusiones . . . 80

A. Cuadros de ajuste en variogramas **82**

B. Mapas de interpolación espacial **86**

C. Intervalos de predicción **91**
C.1. Intervalos de modelos iniciales . . . 91
C.2. Intervalos en modelos finales . . . 96

D. Gráficos para riesgo relativo y riesgo estimado por modelo Lee-Mitchel **100**

E. Residuos de modelo BYM **107**

Resumen

El Dengue es la infección viral transmitida por artrópodos más importante de las últimas décadas para los seres humanos. Se ha estimado que aproximadamente 2500 millones de personas, en cerca de 100 países, están en riesgo de desarrollar dicha enfermedad [12]. El riesgo aumenta para las personas que habitan en las regiones tropicales o subtropicales del mundo. El mosquito *Aedes aegypti*, su principal vector de transmisión, es una especie endémica para la mayoría de estas regiones. Los modelos teóricos que existen para ajustar la dinámica de transmisión del mosquito dan gran importancia a variables como la temperatura y la precipitación, la finalidad de dichos estudios es determinar los patrones de transmisión; sin embargo ha faltado evidencia empírica.

Para desarrollar 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 [12]. En esta tesis se utiliza un modelo espacial Autoregresivo Condicional (CAR, por sus siglas en inglés) para determinar el riesgo relativo en cada municipio de Puerto Rico, explícitamente en los municipios que pertenecen a la Isla Principal. El modelo propuesto intenta estudiar la relación de variables como: temperatura, precipitación y altitud, entre otras, con el riesgo relativo de transmisión del dengue. Se aplica esta metodología a datos recolectados en la Isla de Puerto Rico en el año 2014, datos que están desagregados por municipio.

En la primera parte de esta tesis se da una descripción bibliográfica sobre los modelos espaciales CAR (Conditional Auto-Regressive) y SAR (Simultaneously Auto-Regressive); también se explican dos métodos de simulación para Cadenas de Markov de Monte Carlo (MCMC, por su siglas en inglés), estos son: el Muestreo de Gibbs y el método de Metropolis-Hastings. Otros dos temas a tratar son Estimación Bayesiana y Estimación Clásica. Con especial énfasis se va a describir el método de interpolación espacial óptima, también conocida como kriging. Se habla de los índices de correlación espacial de Moran y Geary, y finalmente se da una descripción sobre algunos criterios para la bondad de ajuste de modelos espaciales, por ejemplo: Deviance Information Criterion (DIC)y Widely Applicable Bayesian Information Criterion (WAIC).

En la segunda parte de la tesis se utilizan algunas de las herramientas mencionadas anteriormente para hacer un análisis descriptivo de los datos que se utilizan en el modelo. En particular, se justifica el uso del modelo espacial haciendo uso de los Índices de Moran y de Geary para estudiar la autocorrelación espacial de la variable en estudio y de las covariables. Otro tema a tratar en esta segunda parte es el de realizar la interpolación espacial a los datos o variables climatológicas.

La parte final del proyecto se centra en la aplicación del Modelo Espacial CAR, al caso particular de la variable de Riesgo Relativo de cada Municipio de la Isla Principal de Puerto Rico, y se estudia la bondad de ajuste del mismo cuando se incluyen variables de índole climatológico y social, además de hacer la comparación de los modelos realizados.

Índice de cuadros

1.1. Principales formas de kriging lineal. 13

2.1. Ajuste en variogramas para precipitación, semana 1. 54
2.2. Pruebas de hipótesis para precipitación, semana 1. 55
2.3. Parámetros de variograma ajustado en precipitación, semana 1. 55
2.4. Ajuste en variogramas para temperatura mínima, semana 1. 57
2.5. Pruebas de hipótesis para temperatura mínima, semana 1. 57
2.6. Parámetros de variograma ajustado en temperatura mínima, semana 1. 58
2.7. Ajuste en variogramas para temperatura máxima, semana 1. 59
2.8. Pruebas de hipótesis para temperatura máxima, semana 1. 59
2.9. Parámetros de variograma ajustado en temperatura máxima, semana 1. 60

3.1. Modelos de ajuste a comparar. 61
3.2. Resumen de la distribución de datos fijos. 62
3.3. Resumen de la distribución de datos variables semanales. 63
3.4. Índices de Moran asociados a 3 variables de los modelos finales. 64
3.5. Estimadores para modelo lineal generalizado, primer ajuste. 66
3.6. Intervalos de predicción para modelo lineal generalizado, primer ajuste. 66
3.7. DIC y WAIC para modelos GLM e Independiente, primer ajuste. 67
3.8. Intervalos de predicción para parámetros en GLM reducido. 68
3.9. Parámetros de ajuste para GLM. 69
3.10. Valor Z en la prueba de Geweke. 70
3.11. Criterio de Información de Watanabe-Akaike. 72
3.12. Mediana de parámetros de efectos aleatorios para modelo BYM. 74
3.13. Parámetros de regresión para covariables, modelo BYM. 75
3.14. Intervalos de predicción en modelo BYM reducido, 95 %. 77
3.15. Pruebas de significancia espacial en residuos de modelo BYM. 77
3.16. Resultado de modelo Lee-Mitchell. 80

A.1. Ajuste en variogramas para precipitación, semana 1-4. 82

A.2. Ajuste en variogramas para precipitación, semanas 31 a 34. 83
A.3. Ajuste en variogramas de temperatura mínima. 84
A.4. Ajuste en variogramas de temperatura máxima. 85

C.1. Intervalo de predicción para parámetros en GLM. 91
C.2. Intervalo de predicción para parámetros en modelo independiente. 92
C.3. Intervalo de predicción para parámetros en modelo intrínseco. 93
C.4. Intervalo de predicción para parámetros en modelo Besag-York-Mollie. 94
C.5. Intervalo de predicción para parámetros en modelo Leroux. 95
C.6. Intervalo de predicción para parámetros en GLM reducido. 96
C.7. Intervalos de predicción para parámetros en modelo independiente reducido. 97
C.8. Intervalo de predicción para parámetros en modelo intrínseco reducido. 98
C.9. Intervalos de predicción para parámetros en modelo Leroux reducido. 99

Índice de figuras

1.1. Representación esquemática del variograma típico. 8

2.1. Mapa de colores para riesgo relativo en semana 1. 43
2.2. Altitud usada para cada municipio de Puerto Rico. 44
2.3. Porcentaje de pobreza según Censo de 2010. 45
2.4. Serie del EVI en 2014 para Adjuntas. 47
2.5. Histogramas para Índice de vegetación. 47
2.6. Diagramas de Caja para Índice de vegetación. 48
2.7. Mapa para el Índice de vegetación. 48
2.8. Ubicación espacial de las estaciones y los centros población. 49
2.9. Histograma para los datos de precipitación. 50
2.10. Histograma para los datos de temperatura máxima. 51
2.11. Histograma para los datos de temperatura mínima. 51
2.12. Datos de lluvia transformados en semana 1. 53
2.13. Datos de lluvia versus coordenadas en la semana 1. 53
2.14. Precipitación en semana 1. 55
2.15. Datos temperatura mínima en semana 1. 56
2.16. Temperatura mínima respecto a las coordenadas, semana 1. 57
2.17. Temperatura mínima en semana 1. 58
2.18. Datos de temperatura máxima en semana 1. 58
2.19. Temperatura máxima respecto a las coordenadas, semana 1. 59

3.1. Traceplot para parámetros de Prec y Tmin en semana 1, modelo BYM. 71
3.2. Comportamiento de la media en parámetros de Prec y Tmin, modelo BYM. 72
3.3. Gráficos de dispersión de variables vs Riesgo Relativo, semana 1. 73
3.4. Casos, casos estimados y residuos de predicción en semana 1. 75
3.5. Efecto aleatorio espacial de riesgo relativo, semana 1 a semana 4. 76
3.6. Superficie de riesgo de los datos en semana 1. 78
3.7. Superficie de riesgo estimado en semana 1 con disimilitud de pobreza. 79

3.8. Superficie de riesgo estimado en semana 1 con disimilitud de altitud. 79

B.1. Precipitación en semana 2. 86
B.2. Precipitación en semana 3. 86
B.3. Precipitación en semana 4. 87
B.4. Precipitación en semana 31. 87
B.5. Precipitación en semana 32. 87
B.6. Precipitación en semana 33. 88
B.7. Precipitación en semana 34. 88
B.8. Temperatura mínima en semana 2. 88
B.9. Temperatura mínima en semana 3. 89
B.10. Temperatura mínima en semana 4. 89
B.11. Temperatura mínima en semana 31. 89
B.12. Temperatura mínima en semana 32. 90
B.13. Temperatura mínima en semana 33. 90
B.14. Temperatura mínima en semana 34. 90

D.1. Riesgo relativo en semana 2. 100
D.2. Riesgo relativo estimado y fronteras en semana 2. 100
D.3. Riesgo relativo en semana 3. 101
D.4. Riesgo relativo estimado y fronteras en semana 3. 101
D.5. Riesgo relativo en semana 4. 102
D.6. Riesgo relativo estimado y fronteras en semana 4. 102
D.7. Riesgo relativo en semana 31. 103
D.8. Riesgo relativo estimado y fronteras en semana 31. 103
D.9. Riesgo relativo en semana 32. 104
D.10. Riesgo relativo estimado y fronteras en semana 32. 104
D.11. Riesgo relativo en semana 33. 105
D.12. Riesgo relativo estimado y fronteras en semana 33. 105
D.13. Riesgo relativo en semana 34. 106
D.14. Riesgo relativo estimado y fronteras en semana 34. 106

E.1. Dispersión de residuos de modelo BYM para las semanas 1 a 4. 107

E.2. Dispersión de residuos de modelo BYM para semanas 31 a 34. 107

Capítulo 1

Marco Teórico

1.1. Inferencia Bayesiana

La inferencia bayesiana supone que las cantidades desconocidas o parámetros, son variables aleatorias, y que los datos, una vez que se observan, son fijos. Por esta razón, la estimación bayesiana consiste en encontrar una distribución de probabilidad completa para dichos parámetros

Se le conoce como inferencia bayesiana al proceso de ajustar un modelo de probabilidad sobre un conjunto de datos, luego resumir dicho ajuste o resultado por medio de una distribución de probabilidad dada para los parámetros del modelo, también sobre las cantidades sin observar aún. El análisis bayesiano de datos es un conjunto de métodos prácticos que permite realizar inferencia a partir de observaciones, usando modelos de probabilidad tanto en la fase previa como en la posterior de la información. La característica esencial en los métodos bayesianos, es que usan explícitamente modelos de probabilidad, con el fin de cuantificar la incertidumbre que hay en la inferencia basada en el análisis preliminar de los datos.

El proceso de análisis bayesiano se puede idealizar haciendo una partición de tres fases según [7]. Estas fases son:

a. Definir un modelo de probabilidad completo: una distribución de probabilidad conjunta para todas las cantidades observables y las no observables en el problema. Dicho modelo debe ser consistente con lo que se conoce del problema científico subyacente, así como con el proceso de recolección de datos.

b. Condicionamiento sobre los datos observados: calcular e interpretar una distribución posterior apropiada, es decir, la distribución de probabilidad condicional de los datos no observados de último interés.

c. Evaluar el ajuste del modelo y las implicaciones de la distribución posterior resultante: ¿se ajusta el modelo a los datos?, ¿son razonables las conclusiones obtenidas?, y, ¿qué tan sensibles

son los resultados con respecto al modelo de probabilidad de la fase (a)?

Como se mencionó antes, la inferencia bayesiana tiene como rasgo o característica central, el poder cuantificar de manera directa la incertidumbre, lo que significa que en muchas ocasiones estos modelos se comportan mejor en situaciones de sobreajuste, y también el poder especificar modelos de probabilidad de varios niveles (capas), lo que se conoce como Modelos Jerárquicos.

1.2. Interpolación Espacial Óptima

En esta sección se presentan algunos elementos que son esenciales para los modelos espaciales, el análisis clásico y el análisis de datos georeferenciados. El concepto fundamental que subyace en la teoría es que se tiene un proceso estocástico $\{Y(x) : x \in D\}$, donde D es un subconjunto del espacio euclídeo $\mathbb{R}^2$, en este caso particular, D representaría el conjunto de coordenadas de Puerto Rico (Longitud, Latitud).

La finalidad del capítulo 2 sección 2.3 es realizar interpolación espacial óptima, a las variables climáticas, creando modelos que se ajusten a las observaciones, así como hacer inferencia sobre estos procesos espaciales en puntos de D donde no hay observaciones. Para el kriging, el proceso estocástico viene dado por:

$$Y(x) = \mu(x) + S(x) + \xi, \tag{1.1}$$

donde:

- x = coordenadas o ubicación espacial, es decir: $x \in D$,
- Y = variable de estudio y observada,
- μ = componente de la media del modelo,
- S = proceso estacionario con distribución normal, con varianza σ^2 llamada "silo parcial", y una función de correlación parametrizada en su forma más simple por ϕ (parámetro de rango),
- ξ = término que representa el error, con varianza τ^2 (varianza nugget).

Sin embargo, para explicar el proceso de interpolación, primero se deben describir algunos de los conceptos que son necesarios para dicho proceso, además esta descripción es más de índole explicativa y sin rigurosidad matemática, para ver de manera más rigurosa la teoría del kriging se invita al lector a revisar [39].

1.2.1. Datos de tipo espacial

En [20] se clasifica a los conjuntos de datos espaciales, en uno de los siguientes 3 tipos básicos:

- **Puntualmente referenciados:** se refiere a los datos donde $Y(x)$ es un vector aleatorio en la ubicación $x \in \mathbb{R}^2$, donde x varía de forma continua en D, un subconjunto fijo de $\mathbb{R}^2$. Son conocidos como datos geoestadísticos o geocodificados, se asume que la covarianza entre dos variables aleatorias en ubicaciones distintas depende de la distancia entre ellas. Un ejemplo utilizado con frecuencia para la covarianza es el modelo exponencial, donde la covarianza entre dos medidas es una función exponencial de la distancia, es decir,

$$Cov[Y(x_i), Y(x_j)] = C(d_{ij}) = \sigma^2 e^{-\frac{d_{ij}}{\phi}}$$

  para $i \neq j$, donde d_{ij} representa la distancia entre x_i y x_j, σ^2 es el parámetro de "silo parcial", y ϕ es el parámetro del rango. Es claro que $d_{ij} = 0$ si $i = j$, por lo que $C(d_{ii}) = Var[Y(x_i)]$, pero esta varianza se expande con frecuencia al valor $\sigma^2+\tau^2$, donde τ^2 es llamado la varianza nugget (efecto nugget), y a la varianza total se le llama "silo". Si se agrega un modelo de distribución conjunta al supuesto anterior de covarianza, entonces se puede realizar inferencia usando una función de verosimilitud. Por lo general, para estos supuestos de varianza y covarianza es conveniente asumir una distribución normal multivariada para los datos, ya que así se dará un vínculo entre modelos que son conjugados, esto es que la forma funcional es similar. Es decir, suponiendo que se tienen las observaciones $Y \equiv \{Y(x_i)\}$, para locaciones conocidas x_i, $i = 1, \cdots, n$, se asume luego que:

$$Y|\mu, \Theta \sim \mathcal{N}_n[\mu, \Sigma(\Theta)].$$

  Para especificar la covarianza, basta con $\Theta = (\tau^2, \sigma^2, \phi)'$, ya que la matriz depende directamente de estos parámetros. La elección más simple para Σ corresponde a la que conlleva a modelos "isotrópicos", donde se asume que la correlación espacial es una función que solo depende de la distancia.

- **Datos de Área:** En este tipo de datos se acostumbra a denotar las regiones por B_i, y los datos propiamente son sumas o promedios de variables en esas regiones. Ahora, si se desea introducir una asociación espacial, por lo general se define un estructura de vecindarios basado en la

estructura u orden de las regiones en el mapa. Una vez que se define dicha estructura, se consideran modelos semejantes a los modelos de series de tiempo autoregresivos. Dos modelos muy populares que incorporan esta información de vecindarios, y que serán explicados con más detalle en la sección 1.4, son los modelos Simultaneamente y Condicionalmente Autoregresivos (SAR y CAR respectivamente, por sus siglas en inglés). El modelo SAR se basa en verosimilitud, por lo que es conveniente en el sentido computacional si se usan métodos con funciones de verosimilitud, mientras que el modelo CAR se basa en métodos bayesianos, así que es mejor usar el Muestreo de Gibbs en conjunto con el ajuste de modelos bayesianos, y en ese sentido se usa con frecuencia un vector de efectos aleatorios variando espacialmente para incorporar correlación espacial, $\phi = (\phi_1, \cdots, \phi_n)'$. Por ejemplo, si $Y_i \equiv Y(B_i)$, se podría asumir que $Y_i \sim \mathcal{N}(\phi_i, \sigma^2)$, siendo independientes, de esa forma, el modelo CAR impuesto sería:

$$\phi_i|\phi_{-i} \sim \mathcal{N}\left(\mu + \sum_{j=1}^{n} a_{ij}(\phi_i - \mu), \tau_i^2\right),$$

donde:

- $\phi_{-i} = \{\phi_j : j \neq i\}$,
- τ_i^2 es la varianza condicional,
- a_{ij} son constantes conocidas o no, tales que $a_{ii} = 0,\ i = 1, \cdots, n$.

En la subsección 1.4.2 se tratará con más detalle este tema.

- **Datos de procesos puntuales:** En un modelo de procesos puntuales, el dominio espacial D es aleatorio, de modo que los elementos del conjunto de índices D, representan la ubicación de eventos aleatorios que constituyen el patrón de puntos espaciales. $Y(x)$ con frecuencia es igual a la constante 1 para cada x en D (indicando la ocurrencia del evento en D), además si se toman algunas características adicionales en cada punto (información adicional covariada) se dice que los datos constituyen un proceso puntual marcado.

  Algunas preguntas de interés en este tipo de datos por lo general se centran en si los datos son agrupados (clustered) en mayor o menor cuantía de lo que puede esperarse si la ubicación de cada punto fuera totalmente determinada por el azar. Estocásticamente, dicha uniformidad se describe con frecuencia usando un *Proceso de Poisson Homogéneo*, lo que implica que el número de ocurrencias en la región A es dada por $\lambda|A|$, donde λ es un parámetro de intensidad del

proceso y $|A|$ es el área de la región. Este tipo de datos no son de interés para el presente trabajo, para un trato más profundo del tema, el lector puede referirse a [20].

1.2.2. El supuesto de estacionariedad

Si no se hacen supuestos restringiendo la clase de campos aleatorios que deseamos considerar, hacer inferencias sobre sus leyes de probabilidad a partir de una simple observación, no tiene sentido. Un supuesto común utilizado para simplificar las cosas, es el de suponer que la estructura probabilística luce similar en diferentes lugares del espacio $\mathbb{R}^d$, con $d \in \mathbb{N}$.

Se dice que un proceso $S(x)$ es estrictamente estacionario [39], si para cada valor k, para todas las ubicaciones $x_1, x_2, \cdots, x_k$, para todos los subconjuntos $C_1, C_2, \cdots . C_k$, y para cualquier vector $h \in \mathbb{R}^d$, se cumple que:

$$P(S(x_1) \in C_1, \cdots, S(x_k) \in C_k) = P(S(x_1 + h) \in C_1, \quad \cdot, S(x_k + h) \in C_k). \tag{1.2}$$

El proceso es estacionario en el sentido que la distribución conjunta de este, evaluada en cualquier conjunto de puntos, no cambia si todos los puntos se mueven en la misma dirección. De hecho,

$$E[S(x)] = E[S(x + h)], \ \ \forall h \in \mathbb{R}^d. \tag{1.3}$$

Por lo anterior se puede concluir que $E[S(x)] = \mu$, es decir, la media del proceso es constante. Además, para cualesquiera dos puntos $x, y \in \mathbb{R}^d$, y cualquier vector $h \in \mathbb{R}^d$, se cumple que:

$$\sigma(x, y) = \sigma(x + h, y + h), \tag{1.4}$$

en particular, si $h = -y$, se tiene que $\sigma(x, y) = \sigma(x - y, 0)$, por lo que la función de covarianza se puede ver como una función de $x - y$ solamente. Para indicar lo anterior, se define $h = x - y$, y se escribe:

$$\sigma(x, y) = \sigma(x - y) = \sigma(h). \tag{1.5}$$

Por definición se puede ver que $\sigma(x, y) = \text{Cov}(S(x), S(y)) = \text{Cov}(S(y), S(x)) = \sigma(y, x)$. En particular, para un proceso estacionario se tiene que:

$$\sigma(x - y) = \sigma(y - x)$$

Un proceso estacionario de segundo orden (débilmente estacionario), es cualquier proceso con media constante, y que cumple (1.5).

1.2.3. Variograma y semivariograma

Existe otro tipo de estacionariedad, llamada estacionariedad intrínseca [20]. Se asume en este caso que $E[S(x+h) - S(x)] = 0$, y se define:

$$E[S(x+h) - S(x)]^2 = \text{Var}[S(x+h) - S(x)] = 2\gamma(h). \tag{1.6}$$

La ecuación (1.6) tiene sentido solamente si el lado izquierdo depende solo de h, no de la elección particular de x, si ese fuera el caso, decimos que el proceso es intrínsecamente estacionario. A la función $2\gamma(h)$ se le llama variograma, y a $\gamma(h)$ se le llama semivariograma.

Es fácil obtener una relación entre el variograma y la función de covarianza,

$$\begin{aligned} 2\gamma(h) &= Var[S(x+h) - S(x)] \\ &= Var[S(x+h)] + Var[S(x)] - 2Cov[S(x+h), S(x)] \\ &= C(0) + C(0) - 2C(h) \\ &= 2(C(0) - C(h)). \end{aligned}$$

Por lo tanto:

$$\gamma(h) = C(0) - C(h). \tag{1.7}$$

Es claro de la ecuación anterior, que si se tiene la función de covarianza, se puede tener el variograma. Sin embargo, para el recíproco se necesita hacer un supuesto más: si el proceso espacial es ergódico, entonces $\lim_{|h|\to\infty} C(h) = 0$, esto significa que la covarianza entre dos puntos se anula, cuando estos puntos se alejan lo suficiente en el espacio. Con dicha condición, se obtiene que $\lim_{u\to\infty} \gamma(u) = C(0)$, por lo que si despejamos $C(h)$ en (1.7) y escribimos $C(0)$ como un límite, se obtiene:

$$C(h) = \lim_{u\to\infty} \gamma(u) - \gamma(h). \tag{1.8}$$

En general, el límite del lado derecho no necesariamente existe, pero en caso de existir, se concluye que el proceso es débilmente estacionario, con la función de covarianza dada por la ecuación (1.8).

Por lo anterior, para un proceso estacionario o intrínseco el variograma se reduce a una función de h. Las propiedades del segundo momento de un proceso estocástico estacionario, $S(x)$, se pueden describir por su función de covarianza, $C(h)$, o por su variograma, $\gamma(h)$; otra equivalencia entre estos se puede ver en la ecuación:

$$\gamma(h) = C(0) - C(h) = \sigma^2(1 - \rho(h)) \tag{1.9}$$

donde $\sigma^2 = Var[S(x)]$ y $\rho(h) = Corr[S(s+h), S(x)]$.

Con frecuencia, se puede asumir que las propiedades del segundo momento de un proceso dependen solamente de la distancia entre dos puntos, $|h|$, y no de la dirección del vector formado entre ellos. Así, un proceso estacionario de segundo orden es isotrópico si $C(h) = C(|h|)$, mientras que un proceso intrínsecamente estacionario es isotrópico si:

$$\gamma(h) = \gamma(|h|).$$

Mientras que un proceso que no es isotrópico, se dice que es anisotrópico. En el presente trabajo se asumirá que los procesos son isotrópicos.

Los procesos isotrópicos son populares debido a su simplicidad, interpretabilidad, y en particular, debido a que existen fórmulas paramétricas relativamente simples disponibles como candidatos para el semivariograma.

La interpolación que se realizará para la temperatura y la precipitación usando kriging, se basarán en un variograma; sin embargo, para hacer esto se debe ajustar un variograma empírico (desde los datos) con alguno de los modelos que se verán más adelante. Suponga que los datos $(x_i, y_i) : i = 1, \cdots, n$ son generados por un proceso estacionario,

$$Y_i = S(x_i) + Z_i, \tag{1.10}$$

donde los Z_i son mutuamente independientes e idénticamente distribuidos, con media cero y varianza τ^2. Se define el variograma del proceso de observaciones (variograma teórico, [16]) como:

$$\gamma_Y(h_{ij}) = \frac{1}{2}E[(Y_i - Y_j)^2], \tag{1.11}$$

donde $h_{ij} = |x_i - x_j|$. Por lo anterior, se concluye que:

$$\gamma_Y(h) = \tau^2 + \sigma^2(1 - \rho(h)). \tag{1.12}$$

Según [16] el variograma típico es una función monótona creciente, con las siguientes características: el intercepto, τ^2, que corresponde a la varianza nugget, la asíntota, $\tau^2 + \sigma^2$, que corresponde a la varianza del proceso de observaciones, también llamado sill (o silo, en español), la cual es la suma de la varianza nugget y la varianza de la señal, σ^2. La forma en la cual el variograma crece desde el intercepto a la asíntota, depende de la función de correlación, $\rho(h)$; entre sus características más relevantes están, su comportamiento cerca de $h = 0$, el cual se refiere a la suavidad analítica del proceso subyacente, y qué tan rápido $\rho(h)$ se aproxima a cero cuando $|h|$ crece, que refleja la

extensión física de la correlación espacial en el proceso. Cuando $\rho(h) = 0$ para algún valor finito de $|h|$, este valor es conocido como el rango del variograma. Así, cuando $\rho(h)$ se aproxima a cero asintóticamente cuando $|h|$ aumenta, el rango no está definido. Por eso, en tales casos se define el "practical range"(rango práctico o simplemente rango en español) como la distancia $|h_0|$ en la cual se cumple que $\rho(h_0) = 0{,}05$. A continuación se presenta una representación esquemática de un variograma típico:

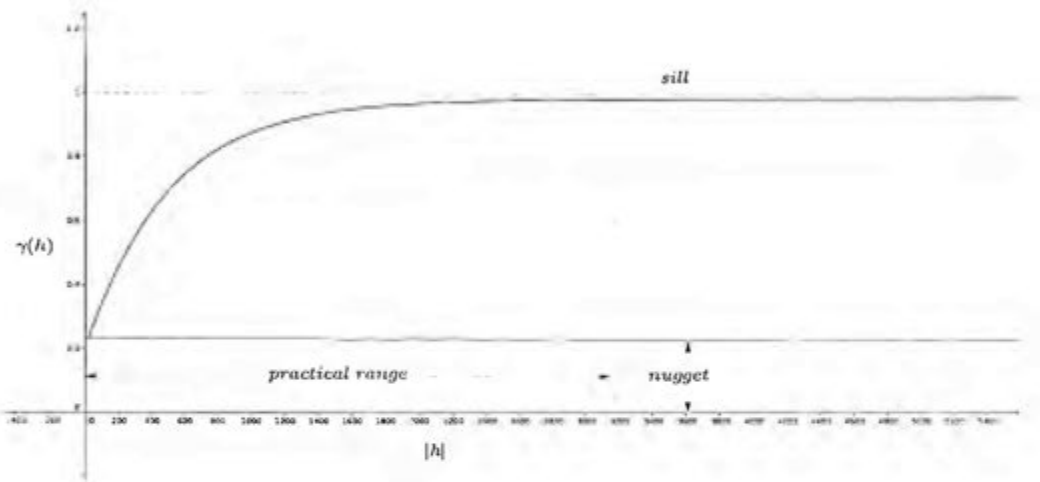


Figura 1.1: Representación esquemática del variograma típico.

La varianza nugget, que en el contexto actual es igual al intercepto de $\gamma(h)$, es un parámetro importante para la predicción espacial. El valor de τ^2 afectará el grado al cual la superficie predecida, $S(x)$, rastreará los datos observados Y_i. En particular, hacer $\tau^2 = 0$ forzará las predicciones espaciales a interpolar los datos. Por lo anterior, decidir si $\tau^2 = 0$ o estimar un valor positivo de este parámetro, es importante al elegir alguna de las familias de la sección (1.2.4).

Desde una perspectiva analítica de datos, la definición de variograma teórico en (1.11) es importante ya que implica que, siempre en el supuesto de estacionariedad, las cantidades observadas $\gamma_{ij} = \frac{1}{2}(Y_i - Y_j)^2$ son estimadores no sesgados de las correspondientes ordenadas del variograma, $\gamma_Y(h_{ij})$. La colección de pares de distancias y las debidas ordenadas del variograma $(h_{ij}, \gamma_{ij}) : j > i$ es llamada variograma empírico de los datos $(x_i, Y_i) : i = 1, \ldots, n$. [16]

El estimador clásico del variograma propuesto por Matheron en 1962 [14] es:

$$2\widehat{\gamma}(h) = \frac{1}{|N(h)|} \sum_{N(h)} (Y_i - Y_j)^2 \tag{1.13}$$

donde la suma es sobre $N(h) = \{(i,j) : x_i - x_j = h\}$ y $|N(h)|$ es el número de elementos distintos

de dicho conjunto. Es un estimador insesgado, sin embargo tiene propiedades muy pobres [14], por ejemplo es fuertemente afectado por observaciones atípicas debido al término $(\cdot)^2$ de (1.13).

1.2.4. Algunas familias de funciones de covarianza

La forma más común del comportamiento empírico, para una estructura de covarianza estacionaria, es que la correlación entre $S(x)$ y $S(y)$ va a disminuir conforme la distancia, $h = |x - y|$, aumenta. Por lo que es natural buscar modelos de correlación teóricos, con una estructura que se comporte así.

El que algunas familias paramétricas de funciones sean definidas positivas, es una condición necesaria y suficiente para que éstas definan una clase de familias de covarianza; sin embargo, dicha condición no es algo sencillo de probar directamente. Por este motivo, es razonable contar con un conjunto de familias de funciones que son definidas positivas, además que sean flexibles como para cumplir con las necesidades de algunos datos geoestadísticos. Se presentan algunos modelos isotrópicos de covarianza y variograma como funciones de $|h|$ (en adelante, h, entendiendo que es función sólo de la distancia), para conocer más modelos se puede revisar [13]. Para variogramas asociados con una covarianza, se da la forma analítica de la covarianza y se deduce el variograma desde $\gamma(h) = C(0) - C(h)$, además para los modelos descritos se toma $C(0) = 1$

Modelos esféricos y sus derivados

Por autoconvolución de la función indicadora de la esfera de $\mathbb{R}^d$ con diámetro ϕ [13], es decir, en términos del módulo $r = |h|$, de la función

$$w_d(r) = \begin{cases} 1 & \text{si } r \leq \phi/2, \\ 0 & \text{si } r > \phi/2 \end{cases} \tag{1.14}$$

se obtiene el covariograma esférico de $\mathbb{R}^d$, el cual se puede considerar como una función de $r = |h|$:

$$g_d(r) = \begin{cases} \phi^d v_{d-1} \int_{r/\phi}^1 (1-u^2)^{(d-1)/2} du & \text{si } r \leq \phi, \\ 0 & \text{si } r \geq \phi. \end{cases} \tag{1.15}$$

En la fórmula anterior v_d es el volumen de la bola de diámetro unitario de $\mathbb{R}^d$, el cual viene dado por

$$v_d = \frac{\pi^{d/2}}{2^{d-1} d \Gamma(d/2)}.$$

Los modelos usados en la práctica, corresponden a $d = 1, 2, 3$:

- *Modelo triangular,* válido en $\mathbb{R}$:

$$C_1(|h|) = \begin{cases} 1 - \frac{|h|}{\phi} & \text{si } |h| \leq \phi, \\ 0 & \text{si } |h| \geq \phi. \end{cases}$$

- *Modelo circular,* válido en $\mathbb{R}^2$:

$$C_2(|h|) = \begin{cases} \frac{2}{\pi}\left(\arccos(|h|/\phi) - \frac{|h|}{\phi}\sqrt{1 - \frac{|h|^2}{\phi^2}}\right) & \text{si } |h| \leq \phi, \\ 0 & \text{si } |h| \geq \phi. \end{cases}$$

- *Modelo esférico,* válido en $\mathbb{R}^3$

$$C_3(|h|) = \begin{cases} 1 - \frac{3|h|}{2\phi} + \frac{|h|^2}{2\phi^3} & \text{si } |h| \leq \phi, \\ 0 & \text{si } |h| \geq \phi. \end{cases}$$

Los variogramas correspondientes se comportan de forma lineal cerca del origen y alcanzan el sill en $|h| = \phi$, por lo que el parámetro de escala, ϕ, coincide con el rango.

Modelo exponencial

En [13] se define el modelo exponencial con parámetro de escala $\phi > 0$ por:

$$C(|h|) = exp\left(-\frac{|h|}{\phi}\right). \tag{1.16}$$

Este modelo es una covarianza en $\mathbb{R}^d$ para cualquier valor natural de d, ya que su densidad espectral correspondiente es (ver [39], sección 2.5):

$$f_d(\rho) = \frac{2^d \pi^{(d-1)/2} \Gamma((d+1)/2) \phi^d}{(1 + 4\pi^2\phi^2\rho^2)^{(d+1)/2}} \tag{1.17}$$

En este caso, el variograma alcanza su sill solo de forma asintótica cuando $|h| \to \infty$, y su rango práctico (95 % del sill) es cercano a 3ϕ.

Modelo de Cauchy y Clase de Cauchy

En [13] el modelo de Cauchy, con parámetro de escala ϕ, es dado por:

$$C(|h|) = \left(1 + \frac{|h|^2}{\phi^2}\right)^{-\beta/2} \quad (\phi > 0,\ \beta \geq 0), \tag{1.18}$$

el cual es un modelo válido en $\mathbb{R}^d$ para todo d. Este tipo de modelo es muy regular cerca del origen, ya que su expansión en series de Taylor contiene solamente términos pares, y alcanza el sill muy lentamente. La clase de Cauchy generaliza el modelo de Cauchy a comportamientos en $|h|^\alpha$ en el origen, es decir que su expansión en serie de Taylor en $h = 0$ se presenta en potencias de $|h|^\alpha$, con una covarianza de la forma:

$$C(|h|) = \left(1 + \frac{|h|^\alpha}{\phi^\alpha}\right)^{-\beta/\alpha} \quad (\phi > 0, 0 < \alpha \leq 2,\ \beta \geq 0), \tag{1.19}$$

donde α es un parámetro de la forma mientras β parametriza la dependencia en distancias grandes.

Modelo Matérn

La función isotrópica (1.18) es positiva e integrable en $\mathbb{R}^d$ para todo d. Por lo tanto, su transformada de Fourier d-dimensional es una covarianza en $\mathbb{R}^d$ (ver [13], secciones 2.3.3, 2.3.4). Al considerar el caso $\beta = 2\kappa + d$, se obtiene el modelo Matérn (también conocido como el modelo K-Bessel):

$$C(|h|) = \frac{1}{2^{\kappa-1}\Gamma(\kappa)} \left(\frac{|h|}{\phi}\right)^\kappa K_\kappa \left(\frac{|h|}{\phi}\right) \quad (\phi > 0,\ \kappa \geq 0) \tag{1.20}$$

donde K_κ es la función de Bessel modificada de segundo tipo de orden κ en $\mathbb{R}$ dada por:

$$K_\kappa(x) = \frac{\pi}{2} \frac{\mathbb{I}_{-\kappa}(x) - \mathbb{I}_\kappa(x)}{\sin(\kappa\pi)}.$$

siendo $\mathbb{I}_\kappa(x)$ la función de Bessel modificada de primera especie y de orden κ, la cual viene dada por:

$$\mathbb{I}_\kappa(x) = \sum_{j=0}^{\infty} \frac{1}{j!\Gamma(j + \kappa + 1)} \left(\frac{x}{2}\right)^{2j+\kappa},$$

κ es un parámetro de la forma, el cual determina la suavidad analítica del proceso subyacente $S(x)$. Específicamente, $S(x)$ es $\lceil |\kappa - 1| \rceil$-veces diferenciable en media cuadrática. Según [16], este modelo puede tener cualquier tipo de comportamiento cerca del origen, pues su principal término irregular se comporta como $|h|^{2\kappa}$ si κ no es un entero y como $|h|^{2\kappa}\log(|h|)$ si κ es un entero. Además, para $\kappa = 0{,}5$ la función de covarianza Matérn se reduce a la familia exponencial, mientras que si hacemos $\kappa \to \infty$, el modelo que se obtiene es conocido como gaussiano.

Note que los parámetros ϕ y κ son no ortogonales, en el siguiente sentido. Si la estructura de covarianza es Matérn con parámetros ϕ y κ, entonces la aproximación que mejor se ajusta con orden $\kappa^* \neq \kappa$ también tendrá $\phi^* \neq \phi$. Es decir, los parámetros de escala que corresponden a distinto orden de la covarianza Matérn, no son directamente comparables. La relación entre el "practical range" y el parámetro ϕ depende por lo tanto de κ.

1.2.5. Ajustando un modelo de variograma

Después de haber visto una pequeña selección de modelos para el variograma, uno debería preguntarse cómo escoger alguno de los modelos para algún conjunto de datos dado, o si los datos pueden distinguirlos. Históricamente, según [20] un modelo de variograma se elige graficando el *variograma empírico*, un estimador simple no parámetrico del variograma, y luego se compara a varias formas teóricas disponibles. El variograma empírico habitual es:

$$\widehat{\gamma}(t) = \frac{1}{2|N(t)|} \sum_{(x_i,x_j)\in N(t)} [Y(x_i) - Y(x_j)]^2 \tag{1.21}$$

donde $N(t)$ es el conjunto de pares de puntos tales que $||x_i - x_j|| = t$, y $|N(t)|$ es el número de pares en dicho conjunto. Note que, al menos que las observaciones caigan en una malla regular, las distancias entre los pares de ubicaciones podrían ser todas diferentes, por lo que este no será un estimador muy útil tal como está. En su lugar, se puede definir una rejilla para las distancias en intervalos $I_1 = (0, t_1)$, $I_2 = (t_1, t_2)$, y así sucesivamente hasta $I_K = (t_{K-1}, t_K)$ para alguna rejilla (posiblemente regular) $0 < t_1 < \cdots < t_K$. Representando los valores de t en cada intervalo por su punto medio, se puede alterar la definición de $N(t)$ a:

$$N(t_k) = \{(x_i, x_j) : ||x_i - x_j|| \in I_k\},\ k = 1, ..., K.$$

1.2.6. Kriging: el método y sus ecuaciones

Un problema central en estadística espacial es la estimación de una variable de interés sobre cierto dominio, sobre la base de valores observados en un número limitado de puntos. Típicamente, el investigador puede querer construir un modelo de cuadrícula con el objetivo de dibujar un mapa de contornos, hacer un inventario y calcular cantidades (contaminantes, especies animales, etc) dentro de ciertas unidades de área. Con más generalidad, la cantidad deseada puede ser una función de la variable observada, funciones que se restringen a tipo lineal (kriging lineal).

Desde un punto de vista determinístico ese es un problema de interpolación. La variable de interés es aproximada por una función paramétrica cuya forma se postula por adelantado, ya sea explícita o implícitamente. Los parámetros son seleccionados para optimizar algún criterio de mejor ajuste a los datos puntuales.

En el presente trabajo se realizará con un enfoque probabilístico, conocido como *kriging*, un término usado por George Matheron en 1963, en honor a Danie G. Krige. Este método produce una

interpolación basada en un modelo de variograma o de covarianza derivado desde los datos, en lugar de un modelo a priori de la función de interpolación.

En [13] se presenta la siguiente tabla, con las principales formas de kriging lineal, al decir lineal se refiere al hecho de que los predictores son funciones lineales de las observaciones [39]:

Tipo de kriging	Media	Modelo derivado	Pre-requisito
Simple (SK)	Conocida	Ninguno	Covarianza
Ordinario (OK)	Desconocida	Constante	Variograma
Universal (UK)	Desconocida	Función de las coordenadas	Variograma
Externo (KED)	Desconocida	Variable externa	Variograma

Cuadro 1.1: Principales formas de kriging lineal.

Antes de comenzar a explicar los métodos, se debe aclarar un poco la terminología, la palabra predicción la establecemos como la determinación del valor de cierta cantidad aleatoria, mientras que estimación se refiere a la inferencia sobre algún parámetro fijo pero desconocido de algún modelo. Sin embargo, acá serán usadas indistintamente refieriéndose a estimación.

Notación y supuestos

- Se le denota por $\{Y(x) : x \in D \subset \mathbb{R}^d\}$ a la función aleatoria usada como un modelo para la variable regionalizada de interés, $\{y(x) : x \in D \subset \mathbb{R}^d\}$ denota a la realidad, es decir, $y(x)$ es una realización de $Y(x)$.

- T denota el conjunto de puntos donde $Y(x)$ fue muestreado. En nuestro caso, T es finito y consistirá de N puntos denotados por:

$$T = \{x_n : n = 1. \cdots, N\}.$$

- El valor de las funciones en puntos muestreados, son referenciados por los subíndices de dichos puntos, por ejemplo:

$$Y_i = Y(x_i), \text{ los datos,}$$
$$\mu_i = \mu(x_i), \text{ valor medio de } Y_i,$$
$$\sigma_{ij} = \sigma(x_i, x_j) \text{ covarianza entre } Y_i \text{ y } Y_j.$$

- Los estimadores hechos con kriging son marcados con un asterisco de superíndice(*). Explícitamente, el estimador de $Y(x_0)$ es de la forma:

$$Y^*(x_0) = \sum_{i=1}^{n} \lambda_i(x_0) Y(x_i) + \lambda_0(x_0)$$

donde $\lambda_i(x_0)$ es un peso puesto sobre $Y(x_i)$ y $\lambda_0(x_0)$ es una constante que depende de x_0. Se debe tener en mente que los pesos λ_i dependen de la ubicación x_0 donde la función está siendo estimada.

- En cuanto a los vecindarios, se debe aclarar que la teoría se deriva siempre como si todos los N puntos de la muestra serán usados para hacer la estimación, este caso es llamado interpolación con vecindario global. En la práctica, N puede ser lo suficientemente grande como para permitir el uso de "vecindarios movibles" o "vecindarios locales", es decir, usar solamente un subconjunto de los datos para hacer la estimación en cada nodo de la malla que se desea interpolar. Se debe tener presente que eso podría alterar la relación entre las estimaciones de diferentes nodos, e incluso introducir discontinuidad.

Kriging con media conocida

En esta parte se verá el caso conocido como Kriging Simple [13], donde conocemos la parte media del modelo. Conocer la media nos hace la teoría muy simple y además dota el estimador de kriging con todas las propiedades agradables (consistente, eficiente, ver [13]). En el caso de un proceso gaussiano, esta coincide con la esperanza condicional $E[Y_0|Y_1, \quad , Y_N]$, que es el estimador ideal de Y_0 en el sentido de media cuadrática. En todas las circunstancias el error $Y^* - Y_0$ no está correlacionado con Y_i, para todo i, ni con Y^*.

En el mundo real la media puede ser conocida solamente si hay repeticiones del fenómeno, como lo es el caso de procesos espacio-temporales, o cuando el número de datos es tan grande que se puede estimar la media casi a la perfección.

Para derivar las ecuaciones, se va a considerar el caso de estimación puntual. Se quiere estimar $y_0 = y(x_0)$ a partir de N observaciones $y_1, \cdots, y_N$ usando el estimador afín:

$$y^* = \sum_i \lambda_i y_i + \lambda_0$$

interpretado en el modelo como una realización de la variable aleatoria

$$Y^* = \sum_i \lambda_i Y_i + \lambda_0.$$

La constante λ_0 y los pesos λ_i son elegidos para minimizar en el modelo el error medio cuadrático esperado $E[Y^* - Y_0]^2$. Primero nos concentramos en λ_0; el error medio cuadrático (MSE, por sus siglas en inglés: mean squared error) se puede escribir como

$$E[Y^* - Y_0]^2 = Var[Y^* - Y_0] + (E[Y^* - Y_0])^2.$$

Como las varianzas no son sensibles a las traslaciones, solamente el último término del lado derecho involucra λ_0. Para minimizar el MSE, es necesario elegir λ_0 de forma que cancele el sesgo $E(Y^* - Y_0)$:

$$\lambda_0 = \mu_0 - \sum_i \lambda_i \mu_i.$$

El estimador Y^* se transforma en

$$Y^* = \mu_0 + \sum_i \lambda_i (Y_i - \mu_i) \tag{1.22}$$

Esto equivale a estimar la variable de media nula $Z(x) = Y(x) - \mu(x)$ por el estimador lineal:

$$Z^* = \sum_i \lambda_i Z_i$$

y añadir después la media. Con lo anterior se establece que el caso de media conocida es equivalente al caso de una media nula y $\lambda_0 = 0$, por lo que en adelante, en esta parte, se considera que $Y(x)$ tiene media cero. El MSE se puede expandir en términos de la covarianza centrada $\sigma(x, y)$ de $Y(x)$,

$$E[Y^* - Y_0]^2 = \sum_i \sum_j \lambda_i \lambda_j \sigma_{ij} - 2 \sum_i \lambda_i \sigma_{i0} + \sigma_{00}.$$

El mínimo de esta función cuadrática se obtiene cancelando sus derivadas parciales con respecto a los pesos λ_i, es decir:

$$\frac{\partial}{\partial \lambda_i} E[Y^* - Y_0]^2 = 2 \sum_j \lambda_j \sigma_{ij} - 2\sigma_{i0}.$$

El que lo anterior alcance un mínimo, lo garantiza la propiedad de positividad de la función de covarianza, y el MSE es una función convexa. Los λ_i son soluciones del sistema lineal de N ecuaciones:

$$\sum_j \lambda_j \sigma_{ij} = \sigma_{i0}, \quad i = 1, \ldots, N \tag{1.23}$$

en notación matricial lo anterior se escribe:

$$\Sigma \Lambda = \sigma_0$$

donde $\Sigma = (\sigma_{ij})$ es la matriz $N \times N$ de covarianzas de los datos, $\sigma_0 = (\sigma_{i0})$ es un vector en $\mathbb{R}^N$ de covarianzas entre los datos y el objetivo o estimación, y $\Lambda = (\lambda_i)$ es un vector en $\mathbb{R}^N$ de soluciones. El sistema (1.23), llamado sistema de kriging simple (SK), tiene solución única, dado que la matriz Σ es no singular. Este será siempre el caso si la matriz de covarianzas es estrictamente definida positiva y si todos los puntos muestrales son distintos, lo cual se asumirá.

Kriging con media desconocida

El enfoque de kriging que se presenta en este apartado provee una solución óptima que involucra solamente una estimación por paso. El caso más sencillo se da cuando la media es constante, $\mu(x) = \mu$, llamado Kriging Ordinario (OK, por sus siglas en inglés). El modelo general, conocido como Kriging Universal, asume que la parte media del modelo se puede representar como una función de la superficie [13]:

$$\mu(x) = \sum_{k=0}^{L} a_k f^k(x) \tag{1.24}$$

donde las $f^k(x)$ son funciones básicas conocidas y los a_k son coeficientes fijos pero desconocidos. Usualmente se tiene que $f^0 = 1$, lo que garantiza que el caso de media constante está incluido en el modelo. Las otras funciones son por lo general monomios de bajo grado en las coordenadas de x (en la práctica el grado no excede 2). Note que (1.24) se puede ver como una aproximación local para $\mu(x)$; esto es, los coeficientes a_k pueden variar en el espacio, pero suficientemente lento para ser considerados constantes dentro de los vecindarios de estimación.

En el modelo de kriging universal se descompone la variable $Y(x)$ en la suma

$$Y(x) = \mu(x) + S(x),$$

de una función determinista suave $\mu(x)$ describiendo el aspecto sistemático del fenómeno, y llamada desviación (*drift*), y una función aleatoria de media cero $S(x)$, llamada residuo y que captura las fluctuaciones erráticas.

Kriging Ordinario

Primero vamos a ver como, el no conocer la media, afecta el problema de estimación en el caso de media constante $\mu(x) = a_0$. Considerando de nuevo el estimador afín $Y^* = \sum_i \lambda_i Y_i + \lambda_0$, para el cual según [13], el error medio cuadrático se puede escribir como:

$$E[Y^* - Y_0]^2 = Var[Y^* - Y_0] + \left[\lambda_0 + \left(\sum_i \lambda_i - 1\right) a_0\right]^2.$$

Solamente el término del sesgo en el lado derecho involucra λ_0, pero esta vez no se puede minimizar sin conocer a_0. Una solución podría ser reemplazar a_0 por un estimador $\widehat{a}_0$ y resolver para λ_0, pero dicho estimador necesariamente será dependiente de los datos, por lo que λ_0 podría resultar no constante. La única solución real es establecer $\lambda_0 = 0$ e imponer la condición sobre los pesos λ_i de que $\sum_i \lambda_i - 1 = 0$. Sujeto a esta condición, el error medio cuadrático es igual a la varianza del error $Y^* - Y_0$ y depende solo de las covarianzas:

$$Var[Y^* - Y_0] = \sum_i \sum_j \lambda_i \lambda_j \sigma_{ij} - 2 \sum_i \lambda_i \sigma_{i0} + \sigma_{00}.$$

El problema se puede reformular de la siguiente manera: Encuentre N pesos λ_i que sumen 1 y minimicen $Var[Y^* - Y_0]$. Lo anterior es resuelto por el método de multiplicadores de Lagrange. Considere la función:

$$Q = Var[Y^* - Y_0] + 2\mu \left(\sum_i \lambda_i - 1\right)$$

donde μ es el multiplicador de Lagrange, y determine el mínimo sin restricciónes de Q igualando sus derivadas parciales a 0:

$$\frac{\partial Q}{\partial \lambda_i} = 2 \sum_j \lambda_j \sigma_{ij} - 2\sigma_{i0} + 2\mu = 0, \quad i = 1, \cdots, N$$

$$\frac{\partial Q}{\partial \mu} = 2 \left(\sum_i \lambda_i - 1\right) = 0$$

Que el extremo sea ciertamente un mínimo, se garantiza de nuevo por la convexidad de $Var[Y^* - Y_0]$ como función de λ_0. Lo anterior nos deja el siguiente sistema lineal de $N+1$ ecuaciones e incógnitas:

Sistema de Ecuaciones de Kriging Ordinario

$$\begin{cases} \sum_j \lambda_j \sigma_{ij} + \mu = \sigma_{i0}, \ i = 1, \cdots, N \\ \sum_i \lambda_i = 1 \end{cases} \tag{1.25}$$

La varianza se obtiene premultiplicando las primeras N ecuaciones de (1.25) por λ_i, sumando sobre i, y luego usando la última ecuación. El resultado será:

$$\sigma^2_{OK} = E[Y^* - Y_0]^2 = \sigma_{00} - \sum \lambda_i \sigma_{i0} - \mu \tag{1.26}$$

El sistema lineal tiene solución única si y solo si la matriz de covarianza $\Sigma = [\sigma_{ij}]$ es definida positiva, lo que será el caso si se usa una función de covarianza definida positiva y todos los puntos de los datos son distintos.

Kriging Universal

Se quiere estimar $Y_0 = Y(x_0)$ usando un estimador lineal $Y^* = \sum_i \lambda_i Y_i$, y se busca minimizar el error cuadrático medio, el cual, como es usual lo podemos descomponer como [13]:

$$E[Y^* - Y_0]^2 = Var[Y^* - Y_0] + (E[Y^* - Y_0])^2.$$

Ahora la parte media del modelo no la asumimos constante, pero sí de la forma (1.24). El sesgo se puede expandir como:

$$E[Y^* - Y_0] = \sum_i \lambda_i \sum_k a_k f_i^k - \sum_k a_k f_0^k,$$

donde se usó la notación $f_i^k = f^k(x_i)$, y la convención de que la suma sobre k se extiende sobre todos los posibles valores $k = 0, 1, \cdots, L$. Al cambiar el orden de la suma en k e i, se obtiene

$$E[Y^* - Y_0] = \sum_k a_k \left(\sum_i \lambda_i f_i^k - f_0^k \right).$$

Para minimizar $E[Y^* - Y_0]^2$, se debe hacer $(E[Y^* - Y_0])^2$ cero, independientemente de los coeficientes a_k, lo que implica anular sus factores en lo anterior. Eso conduce a las $L+1$ condiciones:

$$\sum_i \lambda_i f_i^k = f_0^k, \quad k = 0, 1, \cdots, L, \tag{1.27}$$

conocidas como condiciones de universalidad, de donde sale el nombre de kriging universal. Sujeto a esas condiciones, el error medio cuadrático es igual a la varianza del error $Y^* - Y_0$:

$$Var[Y^* - Y_0] = \sum_i \sum_j \lambda_i \lambda_j \sigma_{ij} - 2 \sum_i \lambda_i \sigma_{i0} + \sigma_{00}.$$

Usando multiplicadores de Lagrange, se minimiza:

$$Q = Var[Y^* - Y_0] + 2 \sum_{k=0}^{L} \mu_k \left[\sum_i \lambda_i f_i^k - f_0^k \right]$$

donde μ_k, $k = 0, \cdots, L$, son $L+1$ incógnitas adicionales, los multiplicadores de Lagrange, y se determina el mínimo sin restricción de Q igualando sus derivadas parciales a cero:

$$\frac{\partial Q}{\partial \lambda_i} = 2\sum_j \lambda_j \sigma_{ij} - 2\sigma_{i0} + 2\sum_k \mu_k f_i^k = 0, \quad i = 1, \cdots, N$$

$$\frac{\partial Q}{\partial \mu_k} = 2\left[\sum_i \lambda_i f_i^k - f_0^k\right] = 0, \quad k = 0, \cdots, L.$$

Lo anterior conduce al siguiente conjunto de $N+L+1$ ecuaciones lineales con $N+L+1$ incógnitas:

Sistema de Ecuaciones de Kriging Universal

$$\begin{cases} \sum_j \lambda_j \sigma_{ij} + \sum_k \mu_k f_i^k = \sigma_{i0}, \ i = 1, \cdots, N \\ \sum_i \lambda_i f_i^k = f_0^k, \ k = 0, \cdots, L. \end{cases} \tag{1.28}$$

En notación matricial el sistema (1.28) es de la forma $\mathbf{Aw} = \mathbf{b}$ con la siguiente estructura:

$$\begin{pmatrix} \Sigma & F \\ F' & 0 \end{pmatrix} \begin{pmatrix} \lambda \\ \mu \end{pmatrix} = \begin{pmatrix} \sigma_0 \\ f_0 \end{pmatrix} \tag{1.29}$$

donde Σ, λ y σ_0 se definen como en kriging simple y donde:

$$F = \begin{pmatrix} 1 & f_1^1 & \cdots & f_1^L \\ 1 & f_2^1 & \cdots & f_2^L \\ \vdots & \vdots & \cdots & \vdots \\ 1 & f_N^1 & \cdots & f_N^L \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_0 \\ \mu_1 \\ \vdots \\ \mu_L \end{pmatrix}, \quad f_0 = \begin{pmatrix} 1 \\ f_0^1 \\ \vdots \\ f_0^L \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} \Sigma & F \\ F' & 0 \end{pmatrix}.$$

La varianza kriging se obtiene premultiplicando las primeras N ecuaciones de (1.28) por λ_i, sumando sobre i, y luego usando las últimas $L+1$ ecuaciones. El resultado es la varianza de kriging universal:

$$\sigma_{UK}^2 = E[Y^* - Y_0]^2 = \sigma_{00} - \sum_i \lambda_i \sigma_{i0} - \sum_k \mu_k f_0^k \tag{1.30}$$

también en forma matricial:

$$\sigma_{UK}^2 = \sigma_{00} - \lambda' \sigma_0 - \mu' f_0 = \sigma_{00} - w'b.$$

El sistema lineal (1.28) tiene solución única si y solo si la matriz $\mathbf{A}$ es no singular. Eso se cumple bajo el siguiente conjunto de condiciones: (1) que la submatriz Σ es estrictamente definida positiva,

(2) que la submatriz F sea de rango completo. Se asegura que Σ es definida positiva al utilizar una función de covarianza definida positiva y además eliminando los puntos duplicados en los datos. La condición sobre F expresa que las $L+1$ funciones básicas $f^k(x)$ son linealmente independientes sobre el espacio:

$$\left(\sum_k c_k f^k(x) = 0 \ \forall x\right) \Longrightarrow c_k = 0, \ \ k = 0, \cdots, L.$$

1.3. Modelos lineales generalizados mixtos (glmm)

Los modelos lineales generalizados (GLM) son una extensión de la regresión ordinaria (ver [1], en capítulo 4), ya que permiten variables de respuesta que no son distribuidas normalmente y una función de enlace para su media. Los modelos lineales generalizados mixtos (GLMM, siglas en inglés) son una extensión adicional, ya que permite la presencia de efectos aleatorios y efectos fijos en el predictor lineal.

Sea y_{ik} la observación k en el clúster i, $i = 1, \cdots, K_i$, sea x_{ik} un vector columna que contiene los valores de las variables explicativas para dicha observación, sea u_i un vector de efectos aleatorios para el mismo clúster, z_{ik} un vector columna de variables explicativas para el efecto aleatorio y finalmente, sea $\mu_{ik} = E[Y_{ik}|u_i]$ el predictor lineal, un GLMM tiene la forma:

$$g(\mu_{ik}) = x_{ik}^T\beta + z_{ik}^T u_i, \tag{1.31}$$

donde $g(\cdot)$ es una función de enlace y β es un conjunto de parámetros para los efectos fijos, además se asume que el vector u_i tiene una distribución normal multivariada, $\mathcal{N}(0, \Sigma)$.Los elementos en la diagonal de la matriz de covarianzas Σ son desconocidos, y es posible que los demás elementos en la misma tampoco se conozcan, es decir la correlación entre las entradas de u_i puede no conocerse. Condicionado en u_i, el modelo de ajuste estándar (GLM) trata los $\{y_{ik}\}$ como independientes para i y para k. La variabilidad entre los elementos u_i induce asociaciones no negativas entre las respuestas (ver [1], sección 11.2.2), para la distribución marginal promediada sobre los casos u ocurrencias, lo anterior es causado por los efectos aleatorios compartidos para cada observación en un clúster.

En la ecuación (1.31) el efecto aleatorio entra al modelo en la misma escala que los términos predictores. Por ejemplo, en ocasiones dicho efecto representa heterogeneidad causada por la omisión de ciertas variables explicativas. Considere el caso especial con efecto aleatorio univariado y $z_{ik} = 1$.

Con u_i reemplazado por $u_i^* \sigma$, donde $\{u_i^*\}$ son normales estándar, el GLMM toma la forma:

$$g(\mu_{ik}) = x_{ik}^T \beta + u_i^* \sigma.$$

Esto tiene la forma de un GLM ordinario con valores no observados $\{u_i^*\}$ de una covariable particular. Así, los GLMM están relacionados a métodos para tratar con predictores sin medir u otras formas de datos perdidos. También, los efectos aleatorios se pueden ver aveces como una medida de error de las variables explicativas. Para más detalles, el lector puede revisar los capítulos 4 y 13 de [1].

1.4. Los Modelos para Datos de Área: CAR y SAR

1.4.1. Herramientas de exploración para datos de área

Se comienza con la presentación de algunas herramientas que pueden ser útiles en la exploración inicial de datos de unidades de área. Un concepto primario en este tipo de datos es el de matriz de proximidades, W (ver [20]). Dadas las medidas $Y_1, Y_2, \cdots, Y_n$ asociadas con las unidades de área $1, 2, \cdots . n$, las entradas w_{ij} en W conectan espacialmente las unidades i y j en algún sentido, habitualmente se tiene que $w_{ii} = 0$. Las posibilidades para definir dicha matriz incluyen por ejemplo una matriz binaria, es decir, con entradas 0 o 1, donde $w_{ij} = 1$ si i y j comparten la frontera, o comparten un vértice en el caso de que la región es dada en una retícula regular. Otra alternativa puede ser el uso de distancias entre las unidades(o sus centroides), por ejemplo una función decreciente de distancias entre los centroides de las unidades (como en un cantón u otra región del mapa). Sin embargo, aún con el uso de distancias se puede crear W como una matriz binaria, por ejemplo, para un i fijo, $w_{ij} = 1$ si j es uno de los K vecinos más cercanos en términos de distancia. Todas las elecciones anteriores sugieren que W podría ser simétrica, pero, para unidades de área irregulares, el último ejemplo nos da una forma donde no necesariamente es el caso. Los w_{ij} pueden ser estandarizados usando $\sum_j w_{ij} = w_{i+}$. Si $\widehat{W}$ tiene las entradas $\hat{w}_{ij} = w_{ij}/w_{i+}$, entonces $\widehat{W}$ es estocástico por filas, es decir, $\widehat{W}\mathbf{1} = \mathbf{1}$, pero en este caso $\widehat{W}$ no es necesariamente simétrica.

Como lo sugiere la notación, las entradas de la matriz W se pueden ver como pesos o ponderaciones. Mayor peso se asocia a los vecinos más cercanos de cada municipio, en algún sentido. En este contexto exploratorio W nos da un mecanismo para introducir una estructura espacial en el modelo formal.

Finalmente, al trabajar con distancias se pueden definir intervalos, es decir, $(0, d_1]$, $(d_1, d_2]$, $(d_2, d_3]$, etcétera. Lo anterior habilita la noción de vecindarios de orden 1 de la unidad i, i.e., to-

das las unidades (municipios) a una distancia menor o igual a d_1, vecinos de orden 2, i.e., todas las unidades a una distancia mayor a d_1 pero menor o igual a d_2, y así sucesivamente. De forma análoga a W se puede definir $W^{(1)}$ como una matriz de proximidad de los vecinos de primer orden. Esto es, $w_{ij}^{(1)} = 1$ si las unidades i y j son vecinos de primer orden y $w_{ij}^{(1)} = 0$ si no es el caso. Similarmente se puede definir $W^{(k)}$ como una matriz de proximidad de los vecinos de orden k.

Una herramienta muy importante para el análisis exploratorio de datos de área, según [7], es un mapa para los valores de los datos, ya que este nos puede mostrar si los datos exhiben una asociación espacial fuerte, o por el contrario, no existe tal asociación. Sin embargo, antes de dar conclusiones se deben estudiar todas las posibles covariables que puedan explicar la existencia de asociación espacial, en nuestro modelo la temperatura mínima por ejemplo.

Dos medidas de asociación espacial

Aunque el mapa con los datos es una herramienta muy útil para observar si la variable y las covariables presentan asociación espacial, existen varias medidas que pueden ayudar a dar conclusiones más justificadas para esta asociación. En particular dos de ellas se exponen en este apartado: índice Moran (I de Moran) e índice Geary (C de Geary).

El índice Moran es análogo al estadístico usado para medir asociación en series de tiempo, el coeficiente de autocorrelación. Para una matriz de asociación binaria, Patrick A.P Moran introdujo en 1950 el siguiente coeficiente de autocorrelación [37]:

$$I = \frac{n \sum \delta_{ij}(Y_i - \bar{Y})(Y_j - \bar{Y})}{\sum \delta_{ij} \sum_i (Y_i - \bar{Y})^2},$$

donde $Y_1, \cdots, Y_n$ son las observaciones sobre n regiones y $\delta_{ij} = 1$ si $i \neq j$ e i y j son contiguos, $\delta_{ij} = 0$ en otro caso. En 1973, Cliff y Ord [37] generalizaron esta definición reemplazando δ_{ij} por w_{ij}, siendo W una matriz de vecindarios. Es decir, el I de Moran toma la forma:

$$I = \frac{n \sum w_{ij}(Y_i - \bar{Y})(Y_j - \bar{Y})}{\sum w_{ij} \sum_i (Y_i - \bar{Y})^2}. \tag{1.32}$$

I no tiene soporte estrictamente en el intervalo $[-1, 1]$. El índice es el cociente de formas cuadráticas en $\mathbf{Y}$, el cual da la idea de obtener aproximados para el primer y segundo momentos [7]. Moran muestra bajo el modelo nulo donde los $Y_i's$ son independientes e idénticamente distribuidos, que I está asintóticamente distribuido como normal con media $-1/(n-1)$, y varianza (ver [20], sección 3.1):

$$Var[I] = \frac{n^2(n-1)S_1 - n(n-1)S_2 - 2S_0^2}{(n+1)(n-1)S_0^2}$$

En la ecuación de dicha varianza, se tiene que:

$$\begin{aligned} S_0 &= \sum_{i \neq j} w_{ij}, \\ S_1 &= \frac{1}{2} \sum_{i \neq j} (w_{ij} - w_{ji})^2, \\ S_2 &= \sum_k \left(\sum_j w_{kj} + \sum_i w_{ki} \right)^2. \end{aligned}$$

Sin embargo, dichas fórmulas no son válidas para los residuos de una regresión [37], en [20] el I de Moran se recomienda usar como una medida exploratoria de asociación espacial, en lugar de usarse como una prueba de significancia espacial.

El C de Geary es análogo al estadístico Durbin-Watson usado en series de tiempo para medir asociación. Robert C. Geary propuso en 1954 el índice con la forma [37]:

$$C = \frac{(n-1) \sum_i \sum_j \delta_{ij} (Y_i - Y_j)^2}{\left(\sum_{i \neq j} \delta_{ij} \right) \sum_i (Y_i - \bar{Y})^2},$$

sin embargo este índice se propuso para matrices de pesos binarias, y al igual que el I de Moran, este fue generalizado por Cliff y Ord [37] reemplazando δ_{ij} por w_{ij}, donde W es una matriz de pesos no necesariamente binaria. Con esto, el C de Geary quedó con la forma:

$$C = \frac{(n-1) \sum_i \sum_j w_{ij} (Y_i - Y_j)^2}{\left(\sum_{i \neq j} w_{ij} \right) \sum_i (Y_i - \bar{Y})^2}. \tag{1.33}$$

Este índice es positivo, y tiene media 1 para el modelo nulo; valores entre 0 y 1 indican asociación espacial positiva. Como el índice de Moran, C es un cociente de formas cuadráticas en $\mathbf{Y}$ y, como el índice I, es asintóticamente normal si los Y_i son independientes e idénticamente distribuidos (detalles en [37]).

1.4.2. Modelo Condicional Autorregresivo (CAR)

Los modelos estadísticos usados para generar estimadores estables de tasas de incidencia de enfermedades deberían ser capaces de introducir y acomodar todas las características de los datos, incluyendo el uso de distribuciones que no son normales para datos de conteo, los efectos de las variables explicativas o covariables y la correlación espacial. Existe un enfoque que involucra el uso de modelos lineales generalizados mixtos (GLMM, por sus siglas en inglés), en el cual el modelo lineal generalizado (GLM, por sus siglas en inglés) es aumentado con un efecto aleatorio que no es

observado y es distribuido normalmente, introducido con el objetivo de explicar la correlación espacial y la sobre dispersión. Al condicionar un vector de efectos aleatorios ϕ, el conteo de la incidencia observada y_i sigue un GLM log-lineal con media condicional dada por:

$$\ln(\mu_i) = \ln(E_i) + x_i'\alpha + \phi, \tag{1.34}$$

donde x_i es un vector de variables explicativas para la región i, α es el vector de coeficientes de regresión y E_i es el conteo de incidencias esperado.

Se particiona la región a estudiar en n unidades que no se traslapan, $\mathcal{S} = \{S_1, S_2, \cdots, S_n\}$, es decir, la única intersección posible entre ellas es la frontera, las cuales están vinculadas a un conjunto correspondiente de respuestas $\mathbf{Y} = \{Y_1, Y_2, \cdots, Y_n\}$ y un vector de "offsets" conocidos $\mathcal{O} = \{O_1, O_2, \cdots, O_n\}$. El patrón espacial en la variable de respuesta se modela con una matriz de covariables $\mathbf{X} = \{x_1, x_2, \cdots, x_n\}$ y un componente para la estructura espacial $\psi = \{\psi_1, \psi_2, \cdots, \psi_n\}$, este último se incluye para modelar cualquier autocorrelación espacial que quede remanente después de que el efecto de las covariables se tome en cuenta. El vector de covariables para la unidad S_k se denota por $x_k = \{1, x_{k1}, x_{k2}, \cdots, x_{kp}\}$, donde la primer entrada se refiere a un término para el intercepto. Un modelo lineal generalizado mixto, para los datos espaciales de área viene dado por [28]:

$$\begin{aligned} Y_k|\mu_k &\sim f(y_k|\mu_k, \nu^2), \quad \text{para } k \in \{1, 2, \cdots, n\}, \\ g(\mu_k) &= \mathrm{x}_k^T\beta + O_k + \psi_k, \\ \beta &\sim \mathrm{N}(\mu_\beta, \Sigma_\beta). \end{aligned} \tag{1.35}$$

Las variables de respuesta, Y_k, en general se supone que vienen de una familia exponencial de distribuciones, entre las cuales pueden ser binomial, gaussiana o poisson. El valor esperado de Y_k puede ser denotado por $E(Y_k) = \mu_k$, mientras que ν^2 es un parámetro de escala que se toma en cuenta si la distribución pertenece a familias de localización o si se modela con parámetros de sobredispersión; un ejemplo es la distribución gaussiana. El valor esperado de las respuestas se relacionan con el predictor lineal usando una función de enlace $g(.)$, la cual es la identidad en el caso gaussiano, la función logit en el caso de la distribución binomial y es el logaritmo natural en el caso de poisson. Con lo anterior, se ajustan los modelos de verosimilitud:

- **Binomial** - $Y_k \sim Bin(n_k, \theta_k)$ y $\ln(\theta_k/(1-\theta_k)) = \mathrm{x}_k^T\boldsymbol{\beta} + O_k + \psi_k$,
- **Gaussiano** - $Y_k \sim N(\mu_k, \nu^2)$ y $\mu_k = \mathrm{x}_k^T\beta + O_k + \psi_k$,

- **Poisson** $Y_k \sim Poisson(\mu_k)$ y $\ln(\mu_k) = \mathrm{x}_k^T\beta + O_k + \psi_k$.

El vector de los parámetros de regresión se denota por $\beta = (\beta_1, \cdots, \beta_p)$, y en el modelo se pueden incorporar efectos no lineales usando splines cúbicos naturales o funciones polinomiales básicas para los elementos de $\mathbf{X}$. En el paquete CARBayes [28] se asume una previa normal multivariada para β, la media μ_β y la matriz diagonal de varianzas Σ_β son elegidas bajo algún criterio o información previa que se tenga, aunque dicho paquete sugiere $\mu_\beta = 0$, los elementos en la diagonal de Σ_β todos iguales 1000.

El componente de estructura espacial ψ de los modelos antes escritos, contiene un conjunto de efectos aleatorios $\phi = (\phi_1, \cdots, \phi_n)$, el cual viene de un modelo condicional autoregresivo, que se define en la ecuación (1.36). Esos modelos son un caso especial de un Campo Aleatorio de Markov Gaussiano (GMRF, por sus siglas en inglés), y se puede escribir en la forma general $\phi \sim \mathrm{N}(0, \tau^2\mathbf{Q}^{-1})$, donde $\mathbf{Q}$ es una matriz de precisión que puede ser singular (en dicho caso el modelo se llama intrínseco). Dicha matriz controla la estructura de autocorrelación espacial de los efectos aleatorios, y se basa en una matriz no negativa simétrica de vecindarios o matriz de pesos $\mathbf{W}$.

De hecho, en [25], se define equivalentemente a ϕ de la siguiente manera: siempre teniendo presente que se trabaja sobre un campo aleatorio de markov, se definen las distribuciones condicionales completas,

$$\mathrm{p}(\phi_i|\phi_{-i}, \tau_i^{-1}) = \mathrm{N}\left(\alpha \sum_{i\sim j} b_{ij}\phi_j, \tau_i^{-1}\right), \quad i, j = 1, \cdots, n \tag{1.36}$$

donde $i \sim j$ denota el hecho de que j es vecino de i en algún sentido. Así, por el teorema de Hammersley-Clifford y el Lema de Brook (ver [20], sección 3.2), las distribuciones completas de 1.36 determinan de forma única la distribución conjunta:

$$\phi \sim \mathrm{N}(0, [\mathrm{D}_\tau(I - \alpha\mathrm{B})]^{-1}), \tag{1.37}$$

donde B es una matriz $n \times n$ tal que $b_{ii} = 0$, y $\mathrm{D}_\tau = \mathrm{Diag}(\tau_i)$; según [25], usualmente se asume que $\mathrm{D}_\tau = \tau\mathrm{D}$, donde D es una matriz diagonal $n \times n$. Además, α es un parámetro con el efecto de suavizar, y con frecuencia es interpretado como una medida de asociación espacial. Note por ejemplo, que $\alpha = 0$ genera un modelo independiente, sin embargo, es importante no ver a α como un parámetro de correlación. Es decir, el paráinetro α controla la dependencia espacial, y está entre 0 y 1, pero no se puede ver como un parámetro de correlación de la forma tradicional (ver [42]). En el modelo 1.37, se pueden tomar distintos valores para α, elegir distintas D y B y generar distintas estructuras para el modelo CAR.

El modelo Besag-York-Mollie (1991)

Este modelo, también conocido como convolución, se introdujo en Besag et al [2], siendo el primero de los CAR propuestos. Este contiene dos conjuntos de efectos aleatorios, autocorrelacionado espacialmente e independiente. Dicho modelo presenta la forma:

$$\begin{aligned} \psi_k &= \phi_k + \theta_k \\ \phi_k|\phi_{-k}, \mathbf{W}, \tau^2 &\sim \mathrm{N}\left(\frac{\sum_{i=1}^{n} w_{ki}\phi_i}{\sum_{i=1}^{n} w_{ki}}, \frac{\tau^2}{\sum_{i=1}^{n} w_{ki}}\right) \\ \theta_k &\sim \mathrm{N}(0, \sigma^2) \text{ i.i.d} \\ \tau^2, \sigma^2 &\sim \text{Gamma} - \text{Inversa}(a, b). \end{aligned} \tag{1.38}$$

El efecto aleatorio $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_n)$ consta de elementos independientes de media cero y varianza constante σ^2, mientras que la correlación espacial es modelada por ϕ. Para este último, la esperanza condicional es el promedio de los efectos aleatorios en áreas vecinas, mientras que la varianza condicional es inversamente proporcional al número de vecinos. Lo anterior parece apropiado, ya que si los efectos aleatorios tienen una autocorrelación espacial fuerte, entonces los más cercanos en un área tienen la mayor cantidad de información sobre el valor del efecto aleatorio de sus vecinos.

En el modelo 1.37, esto es equivalente a tomar $\alpha = 1$. Además, con frecuencia se toma $\mathbf{D} = \mathrm{Diag}(m_i)$, donde m_i es el número de vecindarios de la región i, y $\mathbf{B} = \mathbf{D}^{-1}\mathbf{W}$, siendo $\mathbf{W}$ una matriz de vecindarios definida por ceros y unos, 1 en caso de compartir frontera y cero en otro caso, además $w_{ii} = 0$ [25]. Con lo anterior, la ecuación 1.37 se transforma en:

$$\boldsymbol{\phi} \sim \mathrm{N}(0, [\tau(\mathbf{D} - \mathbf{W})]^{-1}) \tag{1.39}$$

Este modelo es simple y sencillo de ajustar, sin embargo, tiene dos grandes inconvenientes:

- La matriz $\tau(\mathbf{D} - \mathbf{W})$ es singular, es decir que dicha distribución es impropia,
- El modelo no contiene parámetros para controlar la fortaleza de la relación espacial entre las regiones.

El modelo Leroux-Lei-Breslow (1999)

El modelo anterior, aunque es sencillo de ajustar, tiene como una de sus limitaciones el no contener parámetros que controlen la fuerza de la asociación espacial, Leroux et al (1999), en [30] proponen un modelo para la dependencia espacial que incluye parámetros separados para la sobre dispersión y la fuerza de la dependencia espacial.

Besag et al [2] propusieron dos modelos para ϕ, el modelo intrínseco y el modelo independiente, pero dadas sus limitaciones, Leroux et al (1999) [30] proponen un nuevo modelo, donde

los modelos anteriores son casos particulares, el modelo se basa en la especificación de la inversa generalizada de la matriz de covarianza $\mathbf{D}$, de la siguiente manera:

$$\sigma^2\mathbf{D}^- = (1-\rho)\mathbf{I} + \rho\mathbf{W}, \tag{1.40}$$

donde $\mathbf{I}$ es la matriz identidad, $\rho \in (0,1)$ es un parámetro asociado a la correlación espacial, y $\mathbf{W}$ es una matriz que define la estructura de vecindarios de la forma:

$$r_{ij} = \begin{cases} n_i & \text{si } i = j \\ -I[i \sim j] & \text{si } i \neq j \end{cases}, \tag{1.41}$$

donde n_i es el número de vecinos que tiene la región $i-$ésima, en este caso $I[\cdot]$ es una función indicadora.

Como se mencionó en el párrafo anterior, esta especificación contiene el caso de independencia ($\mathbf{D} = \sigma^2\mathbf{I}$), obtenido al tomar $\rho = 0$, y también la autoregresión intrínseca, donde la matriz de covarianzas es dada por $\mathbf{D} = \sigma^2\mathbf{W}^-$, tomando $\rho = 1$.

Finalmente, en este caso los momentos condicionales están dados por:

$$\mathbf{E}(\phi_i|\phi_{-i}) = \frac{\rho}{1-\rho+\rho n_i}\sum_{j\sim i}\phi_j,$$

y

$$Var(\phi_i|\phi_{-i}) = \frac{\sigma^2}{1-\rho+\rho n_i}$$

El modelo Lee-Mitchell (2012)

Según Lee y Mitchell en 2012 [28], las previas CAR descritas con anterioridad obligan a un nivel de suavizado espacial igual en el conjunto de efectos aleatorios para todas las regiones. Lo anterior se ilustra, para el modelo de Leroux et al, por la estructura de correlación parcial del mismo, la cual para (ϕ_k, ϕ_j) está dada por:

$$Corr(\phi_k, \phi_j|\phi_{-kj}) = \frac{\rho w_{kj}}{\sqrt{(\rho\sum_{i=1}^{n} w_{ki} + 1 - \rho)((\rho\sum_{i=1}^{n} w_{ji} + 1 - \rho)}}, \tag{1.42}$$

donde la matriz $\mathbf{W}$ en este caso es la matriz de contigüidad mencionada en el modelo Besag-York-Mollié. Para regiones no vecinas los efectos aleatorios son condicionalmente independientes, mientras que para regiones que son vecinas su correlación parcial es controlada por ρ. Pero dicha representación de suavizamiento espacial puede llegar a ser muy simplista en la práctica, ya que el efecto

aleatorio para la superficie es probable que incluya sub-regiones de suavizamiento, así como límites o fronteras donde ocurran cambios abruptos. Lee y Mitchell en 2012 [29] proponen un método para capturar dicha estructura espacial localizada, incluyendo la identificación de límites en los efectos aleatorios para la superficie. La idea consiste en modelar los elementos de $\mathbf{W}$ que corresponden a regiones geográficamente adyacentes como cantidades aleatorias binarias, en lugar de asumir que se fijan en el valor 1. Por otro lado, si las regiones no comparten una frontera en común, entonces el elemento de la matriz $\mathbf{W}$ que las relaciona se fija en cero. En la ecuación (1.42) se puede ver que si w_{kj} es estimado en 1, entonces (ϕ_k, ϕ_j) son espacialmente correlacionados, y son suavizadas en el proceso del modelado. En contraste, si w_{kj} es estimado como 0, entonces no se da ningún suavizamiento entre (ϕ_k, ϕ_j), y son modelados como condicionalmente independientes.

Se denotan los datos usados para cuantificar el riesgo de alguna enfermedad por $y = (y_1, \cdots, y_n)$ y $\mathbf{E} = (E_1, \cdots, E_n)$, el primero siendo el número de casos de enfermedad observados, mientras que los segundos son los casos de enfermedad esperados que ocurran en cada región, los cuales dependen del tamaño y la estructura demográfica de la población viviendo ahí. La medida más simple de riesgo de enfermedad es la razón de incidencia estandarizado, el cual para cada región k está dado por $\widehat{R}_k = y_k/E_k$. Sin embargo, para estimar los efectos de covariables en el riesgo de alguna enfermedad, típicamente se usan modelos jerárquicos bayesianos. En general se definen:

$$\begin{aligned} Y_k|E_k, R_k &\sim \text{Poisson}(E_k R_k), \text{ para } k = 1, \cdots, n \\ \ln(R_k) &= x_k^T \beta + \phi_k \end{aligned} \tag{1.43}$$

El riesgo de enfermedad es denotado por R_k y es representado por las covariables $x_k^T = (x_{k1}, \cdots, x_{kp})$ y los efectos aleatorios $\phi = (\phi_1, \cdots, \phi_n)$, estos últimos permitiendo cualquier sobredispersión y correlación espacial luego de que los efectos covariados se tomaran en cuenta. Aunque existen varias previas CAR para ϕ en la literatura, en [29] se adoptó la previa propuesta por *Leroux y otros* en [30], la cual especifica las n distribuciones condicionales completas:

$$\phi_k|\boldsymbol{\phi}_{-k}, \mathbf{W}, \tau^2, \rho, \mu \sim \mathbf{N}\left(\frac{\rho \sum_{j=1}^n w_{kj}\phi_j + (1-\rho)\mu}{\rho \sum_{j=1}^n w_{kj} + 1 - \rho}, \frac{\tau^2}{\rho \sum_{j=1}^n w_{kj} + 1 - \rho}\right). \tag{1.44}$$

Así, el modelo propuesto en [29] para (1.43) con el fin de capturar la correlación espacial localizada y además identificar fronteras de riesgo, es modelar el conjunto $\{w_{kj}\}$ para áreas geográficamente adyacentes como ceros o unos. Lo anterior se lo proponen mediante el uso de un número reducido de parámetros de regresión $\alpha = (\alpha_1, \cdots, \alpha_q)$ y tratando a $\{w_{kj}\}$ como conjunto en lugar de como valores separados desconocidos [29].

El primer nivel del modelo está dado por:

$$\begin{aligned} Y_k|E_k,R_k &\sim \text{Poisson}(E_kR_k), \text{ para } k=1,\cdots,n \\ \ln(R_k) &= \phi_k \\ \phi_k|\phi_{-k},\mu,\alpha,\tau^2 &\sim \mathbf{N}\left(\frac{0{,}99\sum_{j=1}^{n} w_{kj}(\alpha)\phi_j + 0{,}01\mu}{0{,}99\sum_{j=1}^{n} w_{kj}(\alpha)+0{,}01}, \frac{\tau^2}{0{,}99\sum_{j=1}^{n} w_{kj}(\alpha)+0{,}01}\right). \end{aligned} \tag{1.45}$$

En el modelo no se incluyen covariables, de manera que la estructura espacial de la superficie de efectos aleatorios y la superficie de riesgos es la misma. Además ρ se fija en 0.99, de forma que la estructura de la correlación espacial puede ser determinada localmente por $\{w_{kj}(\alpha)\}$, en lugar de globalmente por ρ. Note que $\rho = 0$ no es elegible pues eso desaparecería los $\{w_{kj}(\alpha)\}$, mientras que un valor de 1 tampoco se elige debido a que esto podría resultar en media y varianza infinita si para alguna región k se cumple que $\sum_j w_{kj}(\alpha) = 0$.

La idea del modelo, es que las fronteras en la superficie de riesgo es probable que ocurran entre poblaciones que son muy distintas, ya que poblaciones homogéneas deberían tener perfiles de riesgo similares. Por lo tanto, se modela la presencia o ausencia de una frontera entre regiones adyacentes usando q métricas de disimilitud no negativas $z_{kj} = (z_{kj1}, \cdots, z_{kjq})$, donde $z_{kji} = |z_{ki} - z_{ji}|/\sigma_i$, $i = 1,\cdots,q$. Usando dicha métrica se propone un modelo de vecindarios dado por:

$$w_{kj}(\alpha) = \begin{cases} 1 & \text{si } exp\left(-\sum_{i=1}^{q} z_{kji}\alpha_i\right) \leq 0{,}5 \text{ y } k \sim j \\ 0 & \text{otro caso} \end{cases} \tag{1.46}$$

Los parámetros de la regresión, α, tienen la restricción de ser no negativos, de manera que cuanto mayor es la disimilitud entre dos zonas, es más probable que haya una frontera entre ellas. Además, no hay un término de intercepto en (1.46), de modo que dos regiones con poblaciones homogéneas no deben tener una frontera entre ellas. Para más detalles el lector puede ver [29].

1.4.3. Modelo Simultáneo Autorregresivo (SAR)

Los modelos simultáneos autoregresivos han sido usados extensamente en el análisis de datos espaciales en diversas áreas, es otra alternativa disponible para modelar los datos de riesgo relativo y todas las covariables, sin embargo, no será considerado en la tesis ya que el enfoque más común en la inferencia de estos modelos es con máxima verosimilitud, mientras que un enfoque bayesiano ha quedado atrás [15], además de lo extensivo que se puede volver el trabajo, por lo que sólo vemos su definición.

Asumimos el estudio del modelo:

$$\begin{aligned} \mathbf{Y} &= \mathbf{BY} + \phi, \text{ o, de forma equivalente,} \\ (\mathbf{I} - \mathbf{B})\mathbf{Y} &= \phi \end{aligned} \tag{1.47}$$

Se asume que ϕ induce una distribución sobre $\mathbf{Y}$ [20], imitando los modelos autoregresivos para las series de tiempo (ver [33], sección 5.4), se supone que los ϕ_i toman innovaciones independientes, es más, se asume que $\phi \sim \mathcal{N}\left(\mathbf{0}, \text{Diag}(\sigma_i^2)\right)$. Ahora, en [42] Wall escribe que $\mathbf{Y}_i = \sum_j b_{ij}\mathbf{Y}_j + \phi_i$, $i = 1 \cdots$, n, siendo n el número de elementos de una retícula, y donde $\phi_i \sim \mathcal{N}\left(0, \sigma_i^2\right)$. Por lo tanto, si $(I - B)$ es de rango completo,

$$\mathbf{Y} \sim \mathcal{N}\left(\mathbf{0}, (\mathbf{I} - \mathbf{B})^{-1} D((\mathbf{I} - \mathbf{B})^{-1})^t\right), \tag{1.48}$$

siendo $\mathbf{D} = \text{Diag}(\sigma_i^2)$, además se cumple que $Cov(\phi, \mathbf{Y}) = \mathbf{D}(\mathbf{I} - \mathbf{B})^{-1}$, lo que indica que la respuesta está correlacionada con ϕ y es precisamente por ello que se le llama simultáneo. Note además que si se toma $\mathbf{D} = \sigma^2 I$, la distribución en (1.48) se simplifica a $\mathbf{Y} \sim \mathcal{N}\left(\mathbf{0}, \sigma^2[(\mathbf{I} - \mathbf{B})(\mathbf{I} - \mathbf{B})^t]^{-1}\right)$; para que dicha distribución no sea impropia, se debe cumplir que $(\mathbf{I} - \mathbf{B})$ sea de rango completo.

Dos elecciones para la matriz $\mathbf{I} - \mathbf{B}$ son las que más se discuten en la literatura [42]. La primera de ellas asume que $\mathbf{B} = \rho\mathbf{W}$, donde $\mathbf{W}$ es una matriz de contigüidad, es decir, es una matriz de ceros y unos y tal que $w_{ii} = 0$. En este escenario, ρ es conocido como un parámetro de autocorrelación espacial y, $\mathbf{Y}_i = \sum_{j \sim i} \mathbf{Y}_j + \phi_i$, donde $j \sim i$ denota que el elemento j es vecino del elemento i en la retícula. De hecho, $\mathbf{W}$ puede ser cualquier matriz de proximidad, y $\mathbf{I} - \rho\mathbf{W}$ será no singular si $\rho \in \left(\frac{1}{\lambda_{(1)}}, \frac{1}{\lambda_{(n)}}\right)$, donde $\lambda_{(1)}, \cdots, \lambda_{(n)}$ son los valores propios ordenados de $\mathbf{W}$ [20].

De forma alternativa, $\mathbf{W}$ se puede reemplazar por una matriz $\widetilde{\mathbf{W}}$, obtenida al normalizar cada fila de $\mathbf{W}$, es decir, $\tilde{w}_{ij} = \frac{w_{ij}}{w_{i+}}$. Es claro que esta nueva matriz no es necesariamente simétrica, pero si se cumple que $\widetilde{\mathbf{W}}\mathbf{1} = \mathbf{1}$. Si se define $\mathbf{B} = \alpha\widetilde{\mathbf{W}}$, el valor de α será un parámetro de autocorrelación espacial, y como $\mathbf{W}$ es una matriz de contigüidad, se tiene que $\mathbf{Y}_i = \alpha \sum_{j \sim i} \frac{w_{ij}}{w_{i+}} + \phi_i$. Si se tuviera una retícula muy regular, de modo que los valores de w_{i+} sean esencialmente el mismo, entonces se tendrá que α es un múltiplo de ρ. Ahora, como $\widetilde{\mathbf{W}}$ es estocástica por filas, los valores propios de la misma serán menores o iguales a 1, por lo que $I - \alpha\widetilde{\mathbf{W}}$ será no singular siempre que $\alpha \in (-1, 1)$, lo que justifica el referirse a α como parámetro de autocorrelación [20].

Habitualmente, un modelo SAR se introduce en un contexto de regresión, por lo que los residuos, $U = \mathbf{Y} - \mathbf{X}\beta$, son los que se asumen que siguen dicho modelo, en lugar de $\mathbf{Y}$ mismo. Entonces,

siguiendo el modelo (1.47), si $U = BU + \phi$, se obtiene la forma:

$$\mathbf{Y} = B\mathbf{Y} + (I - B)\mathbf{X}\beta + \phi \tag{1.49}$$

Esta expresión muestra que $\mathbf{Y}$ es modelado por un componente que es un promedio espacial de los vecinos, para cada elemento en la partición, y un componente que es una regresión lineal simple (usual). Si B fuera la matriz nula, entonces se obtendría un modelo de regresión lineal simple para $\mathbf{Y}$, mientras que si $B = I$, se obtiene un modelo puramente espacial.

1.5. Cadenas de Markov de Monte Carlo

La simulación a través de Cadenas de Markov de Monte Carlo es un método general basado en la extracción de muestras de θ utilizando distribuciones posteriores aproximadas, donde θ representa el conjunto total de parámetros del modelo. Las muestras se extraen secuencialmente, a través de métodos iterativos, los cuales garantizan que el conjunto de muestras forme una cadena de Markov.

Según [9], un método de Cadenas de Markov de Monte Carlo para la simulación de una distribución f es cualquier método que produce una cadena de Markov ergódica $(X^{(t)})$, cuya distribución estacionaria es f.

Para referirse a este método de simulación, es necesario hablar un poco de Cadenas de Markov por un lado, y de integración de Monte Carlo por otro, lo cual se hará de manera sintetizada, ya que se sale de los objetivos de la tesis profundizar en dichos temas.

1.5.1. Simulación de Monte Carlo

Dos problemas numéricos importantes que tienen lugar en inferencia estadística son los relacionados con integración o con optimización, y generalmente el problema de optimización se asocia al problema de estimación por máxima verosimilitud, mientras que los de integración se relacionan al enfoque bayesiano, principalmente en la determinación de constantes de normalización (ver [9]).

Los métodos de Monte Carlo se basan en el muestreo de distribuciones de probabilidad. La generación de muestras aleatorias de una distribución uniforme, $U(0, 1)$, es sumamente importante, ya que todos los métodos de muestreo clásico dependen de un generador de números aleatorios que tengan este comportamiento. Actualmente la simulación de números aleatorios se basa en generadores más complejos, los cuales son proporcionados por programas de computadora especializados en estadística, por ejemplo R-CRAN, MATLAB, SAS, entre otros.

Uno de los principales problemas a resolver usando integración de Monte Carlo, es el de calcular, o aproximar la integral:

$$E_f\left[g(X)\right] = \int_{\chi} g(x)f(x)dx. \tag{1.50}$$

Basado en lo anterior, parece natural proponer el uso de una muestra, $(X_1, ..., X_m)$, generada por f para aproximar (1.50) a través del promedio empírico:

$$\bar{g}_m = \frac{1}{m}\sum_{i=1}^{m} g(X_i)$$

ya que $\bar{g}_m$ converge a $E_f[g(X)]$ casi seguramente debido a la ley de los grandes números [9]. Además, si se cumple que $g^2(X)$ tiene una esperanza finita bajo f, se puede evaluar la velocidad de convergencia de $\bar{g}_m$, pues la varianza:

$$\mathsf{var}(\bar{g}_m) = \frac{1}{m}\int_{\chi} [g(x) - E_f[g(x)]]^2 f(x)dx$$

también se puede aproximar por la muestra $(X_1, ..., X_m)$ por medio de la fórmula:

$$v_m = \frac{1}{m^2}\sum_{i=1}^{m}[g(X_i) - \bar{g}_m]^2.$$

Para m suficientemente grande, la variable aleatoria:

$$\frac{\bar{g}_m - E_f[g(X)]}{\sqrt{v_m}}$$

es distribuida aproximadamente como una variable aleatoria $\mathcal{N}(0,1)$, lo que permite la construcción de un test de convergencia y de intervalos de confianza para la aproximación $E_f[g(X)]$. Este enfoque se conoce con frecuencia como el *método de Monte Carlo.*

1.5.2. Cadenas de Markov

Sea $(S, \mathcal{S})$ un espacio medible, en [17] se define un Cadena de Markov con respecto a la filtración $\mathcal{F}_n$ como una sucesión de variables aleatorias, $\{X_n\}_{n\in\mathbb{N}}$, que toman valores en S, si $X_n \in \mathcal{F}_n$ y $\forall \mathrm{B} \in \mathcal{S}$ se cumple:

$$P(X_{n+1} \in \mathrm{B}|\mathcal{F}_n) = P(X_{n+1} \in \mathrm{B}|X_n). \tag{1.51}$$

Es decir, para predecir X_{n+1} dado X_n, es innecesario conocer $X_{n-1}, X_{n-2}, ..., X_0$. Por lo tanto se define la probabilidad de transición como:

Definición 1.6. Una función $p : S \times \mathcal{S} \to \mathbb{R}$ se llama una probabilidad de transición si:

(i) Para cada $x \in S$, $A \to p(x, A)$ es una medida de probabilidad sobre $(S, \mathcal{S})$.

(ii) Para cada $A \in \mathcal{S}$, $x \to p(x, A)$ es una función medible.

Entonces decimos que X_n es una cadena de Markov con respecto a $\mathcal{F}_n$ con probabilidades de transición p_n si:

$$P(X_{n+1} \in \mathrm{B}|\mathcal{F}_n) = p_n(X_n, \mathrm{B}).$$

Cuando S es discreto, las probabilidades de transición se pueden representar fácilmente mediante una matriz, también llamada matriz de transición, y las entradas de dicha matriz son dadas por:

$$P_{xy} = P(X_n = y|X_{n-1} = x), \quad x, y \in S.$$

Es importante notar que la distribución de X_0 (distribución inicial) juega un rol crucial, pues si tenemos la matriz de transición de la cadena de Markov; $\mathbf{P}$, y la distribución inicial $\mu_0 = (\omega_1, \omega_2, ...)$, entonces podemos obtener la distribución de probabilidad marginal de X_1 mediante el producto matricial $\mu_1 = \mu_0\mathbf{P}$, y haciendo repetidas multiplicaciones obtenemos la distribución de probabilidad marginal de X_n, es decir, $X_n \sim \mu_n = \mu_0\mathbf{P}^n$.

Por lo tanto [9], en el caso de una cadena de Markov, si la distribución inicial o el estado inicial son conocidos, la construcción de la cadena de Markov es completamente determinada por su probabilidad de transición, es decir, por la distribución de X_n condicionalmente sobre X_{n-1}.

1.6.1. Muestreo de Gibbs y el Algoritmo Metropolis-Hastings

El muestreo de Gibbs se ha convertido en uno de los métodos computacionales más populares de la inferencia bayesiana. Técnicamente, este puede ser visto como un método especial que supera el problema de la dimensionalidad por medio del condicionamiento. La idea básica es la misma que hay en los métodos de optimización condicional iterativos, suponga que se quiere generar números aleatorios desde una función de densidad $f(x)$, $x \in \mathcal{S} \subseteq \mathbb{R}^d$, se particiona el vector en K bloques, es decir, $x = (x_1, ..., x_K)^t$, con $K \leq d$ y

$$\dim(x_1) + \cdots + \dim(x_K) = d.$$

Denote por:

$$f_k(x_k|x_1, .., x_{k-1}, x_{k+1}, ..., x_K), \quad k = 1, ., K \tag{1.52}$$

el correspondiente conjunto de distribuciones condicionales completas. Bajo ciertas condiciones, este conjunto de condicionales, a su vez, determina la distribución destino (distribución posterior) $f(x)$, de acuerdo con el teorema de Hammersley-Clifford [8].

Teorema 1.6.1. ***(Hammersley-Clifford)*** *Si $f(x) > 0\,\forall x \in S$, entonces la distribución conjunta $f(x)$ es únicamente determinada por las condicionales completas (1.52). Más precisamente:*

$$f(x) = f(y) \prod_{k=1}^{K} \frac{f_{j_k}(x_{j_k}|x_{j_1}, ..., x_{j_{k-1}}, y_{j_{k+1}}, ..., y_{j_K})}{f_{j_k}(y_{j_k}|x_{j_1}, ..., x_{j_{k-1}}, y_{j_{k+1}}, ..., y_{j_K})} \quad (x \in S) \tag{1.53}$$

para toda permutación j en $(1, ..., d)$ y todo $(y \in S)$.

Algoritmicamente, el muestreo de Gibbs es iterativo. Iniciando con un punto arbitrario $x^{(0)} \in S$, con la restricción que $f(x^{(0)}) > 0$, cada iteración del muestreo de Gibbs genera un número aleatorio desde cada $f_k(x_k|x_1, ..., x_{k-1}, x_{k+1}, ..., x_K)$ cambiando $x_1, ..., x_{k-1}, x_{k+1}, ..., x_K$ a sus valores generados más recientes.

Una de las principales ventajas de este algoritmo es que todas las simulaciones son aceptadas, y en cada transición se obtiene un punto diferente de la cadena, lo anterior se debe a que la probabilidad de aceptación es 1 en todo momento. Además, las simulaciones sólo se realizan a través de las densidades condicionales completas, y el que estas sean densidades unidimensionales representan una gran ventaja computacional.

Sin embargo, para poder implementar el algoritmo es necesario que se pueda simular de una manera sencilla cada una de las densidades condicionales completas. En muchas situaciones esto no es posible tan siquiera para una de esas densidades, lo que limita la aplicación de este método. Dicho método se presenta el algoritmo 1.

Algoritmo 1 Muestreo de Gibbs.

Entrada: Una distribución previa: $f^{(0)}(x) > 0$.

Salida: $x^{(t)}$

1. Tome $x^{(0)} = (x_1^{(0)}, \cdots, x_K^{(0)})$ desde $f^{(0)}(x)$.
2. Genere $x_1^{(t)} \sim f_1(x_1|x_2^{(t-1)}, ..., x_K^{(t-1)})$.
3. Genere $x_k^{(t)} \sim f_k(x_k|x_1^{(t)}, ..., x_{k-1}^{(t)}, x_{k+1}^{(t-1)}, ..., x_K^{(t-1)})$.
4. Genere $x_K^{(t)} \sim f_K(x_K|x_1^{(t)}, ..., x_k^{(t)}, ..., x_{k-1}^{(t)})$

Bajo ciertas condiciones de regularidad la distribución de $x^{(t)} = (x_1^{(t)}, ..., x_K^{(t)})'$, denotada por $f^{(t)}(x)$, convergerá a $f(x)$ (ver [19]).

Aunque el muestreo de Gibbs es muy útil en muchos modelos estadísticos, dicho método no puede ser aplicado a una selección de problemas en donde se involucra espacios de múltiples parámetros con dimensionalidad distinta. Es decir, el muestreo de Gibbs es inconveniente para sacar muestras desde algunas distribuciones para las cuales las distribuciones condicionales de algunas o todas las componentes no son estándar. Para esos problemas, el Algoritmo de Metropolis-Hastings, el cual es una generalización del muestreo de Gibbs, es necesario [8].

El algoritmo de Metropolis-Hastings es ciertamente el método McMC por excelencia, fue nombrado en honor a los trabajos realizados por Nicholas Metropolis (1953) y W.Keith Hastings (1970). Sea F una distribución de probabilidad, con función de densidad f, la idea básica de crear una cadena de Markov con núcleo de transición K es tener a F como su distribución invariante, de modo que [8]:

$$F(dy) = \int_{\mathcal{X}} F(dx)K(x, dy),$$

donde $\mathcal{X}$ es el espacio de muestras. El algoritmo parte de la función de distribución F, conocida como función objetivo o de interés. Luego se elige una densidad condicional, $q(y|x)$, definida con respecto a la medida dominante del modelo, en la práctica se busca una densidad q de la cual es sencillo realizar simulación mientras la densidad f debe estar disponible en algún sentido, por ejemplo, el cociente:

$$f(y)/q(y|x),$$

es una constante independiente de x, o constante en términos de x. Así, el algoritmo de Metropolis-Hastings asociado con la densidad f y la densidad condicional q produce una cadena de Markov por medio del algoritmo 2.

Algoritmo 2 Metropolis-Hastings.

Entrada: $x^{(t)}$.

Salida: $X^{(t)}$

1. Genere: $Y_t \sim q(y|x^{(t)})$.
2. Defina:

$$X^{(t+1)} = \begin{cases} Y_t \text{ con probabilidad } \rho(x^{(t)}, Y_t) \\ x^{(t)} \text{ con probabilidad } 1 - \rho(x^{(t)}, Y_t) \end{cases}$$

donde $\rho(x, y) = \min\left(\frac{f(y)}{f(x)}\frac{q(x|y)}{q(y|x)}, 1\right)$

La probabilidad de transición de la cadena Metropolis-Hastings está dado por [9]:

$$K(x, y) = \rho(x, y)q(y|x) + \delta_x(y)(1 - r(x)), \tag{1.54}$$

donde $r(x) = \int \rho(x,y)q(y|x)dy$, y $\delta_x(y)$ es la función delta de Dirac.

La distribución q es llamada la *distribución instrumental*, y la probabilidad $\rho(x,y)$ es conocida como la probabilidad de aceptación Metropolis-Hastings. Dicha probabilidad debe distinguirse de la *tasa de aceptación*, la cual es el promedio de la probabilidad de aceptación sobre las iteraciones,

$$\bar{\rho} = lim_{T\to\infty}\frac{1}{T}\sum_{t=0}^{T}\rho(X^{(t)},Y_t) = \int \rho(x,y)f(x)q(y|x)dydx. \tag{1.55}$$

Esta cantidad permite una evaluación del rendimiento del algoritmo. Por otro lado, el algoritmo 2 satisface la *condición de balance detallado*,

$$f(x)K(y|x) = f(y)K(x|y), \tag{1.56}$$

desde la cual se puede deducir que f es la distribución estacionaria de la cadena $\{X^{(t)}\}$, integrando a ambos lados de la ecuación sobre x (ver [9]).

Según [9], el algoritmo 2 siempre acepta valores y_t tales que el cociente $f(y_t)/q(y_t,x^{(t)})$ es creciente, comparado con el valor previo $f(x^{(t)})/q(x^{(t)},y_t)$. Una característica importante del algoritmo 2 es que éste puede aceptar valores y_t tales que el cociente es decreciente, algo similar a algunos métodos estocásticos de optimización (para más información al respecto se puede referir a la sección 5.3 de [9]). El algoritmo de Metropolis-Hastings depende sólo de los cocientes:

$$f(y_t)/f(x^{(t)}) \text{ y } q(x^{(t)}|y_t)/q(y_t|x^{(t)})$$

y es, por lo tanto, independiente de constantes de normalización, asumiendo, nuevamente que $q(\cdot|x)$ es conocida excepto por una constante que es independiente de x.

Es obvio que la probabilidad $\rho(x^{(t)},y_t)$ está definida solo cuando $f(x^{(t)}) > 0$. Sin embargo, si la cadena comienza con un valor $x^{(0)}$ tal que $f(x^{(0)}) > 0$, se sigue que $f(x^{(t)}) > 0\ \forall t \in \mathbb{N}$ ya que los valores de y_t tal que $f(y_t) = 0$ hacen que $\rho(x^{(t)},y_t) = 0$, y son por lo tanto rechazados por el algoritmo. En lo anterior se usa la convención de que $\rho(x,y) = 0$ siempre que $f(x) = f(y) = 0$, con el fin de eliminar dificultades teóricas [9].

1.6.2. Convergencia del Algoritmo Metropolis-Hastings

En este apartado se establecen las condiciones suficientes bajo las cuales f es la distribución estacionaria del algoritmo, y además la cadena generada es ergódica, para la prueba de los teoremas y más detalles en este tópico el lector puede revisar Robert y Casella (2004) [9].

Definición 1.7. Una cadena de Markov estacionaria $\{X^{(t)}\}$ es reversible si la distribución de $X^{(t+1)}|X^{(t+2)}$ es la misma que la distribución de $X^{(t+1)}|X^{(t)}$.

Teorema 1.7.1. *Sea $\{X^{(t)}\}$ una cadena de Markov con probabilidad de transición K. Suponga que dicha probabilidad satisface la condición de balance detallado (1.56) con una función de densidad de probabilidad f. Entonces se cumple:*

(a) La probabilidad de transición de $\{X^{(t)}\}$ satisface la condición de balance detallado con f,

(b) f es una distribución estacionaria de la cadena $\{X^{(t)}\}$.

El teorema anterior es necesario para poder probar nuestro primer resultado de interés, el cual afirma que la distribución f es estacionaria.

Teorema 1.7.2. *Sea $\{X^{(t)}\}$ una cadena de Markov generada por el algoritmo de Metropolis-Hastings. Entonces se cumple que para cualquier distribución condicional q, que satisface que $supp(f) \subseteq supp(q)$, se cumple:*

(a) La función f es una densidad invariante de la cadena,

(b) La cadena $\{X^{(t)}\}$ es reversible.

Para garantizar que la distribución límite es f y que la cadena sea convergente sin importar cual sea el punto inicial, se debe establecer la ergodicidad. Para lograr lo anterior, es necesario que la cadena sea aperiódica y Harris recurrente, dos conceptos que se definen a continuación:

Definición 1.8. Sea ψ una medida, la cadena de Markov $\{X^{(t)}\}$ con probabilidad de transición $K(x, y)$ es $\psi-$irreducible si, para todo $A \in \mathcal{S}$ con $\psi(A) > 0$, existe un natural n tal que $\forall x \in S,\ K^n(x, A) > 0$.

Definición 1.9. Un conjunto A es Harris recurrente si $\forall x \in A,\ P_x(n_A = \infty) = 1$. Una cadena $\{X^{(t)}\}$ es Harris recurrente si $\exists\ \psi$, una medida tal que la cadena es $\psi-$irreducible y para cada conjunto A para el cual $\psi(A) > 0$, A es Harris recurrente.

En la definición anterior se usó el término n_A, el cual se define como el número de pasadas de $\{X^{(t)}\}$ por A, es decir:

$$n_A = \sum_{t=1}^{\infty} \mathbb{I}_A\left(X^{(t)}\right).$$

Para la aperiodicidad de la cadena es suficiente que se cumpla que $K(x,x) > 0$ para todo $x \in S$, siendo K la probabilidad de transición, sin embargo esto se logra haciendo que se cumpla que:

$$P\left[f(x)q(y|x) \leq f(y)q(x|y)\right] < 1.$$

Ahora, para que la cadena sea Harris recurrente, basta el siguiente teorema:

Teorema 1.9.1. *Si la cadena de Metropolis-Hastings $\{X^{(t)}\}$ es f-irreducible, entonces es Harris recurrente.*

1.10. Comparación y evaluación de modelos

1.10.1. Criterio de información de la devianza (DIC)

El DIC surge como una generalización de los criterios de información de Akaike (AIC) y el bayesiano (BIC), fue propuesto por Spiegelhalter et al en 2002 [38]. Es de gran utilidad en la selección de modelos bayesianos donde la distribución posterior se obtiene a partir de cadenas de Markov con Monte Carlo (MCMC). El ajuste de un modelo se puede medir a través del error cuadrático medio ponderado:

$$T(y,\theta) = \frac{1}{n}\sum_{i=1}^{n}\frac{(y_i - E(y_i|\theta))^2}{Var(y_i|\theta)},$$

donde y representa los datos, y θ es el parámetro en el modelo. Una opción más general, es usar la devianza, la cual se define como -2 veces el logaritmo de la verosimilitud,

$$D(y,\theta) = -2\log(p(y|\theta)), \tag{1.57}$$

y es proporcional al error cuadrático medio si los datos tienen una distribución normal de varianza constante [21]. En el límite, para muestras de gran tamaño, el modelo con la devianza esperada más baja, tendrá una probabilidad posterior más alta. Por lo tanto, parece razonable estimar la devianza esperada como una medida de ajuste del modelo global; la discrepancia entre los datos y el modelo depende en general de θ así como de y. Para obtener un resumen que dependa sólo de y, se puede definir:

$$D_{\widehat{\theta}} = D(y,\widehat{\theta}(y)), \tag{1.58}$$

donde se usa un estimador puntual para θ, tal como la media de las simulaciones posteriores. Desde un punto de vista bayesiano, parece ser más atractivo promediar la discrepancia misma sobre la distribución posterior:

$$D_{ave}(y) = E[D(y,\theta)|\theta], \tag{1.59}$$

la cual podría estimarse usando simulaciones posteriores θ_l:

$$\widehat{D}_{ave}(y) = \frac{1}{L}\sum_{l=1}^{L} D(y,\theta^l). \tag{1.60}$$

La diferencia entre la devianza media posterior en (1.60) y la devianza puntual (1.58),

$$p_D = \widehat{D}_{ave}(y) - D_{\widehat{\theta}}, \tag{1.61}$$

representa el efecto del ajuste del modelo y es usado como una medida del número efectivo de parámetros del modelo bayesiano. En (1.61), p_D representa el decremento en la devianza esperada debido a la estimación de los parámetros en el modelo. Otro enfoque relacionado, hace la estimación del error que podría esperarse al aplicar el modelo ajustado a datos futuros, por ejemplo el error cuadrático medio en la predicción:

$$D_{ave}^{pred}(y) = E[D(y^{rep},\widehat{\theta}(y))], \tag{1.62}$$

donde $D(y^{rep},\theta) = -2\log(p(y^{rep}(\theta)))$, y y^{rep} se define como una réplica independiente de los datos, o de forma predictiva, como los datos que podríamos ver mañana si el experimento que produjo y hoy se repitiera con el mismo modelo y el mismo valor de θ que produjo los datos observados [7]. La esperanza se calcula sobre la distribución de y^{rep} y $\widehat{\theta}$ es un parámetro estimado. En general, la devianza esperada en (1.62) será más alta que la devianza esperada $\widehat{D}_{ave}$, ya que los datos que se predijeron están siendo comparados con un modelo obtenido a partir de los datos. La devianza esperada en (1.62) ha sido sugerida como un criterio de ajuste de modelos, cuando el objetivo es obtener una mejor potencia al hacer predicciones fuera de la muestra, este valor puede ser aproximado por una expresión conocida como criterio de información de la devianza (DIC, por su nombre en inglés):

$$\begin{aligned} DIC &= \widehat{D}_{ave}^{pred}(y) \\ &= 2\widehat{D}_{ave}(y) - D_{\widehat{\theta}} \\ &= 2p_D + D_{\widehat{\theta}} \end{aligned}$$

El lector más interesado en el tema, puede revisar [38], [21].

1.10.2. Criterio de información de Watanabe-Akaike (WAIC)

Este criterio de información tiene un enfoque completamente bayesiano, es usado para estimar la esperanza del logaritmo de la densidad de predicción puntual:

$$elppd = \sum_{i=1}^{n} E_f[\log(p_{pos}(\hat{y}_i))], \tag{1.63}$$

donde p_{pos} se refiere a la distribución posterior, iniciando desde el logaritmo de densidad puntual calculado:

$$clppd = \sum_{i=1}^{n} \log\left(\frac{1}{S}\sum_{s=1}^{S} p(y_i|\theta^s)\right), \tag{1.64}$$

para después añadir una corrección para el número efectivo de parámetros para ajustar debido al sobre-ajuste. En la literatura se proponen 2 ajustes, ambos basados en cálculos puntuales que se pueden ver como aproximaciones a la validación cruzada [44], en el presente trabajo sólo se dará uno de dichos ajustes.

Esta medida usa la varianza de términos individuales en el logaritmo de la densidad predictiva, sumado sobre n datos puntuales:

$$p_{waic} = \sum_{i=1}^{n} Var_{post}(\log p(y_i|\theta)). \tag{1.65}$$

Para obtener (1.65), se calcula la varianza posterior del logaritmo de la densidad predictiva para cada dato puntual y_i, es decir, $V_{s=1}^{S} \log p(y_i|\theta^s)$, donde $V_{s=1}^{S}$ representa la varianza muestral, esto es,

$$V_{s=1}^{S} a_s = \frac{1}{m-1}\sum_{s=1}^{S}(a_j - \bar{a})^2.$$

Sumando sobre todos los datos puntuales y_i se obtiene el número efectivo de parámetros del modelo:

$$cp_{waic} = \sum_{i=1}^{n} V_{s=1}^{S}(\log p(y_i|\theta^s)). \tag{1.66}$$

Después se usa cp_{waic} como una corrección de sesgo para el logaritmo de la densidad predictiva puntual, calculada como:

$$lppd = \sum_{i=1}^{n} \log\left(\frac{1}{S}\sum_{s=1}^{S} p(y_i|\theta^s)\right), \tag{1.67}$$

es decir, se obtiene la medida $\widehat{elppd}_{waic}$ sin sesgo definida por

$$\widehat{elppd}_{waic} = lppd - p_{waic}, \tag{1.68}$$

asumiendo que el número de simulaciones extraídas es suficientemente grande para capturar completamente la distribución posterior.

Así, se define el criterio de información de Watanabe-Akaike (o Criterio de Información Ampliamente Aplicable), como -2 veces el valor de $\widehat{elppd}_{waic}$, por lo tanto:

$$WAIC = \sum_{i=1}^{n} \log E_{\theta|y}(p(y_i|\theta)) + \sum_{i=1}^{n} Var_{\theta|y}(\log p(y_i|\theta)) \tag{1.69}$$

donde el primer elemento en la suma es visto como medida de ajuste del modelo, mientras que el segundo término es una medida de complejidad. En comparación de modelos, se prefieren aquellos que tengan un WAIC menor, para más detalles se puede revisar [21], [43] y [44]. A diferencia del DIC, el WAIC tiene la propiedad deseable de promediar sobre la distribución posterior, en lugar de condicionarlo a un estimador puntual; además el WAIC funciona incluso con modelos singulares, lo que resulta útil en modelos con estructuras jerárquicas y mixtas en los cuales el número de parámetros aumenta con el tamaño de la muestra y donde los estimadores puntuales con frecuencia no tienen sentido. Por tales motivos en el presente trabajo preferimos como medida de comparación de modelos el WAIC.

1.10.3. Diagnóstico de convergencia de Geweke

Este diagnóstico es utilizado como un método para estimar el número de iteraciones que serán quemadas en la simulación, además de proveer información concluyente en cuanto a la convergencia del algoritmo a una distribución estacionaria. El diagnóstico compara la ubicación del parámetro muestreado en dos momentos diferentes de la cadena; si los valores medios de los parámetros en dos momentos distintos son muy cercanos entre si, podemos asumir que las dos partes de la cadena tienen ubicaciones similares en el espacio de estados del parámetro, y se asume que las dos sub-muestras vienen de la misma distribución. Usualmente se compara la última mitad de la cadena, la cual se asume que ha convergido, con algún intervalo más pequeño del inicio de la cadena, este diagnóstico usa la estimación de la densidad espectral para dicho análisis.

Ahora se procede a describir el diagnóstico. Sea $\Theta(X)$ un funcional y denote por $\theta^t = \Theta(X^{t+n_0})$, donde n_0+1 es la iteración inicial desde donde se quiere probar si la cadena ha tenido convergencia. Si se definen $A = \{t : 1 \leq t \leq n_A\}$ y $B = \{t : n^* \leq t \leq n\}$, donde $1 < n_A < n^* < n$ y $n_A + n_B < n$, sean:

$$\bar{\theta}_A = \frac{1}{n_A} \sum_{t \in A} \theta^t, \qquad \bar{\theta}_B = \frac{1}{n - n^* + 1} \sum_{t \in B} \theta^t.$$

Note que $\bar{\theta}_A$ y $\bar{\theta}_B$ son las medias del parámetro del modelo en dos momentos distintos que no traslapan. Además, si el proceso MCMC y el funcional Θ implican la existencia de una densidad espectral $\widehat{S}_\theta(0)$ para esta serie de tiempo sin discontinuidades en 0, entonces $\widehat{S}_\theta^A(0)/n_A$ y $\widehat{S}_\theta^B(0)/(n-n^*+1)$ son las varianzas asintóticas de dichas medias. Por lo tanto, la raíz cuadrada de dichas varianzas asintóticas estiman el error estándar de $\bar{\theta}_A$ y $\bar{\theta}_B$.

El diagnóstico sugiere que si la cadena ha convergido en n_0, es decir, ambas sub-muestras se extrajeron de la distribución estacionaria de la cadena, entonces las dos medias $\bar{\theta}_A$ y $\bar{\theta}_B$ deberían ser iguales, y el estadístico de Geweke es asintóticamente normal estándar,

$$Z_n = \frac{\bar{\theta}_A - \bar{\theta}_B}{\sqrt{\frac{1}{n_A}\widehat{S}_\theta^A(0) + \frac{1}{n-n^*+1}\widehat{S}_\theta^B(0)}} \longrightarrow \mathcal{N}(0,1), \quad \text{si } n \to \infty.$$

Dicho resultado nos permite probar la hipótesis nula de igualdad en la ubicación en el espacio de estados de θ. La hipótesis nula se rechaza si $|Z_n|$ es grande, y eso indicaría que la cadena aún no ha convergido en n_0.

En el presente trabajo se hace el diagnóstico de Geweke usando la configuración predeterminada del paquete coda [34], de R-CRAN [35]. Es decir, las ventanas son $A = 0{,}1n$ y $B = 0{,}5n$, que son los valores sugeridos por Geweke en 1992 [22]. Primero se aplica el estadístico a la cadena completa, si el estadístico Z está fuera del intervalo de confianza de 95 %, se continúa aplicando el diagnóstico después de quemar (burn-in) el 10 %, 20 %, 30 % y 40 %, si se mantiene el comportamientom se reporta no convergencia en la cadena. Para más detalles el lector puede revisar ([22], [27]).

Capítulo 2

Sobre los Datos

2.1. Origen de los datos

El modelo que se construirá incorpora el riesgo relativo de contraer la infección del dengue por municipio de Puerto Rico, este riesgo se define como:

$$RR_i = \left(\frac{c_i}{P_i}\right) / \left(\frac{c_T}{P_T}\right) \tag{2.1}$$

donde:

- 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 (2.1) que se presenta a continuación:

Figura 2.1: Mapa de colores para riesgo relativo en semana 1.

Una descripción de la figura 2.1 se hará en la sección 3.1.

- 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 [10],

- 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 [6],
- 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 covariables a utilizar 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 modelo se usará como el dato para toda el área del municipio. Como es natural, los municipios costeros son los de altitud mas baja, mientras que el centro de Puerto Rico tiene las regiones mas altas, dado que ahí se encuentra la mayor región montañosa de este país, sin embargo, la máxima altitud obtenida es de 602.8 m.s.n.m., para el municipio de Aibonito, esto nos permite decir que el país en promedio es plano, ya que no posee lugares de gran altitud. En la figura (2.2) se presentan los datos,

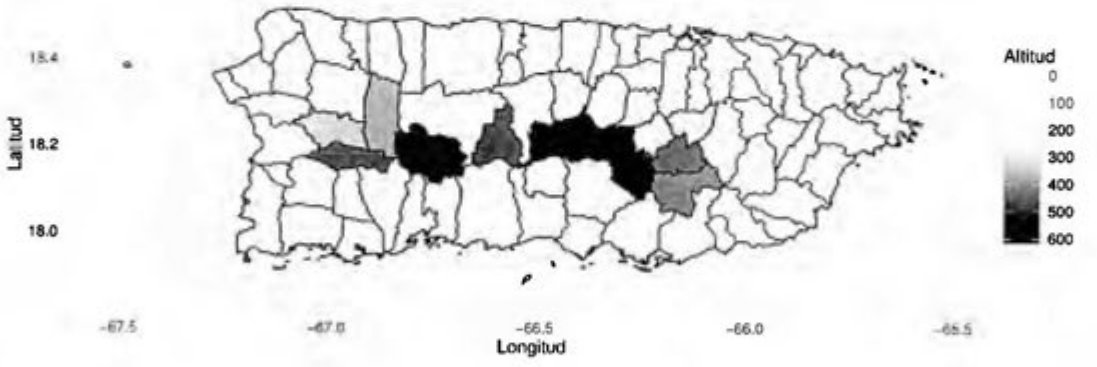


Figura 2.2: Altitud usada para cada municipio de Puerto Rico.

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 Centers for Environmental Information (NCEI)**, el cual pertenece al National Oceanic and Atmospheric Administration de los Estados Unidos de América [32].

3. Porcentaje de pobreza: este porcentaje fué obtenido desde **U.S Census Bureau**, de los Estados Unidos, desde el censo tomado para el año 2010. En la figura (2.3) se presenta un mapa de colores para esta covariable:

Figura 2.3: Porcentaje de pobreza según Censo de 2010.

Se puede observar en la figura 2.3 que las regiones con mayor índice de pobreza se sitúan en el centro del país, tal es el caso de los municipios de Orocovis y Morovis que son vecinos en la región central, además al noreste se encuentra el municipio de Quebradillas, al este centro Maricao, un municipio bastante diferente a sus vecinos en términos de pobreza. Las regiones con mayor desarrollo se sitúan en su mayoría cerca de San Juan, tal es el caso de Guaynabo, Carolina, Trijullo Alto Bayamón, más al sur Gurabo y al este en la región costera los vecinos de Dorado y Toa Baja.

4 Índice de vegetación mejorado (EVI, siglas para Enhanced vegetation index): los datos de esta covariable fueron extraídos utilizando el paquete MODISTools [40] desarrollado para el sofware R-CRAN, el cual extrae los mismos usando internet desde los archivos de **NASA LPDAAC** (Land Processes Distributed Active Archive Center of NASA).

2.2. Indice de Vegetación Mejorado

Este índice obtiene la respuesta de las variaciones estructurales del dosel vegetal, incluyendo el índice de área foliar, tipo y arquitectura del dosel, 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

[24]. Dicho índice se calcula como:

$$EVI = G \cdot \frac{IRp - R}{IRp + C_1 R - C_2 A + L} \tag{2.2}$$

donde:

- A, R, IRp: representan la reflectividad en la banda azul, roja e infrarroja cercana, respectivamente,
- $C_1 = 6$: es un coeficiente de resistencia atmosférica,
- $C_2 = 7,5$: es un coeficiente de resistencia atmosférica,
- $G = 2,5$: es un factor de ganancia,
- $L = 1$: es un factor de corrección.

A diferencia de los datos de temperatura y precipitación, los valores del índice de vegetación se pueden obtener para cada centro de población, sin embargo, los datos de EVI no están disponibles para las fechas que en que los demás datos fueron recolectados, por lo que se interpola la serie temporal que se obtiene con el paquete MODISTools de R-CRAN. Este paquete tiene una función encargada de extraer de forma remota y descargar localmente series temporales de productos MODIS (siglas en inglés para Moderate-Resolution Imaging Spectroradiometer), MODIS es un instrumento científico puesto en órbita terrestre lanzado por la NASA en 1999 a bordo del satélite Terra, en 2002 a bordo del satélite Aqua. Dicha función, a partir de coordenadas, fechas y el tamaño de la retícula descarga los datos, en este caso, el producto obtenido fue el índice de vegetación de cada 16 días y una banda de 250 metros, se decidió descargar los datos para un área de 4 kilómetros cuadrados, siendo el centro de población de cada municipio, también el centro de cada cuadrado 2×2, al hacerlo de este modo se obtienen 81 datos por centro de población, y el valor usado en el modelo será la interpolación temporal de los promedios de estos 81 valores. La serie obtenida para el Municipio de Adjuntas por ejemplo, se muestra en la figura (2.4).

La interpolación se realizó mediante el uso del paquete Zoo ([46]), también de R-CRAN, a través de splines cúbicos mediante la función `na.spline()`. El objetivo principal de esto es obtener el valor del EVI por municipio para cada semana, y utilizar dicho dato en nuestro modelo CAR.

Una vez que se hace la interpolación, se quiere estudiar el comportamiento espacial del EVI para cada semana, en la primer semana por ejemplo se tiene un valor mínimo de 2362.8, obtenido en

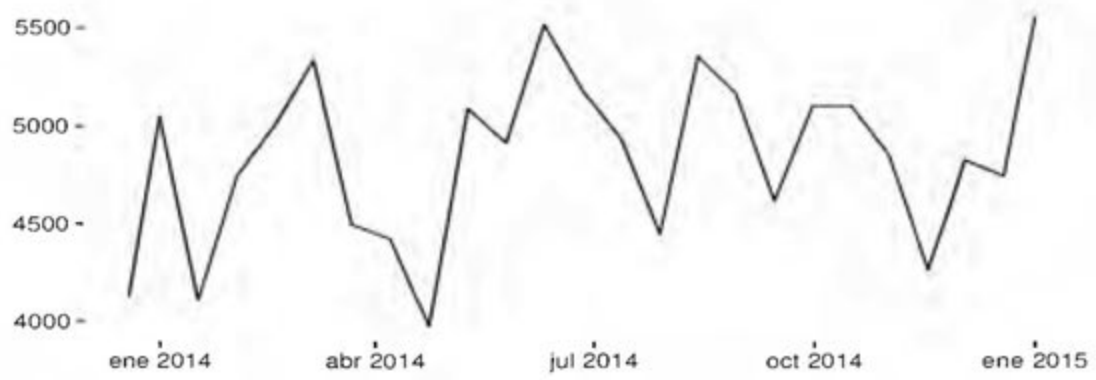


Figura 2.4: Serie del EVI en 2014 para Adjuntas.

el Municipio de Caguas, mientras que el máximo para dicha semana es de 6568.1 en Toa Baja, el promedio y la mediana son de 4144 y 4149 respectivamente, además una desviación estándar de 761.5. En la figura (2.5) podemos observar el histograma para las primeras tres semanas del año 2014. En las tres semanas se puede ver que la mayoría de lo municipios tienen un índice entre los 3000 y los 5000, siendo mas de 50 los municipios con esta característica.

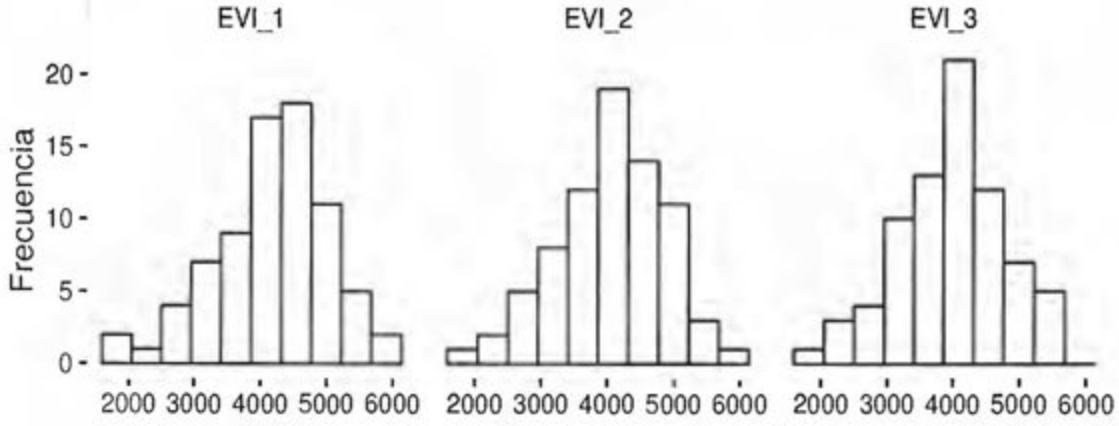


Figura 2.5: Histogramas para Índice de vegetación.

En los diagramas de cajas (boxplot), presentados en la figura (2.6), podemos ver el comportamiento estable de esta variable, donde se muestra además que para las semanas 1 y 3 existen posibles valores atípicos, ya que estos se alejan mucho del valor medio de las muestras.

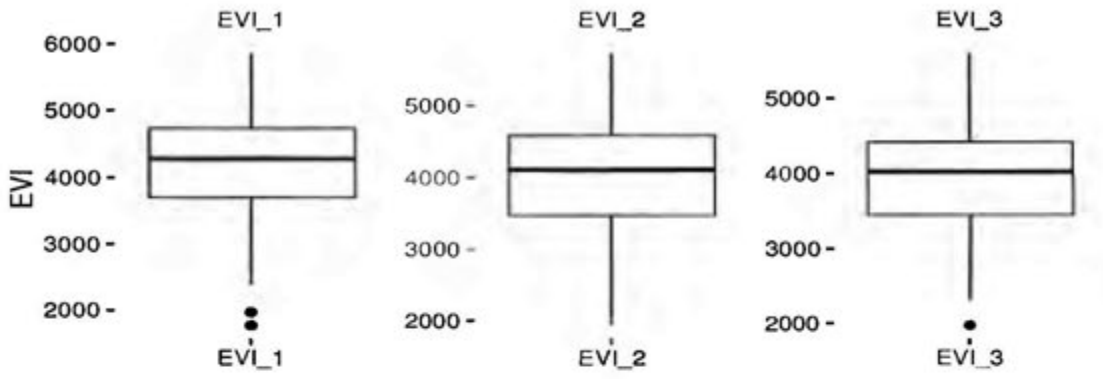


Figura 2.6: Diagramas de Caja para Índice de vegetación.

La información anterior, es mejor descrita mediante un mapa coroplético o de "temperaturas", para la primer semana se muestra dicho mapa en la figura 2.7. Podemos ver en dicho mapa que las regiones mas boscosas se sitúan en el centro del país, que es de hecho la región menos poblada y con mayor altitud del mismo, por otro lado, la región mas habitada de Puerto Rico es San Juan, la capital, y en los datos obtenidos para la primer semana del EVI se nota que es de las regiones con menos índice foliar, lo que parece consistente dada la naturaleza de dicha región.

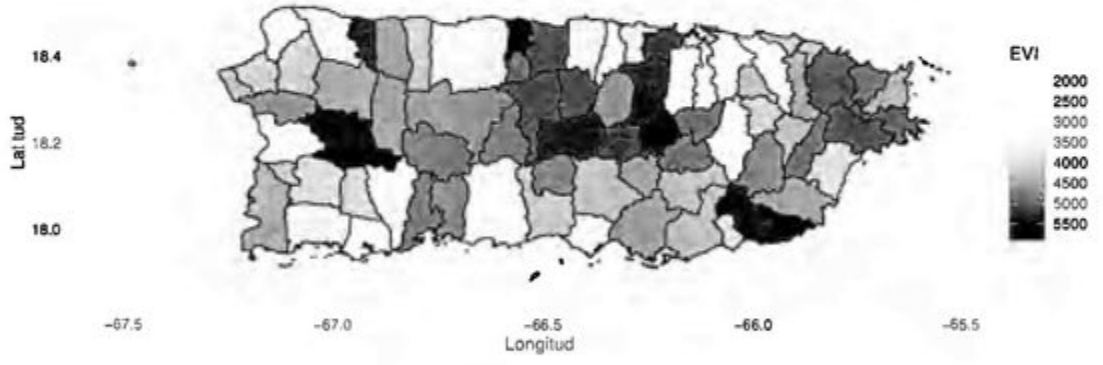


Figura 2.7: Mapa para el Índice de vegetación.

2.3. Interpolación espacial de datos climáticos

Los datos obtenidos para precipitación y temperaturas no se encuentran ubicados espacialmente en los centros de población de cada municipio de Puerto Rico, ya que las estaciones que toman dichas

medidas no necesariamente se sitúan de esta forma. Por lo anterior, se debió aplicar un método de interpolación espacial para estimar estas covariables para cada municipio. En la figura (2.8) se muestra la posición de las estaciones climatológicas y la posición de cada centro de población de los municipios de Puerto Rico, donde es claro que para obtener valores de precipitación, temperaturas u otra variable, es necesario algún tipo de interpolación espacial.

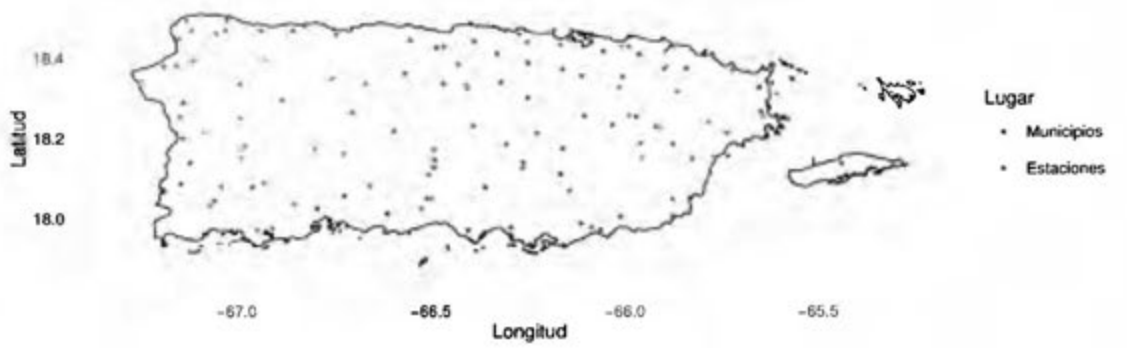


Figura 2.8: Ubicación espacial de las estaciones y los centros población.

Como se expresó en la sección (1.2.5), para realizar esta interpolación primero se debe hacer un ajuste de los parámetros para un variograma, en el presente trabajo se utilizará la función de covarianza exponencial, en las tres covariables, es decir:

$$C(h) = exp\left(\frac{|h|}{\phi}\right),$$

debido a que es una estructura de variograma más simple, y los datos disponibles para ajustar son relativamente pocos.

Es importante aclarar que los datos se obtienen de forma diaria por las estaciones, por lo tanto, en el presente trabajo lo que se utiliza como dato bruto es el promedio de precipitaciones o temperaturas que se midieron en las semanas 1 a 4 y 31 a 34 del año 2014, es decir, para cada estación se calcula el promedio de los datos medidos desde el día 1 al 7, y ese será nuestro dato para la primer semana, desde el día 8 al 14 los de la segunda semana y así sucesivamente. Por lo tanto, el análisis e interpolación que se realiza en este apartado, es a dichos promedios.

2.3.1. Análisis exploratorio de datos, caso no espacial

Los datos de precipitación están medidos en décimas de milímetros, según la documentación de la GHCN (Global Historical Climatology Network) [32]. Para la primera semana de 2014 van desde

0 y 79,2 décimas de mm, en la segunda semana el intervalo va desde 0 a 99.42 décimas de mm de lluvia, en tanto que en la tercera semana las precipitaciones oscilan entre 0 y 141 décimas de mm. El promedio en la precipitación de la semana 1 es de 20.17, la mediana es igual a 13.33 y los datos tienen una desviación estándar de 23.34. Para la segunda semana se tiene un promedio de 19.48, con una desviación estándar de 21.27 décimas de mm. Durante la tercera semana llovió en promedio 11.46 décimas de mm, y esta tuvo una desviación de 24.9. Según los datos, la primera semana fue mas lluviosa que la segunda y tercer semanas en promedio, sin embargo se puede ver que se registran lluvias mas grandes en las semanas 2 y 3. Un histograma para los datos de estas 3 semanas se presenta en la figura (2.9), podemos ver que dichos datos tienen una distribución truncada, ya que sólo toma valores positivos (naturalmente) y también se puede ver que es una distribución asimétrica, lo que nos da suficiente evidencia de que una distribución normal no es una buena elección.

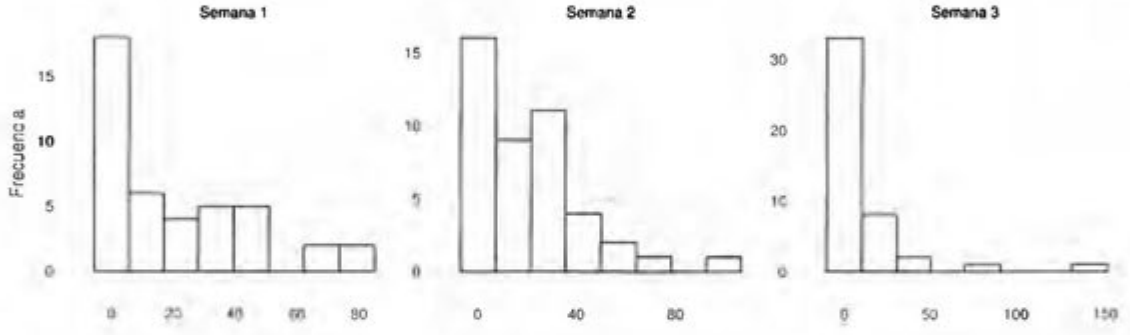


Figura 2.9: Histograma para los datos de precipitación.

Por otro lado, los datos de temperatura máxima de la primer semana entán entre los 23.09° y 34.7° grados celsius, con un promedio de 28.29° y una mediana de 28.97°, la desviación esándar para dicha semana es de 2.6236°. En la segunda semana se tiene que el rango de la temperatura máxima está entre 22.31° y 31.57°, con una mediana de 28.57° y un promedio de 27.74°, en este caso la desviación estándar es de 2.5°. Finalmente, para la tercera semana los promedios se sitúan entre los 22.1° celsius y los 32.14° celsius, tienen un promedio de 28.29°, mientras que la mediana es de 28.97° y la desviación estándar tiene un valor de 2.56°. En la figura (2.10) se presentan los histogramas para los datos de temperatura máxima medida para las primeras 3 semanas de 2014.

Los datos de temperatura mínima para estas 3 semanas tienen un comportamiento semejante a los de temperatura máxima, en la primer semana la temperatura mínima estuvo entre los 14.11° centígrados y 24.53° centígrados, el promedio alcanzó un valor de 19.76°, mientras que la mediana

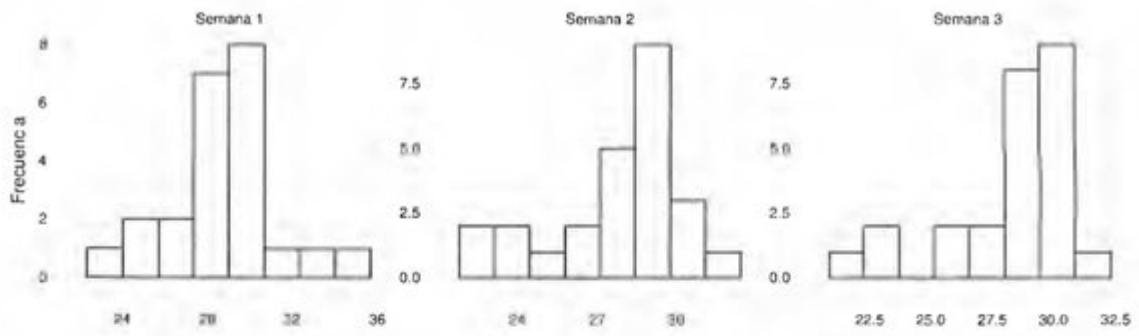


Figura 2.10: Histograma para los datos de temperatura máxima.

fue de 19.64°, además su desviación estándar es de 2.64°. Para la segunda semana la temperatura mínima estuvo en promedio entre 13.17° y 23.89° celsius, y tuvo un promedio de y una mediana de 18.95° y 18.89° respectivamente, mientras su desviación estándar fue de 2.79°. Si observamos los datos de temératura máxima y mínima para las dos primeras semanas podemos concluir que la semana 2 fue en promedio mas fresca. Luego, en la semana 3 este dato se encontró entre 11.37° y 23.09°, teniendo un promedio de 18.31° con una desviación estándar de 3.09°, además dichos datos tienen una mediana de 17.84° celsius. En la figura (2.11) se presenta un histograma para cada una de las semanas de estudio.

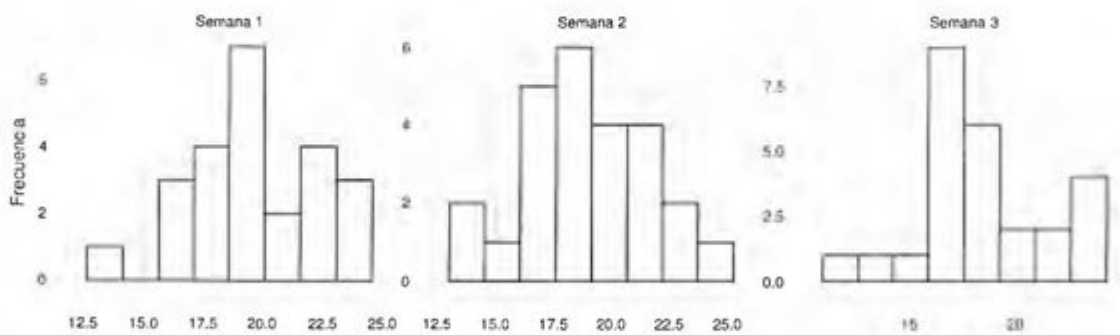


Figura 2.11: Histograma para los datos de temperatura mínima.

2.3.2. Análisis exploratorio de datos, caso espacial

Como bien se sabe, en la recolección, digitación y almacenamiento de datos pueden ocurrir errores, lo que conlleva a la posible presencia de datos con anomalías o también conocidos como datos

atípicos (outliers en inglés), el presente trabajo no se enfoca en el estudio de herramientas para la *detección* de los mismos, sin embargo, se pueden usar gráficos y el ajuste de variogramas para la eliminación de algunos datos que se consideren, desde punto de vista meramente gráfico, atípicos.

El primer escenario del análisis espacial de datos, es la graficación de los mismos en el mapa de Puerto Rico, este elemento es una herramienta muy simple, pero importante, ya que se puede ver si hay respuestas discordantes entre vecinos y también con los supuestos de continuidad e isotropía, también se puede ver si existen tendencias espaciales que puedan sugerir la necesidad de incluir un modelo de superficie con tendencia para una media espacialmente variable, o también se pueden notar comportamientos cualitativamente diferentes en diferentes subregiones. En este gráfico los datos se representan usando círculos, donde el radio de cada círculo es proporcional al valor del dato que representa con respecto a los demás datos.

Caso de precipitación

En cuanto a la variable de precipitación se realizó la transformación:

$$prect = \ln(prec + 1),$$

donde $prect$ es el valor de la variable transformada y $prec$ es el valor del dato original, esto con el fin de suavizar la variable y contrarrestar la asimetría.

En la figura (2.12) se exhibe que la región noreste de Puerto Rico es más lluviosa para esta semana en cuestión. Se puede notar en la figura (2.12) que existen datos muy cercanos espacialmente, lo que se puede traducir en un problema al momento de ajustar una función de covarianza para el variograma, una forma de resolverlo es tomar dichos datos como uno sólo (el promedio por ejemplo), o si tienen valores muy cercanos(como debería ser) tomar sólo uno de los dos, sin embargo, lo anterior se realizará sólo en el caso de que al ajustar los variogramas se presentes soluciones discordantes. No es claro si hay datos atípicos en dicha ilustración, sin embargo se puede ver que hay dos parejas de datos(estaciones climáticas) que son cercanos pero que son discordantes en magnitud, una de ellas situada en las coordenadas $(18{,}19; -67{,}0)$, y la otra aproximadamente a una latitud de 18,42 y longitud $-67{,}1$, esto puede generar problemas al momento de hacer el ajuste en el variograma para precipitación en la semana 1.

Los gráficos de dispersión que se obtienen entre las coordenadas y los datos se pueden ver en la figura (2.13), en estos se puede visualizar una vez más que la región noreste de Puerto Rico presenta más precipitaciones para esta semana en particular, esto nos confirma la necesidad de realizar las

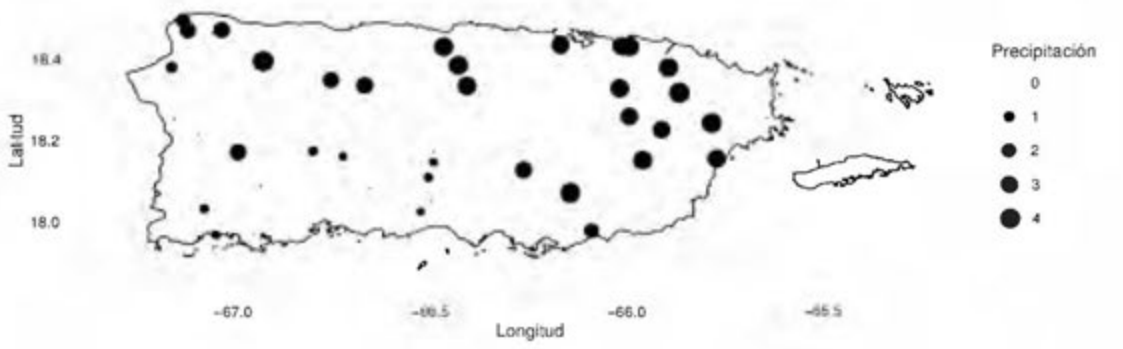


Figura 2.12: Datos de lluvia transformados en semana 1.

pruebas de hipótesis para tomar la decisión de usar un modelo con media variable o uno de media constante en la modelación de precipitación usando kriging. Un elemento más que merece ser mencionado, es que existen muchas estaciones cuya medida de precipitación es nula, eso puede generar predicciones negativas en kriging, ya que este es un proceso de suavizamiento.

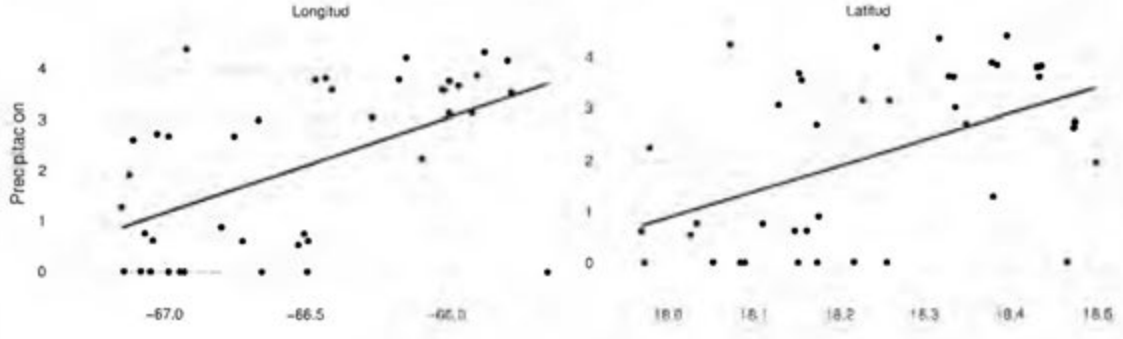


Figura 2.13: Datos de lluvia versus coordenadas en la semana 1.

Se realizó una prueba sobre la bondad de ajuste de los modelos con cociente de verosimilitud usando las devianzas, una prueba de χ^2 [1], se comenzó probando un modelo donde la media es dada por la función $\mu(x) = \beta_0$ (se refiere a la ecuación 1.1), versus un modelo con media dada por $\mu(x) = \beta_0 + \beta_1 \mathtt{Lat}(x)$, o $\mu(x) = \beta_0 + \beta_2 \mathrm{Lon}(x)$, en el caso en que alguno de los modelos saturados mejoran el ajuste, se vuelve a realizar la prueba de alguno de ellos versus uno con media $\mu(x) = \beta_0 + \beta_1 \mathtt{Lat}(x) + \beta_2 \mathrm{Lon}(x)$, como se puede observar son modelos anidados, ya que el primero de ellos es un caso particular de los otros dos tomando $\beta_2 = 0$, o $\beta_1 = 0$, y cualquiera de ellos es un caso particular del último. La prueba tiene como hipótesis nula que el modelo sencillo y parsimonioso

es el mejor, mientras que la hipótesis alternativa es que el modelo "saturado" es el mejor, el estadístico del cociente de verosimilitud es:

$$\Delta \log(L) = -2\log(L_R) - (-2\log(L_A))$$

donde L_R es la verosimilitud del modelo reducido, y L_A es la del modelo aumentado, además tiene k grados de libertad, siendo k el número de parámetros extra que tiene el modelo saturado respecto al modelo simple. Sean MCte y M1st los modelos con media constante y lineal respectivamente, además MLo y MLa modelos donde la media es solo función de la longitud y la latitud respectivamente, es decir:

$$\begin{aligned}
\texttt{MCte}: \mu(x) &= \beta_0, \\
\texttt{MLo}: \mu(x) &= \beta_0 + \beta_2 \texttt{Lon}(x), \\
\texttt{MLa}: \mu(x) &= \beta_0 + \beta_1 \texttt{Lat}(x), \\
\texttt{M1st}: \mu(x) &= \beta_0 + \beta_1 \texttt{Lat}(x) + \beta_2 \texttt{Lon}(x).
\end{aligned}$$

Además, los resultados del ajuste de los variogramas, el cual se hizo mediante el paquete geoR ([36]) de R Cran, para precipitación se resumen en el cuadro (2.1), donde se observa que según el criterio de información de Akaike (AIC), la mejor elección entre los cinco modelos, es el que tiene una media que es una función lineal de las coordenadas, mientras que el criterio de información bayesiano (BIC) parece dar como mejor modelo el que tiene la media como función de la latitud.

Modelo de Media	# de parámetros	AIC	BIC	$\log(L)$
Constante	4	141.7	148.7	-66.86
Longitud	5	143.3	152	-66.64
Latitud	5	139.9	148.6	-64.94
Lineal	6	139	149.4	-63.49

Cuadro 2.1: Ajuste en variogramas para precipitación, semana 1.

Se resumen los resultados de las pruebas de hipótesis en el cuadro (2.2), donde se da evidencia, junto con lo antes mencionado, que el mejor modelo para ajustar estos datos de precipitación, es un modelo con media variable, siendo ésta una función de la coordenada correspondiente a la latitud. Lo anterior lo vemos porque al comparar el modelo con media constante con el de media dependiendo de la longitud, no se rechaza la hipótesis nula, en la cual el modelo simple ajusta mejor el variograma.

Prueba	H_o	H_1	p-valor
MCte vs MLo	$\beta_2 = 0$	$\beta_2 \neq 0$	0.5003
MCte vs MLa	$\beta_1 = 0$	$\beta_1 \neq 0$	0.0499
Mla vs M1st	$\beta_2 = 0$ y $\beta_1 \neq 0$	$\beta_2 \neq 0$ y $\beta_1 \neq 0$	0.0881

Cuadro 2.2: Pruebas de hipótesis para precipitación, semana 1.

Por otro lado, según el test, agregar al modelo simple la coordenada de la latitud, este hace un mejor ajuste, misma conclusión que se puede hacer al examinar el índice BIC (ver [1], sección 6.1) de cada ajuste, nos da el candidato a usar para realizar la interpolación de los datos de precipitación para la semana 1.

Los parámetros obtenidos en dicho ajuste son mostrados en (2.3),

β_0	β_1	Efecto nugget (τ^2)	σ^2	ϕ
-78.04	4.3910	0.575	1.3941	0.3266

Cuadro 2.3: Parámetros de variograma ajustado en precipitación, semana 1.

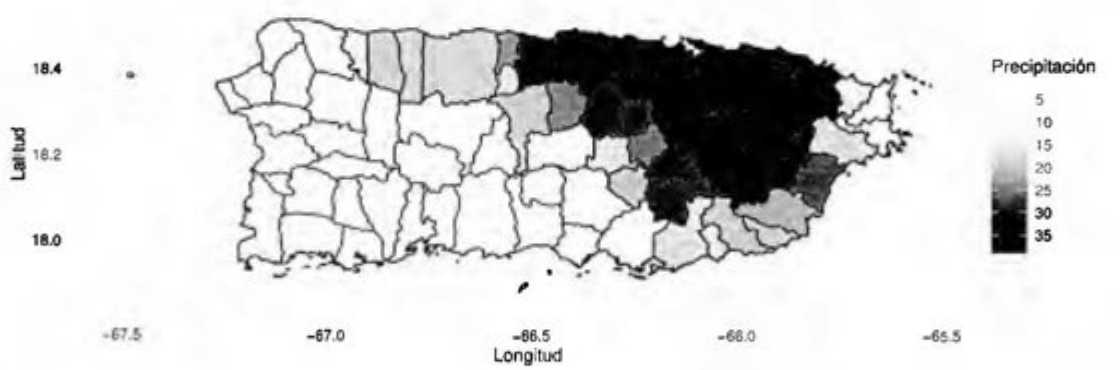


Figura 2.14: Precipitación en semana 1.

En la figura 2.14 se exhibe el resultado de la interpolación mediante kriging, y este, como cualquier método razonable de interpolación, tiene un efecto de suavizamiento [13]. Se puede observar como los municipios que comparten frontera no tienen comportamientos muy distintos, eso se debe al supuesto de continuidad en el espacio que tiene la variable interpolada.

Caso de temperatura mínima

Para la temperatura mínima, se exhiben los datos de las estaciones para la primer semana en la figura (2.15), este gráfico resulta ser poco útil para hacer conclusiones respecto al comportamiento espacial de dicha covariable, un motivo quizá es que el rango de los datos es muy pequeño, por lo que se revisan los de dispersión en la figura (2.16).

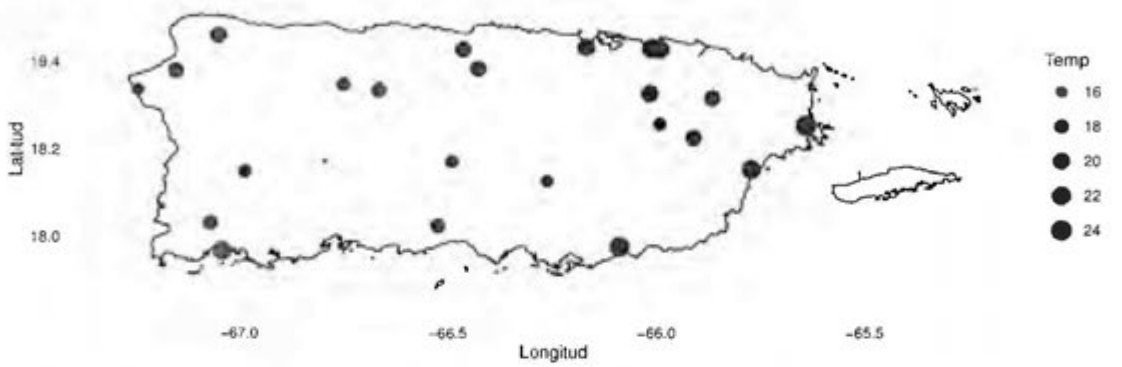


Figura 2.15: Datos temperatura mínima en semana 1.

En los gráficos de dispersión, se puede extraer información mas relevante que en la figura (2.15), por ejemplo, la coordenada de latitud no parece ser significativa para representar esta temperatura, ya que se ve claramente una recta casi horizontal, es decir, moverse de norte a sur o viceversa en Puerto Rico, no hará que la temperatura tenga un gran cambio. Por otro lado, si se camina de este a oeste, parece ser que si hay un cambio más significativo, aunque eso no es concluyente. Esta información es relevante al momento de elegir un modelo para el variograma, ya que da una idea más precisa del comportamiento espacial de esta variable. En el caso de la variable de temperatura mínima, el ajuste de los variogramas dieron los resultados mostrados en la tabla (2.4), es este caso el criterio de Akaike [1] sugiere como mejor modelo el que tiene media como función de la longitud, mientras que el criterio de información bayesiano considera un mejor ajuste con media constante.

Luego de hacer las pruebas de hipótesis, se concluye que el mejor modelo para el variograma de los datos de temperatura, es el de media constante, ya que en ambas pruebas no se rechaza la hipótesis nula, en la cual dicho modelo es el de mejor ajuste, lo que refuerza además la conclusión del BIC.

En la figura (2.17) se muestra el resultado de la interpolación realizada a los datos de la primer semana de temperatura mínima, ahí también se puede observar el efecto de suavizamiento del kriging,

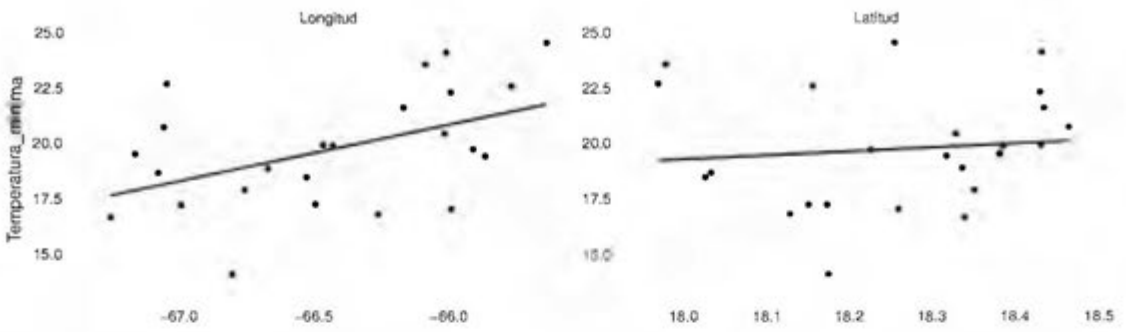


Figura 2.16: Temperatura mínima respecto a las coordenadas, semana 1.

Modelo de Media	# de parámetros	AIC	BIC	$\log(L)$
Constante	4	117.7	122.4	-54.86
Longitud	5	117.6	123.5	-53.8
Latitud	5	119.7	125.6	-54.86
Lineal	6	119.5	126.6	-53.75

Cuadro 2.4: Ajuste en variogramas para temperatura mínima, semana 1.

además que las regiones montañosas de Puerto Rico son las menos calientes para esas fechas, mientras que la zona costera mantiene las temperaturas mas elevadas, dicho comportamiento se mantiene durante las otras semanas de estudio, lo que se puede ver en el Apéndice B. Además, se presentan los parámetros obtenidos para el variograma en el cuadro (2.6).

Caso de temperatura máxima

El gráfico de datos en el mapa de Puerto Rico, no es informativo, en parte se debe a los pocos datos que se tienen para toda la región, además del rango pequeño en el que se sitúan los mismos. Las estaciones que graban temperaturas en Puerto Rico son muy pocas, lo que puede representar ajustes,

Prueba	H_o	H_1	p-valor
MCte vs MLo	$\beta_2 = 0$	$\beta_2 \neq 0$	0.976
MCte vs MLa	$\beta_1 = 0$	$\beta_1 \neq 0$	0.14

Cuadro 2.5: Pruebas de hipótesis para temperatura mínima, semana 1.

Figura 2.17: Temperatura mínima en semana 1.

β_0	β_1	β_2	Efecto nugget (τ^2)	σ^2	ϕ
20.1263	0	0	0.3501	7.1337	0.1721

Cuadro 2.6: Parámetros de variograma ajustado en temperatura mínima, semana 1.

modelos o predicciones no tan fiables, ya que por ejemplo en la primer semana solo se cuentan con 23 observaciones de las 43 estaciones que se tienen, es decir que debemos usar esa cantidad de información para hacer las predicciones en 76 ubicaciones. En la figura (2.18) se presentan los datos localizados en la isla.

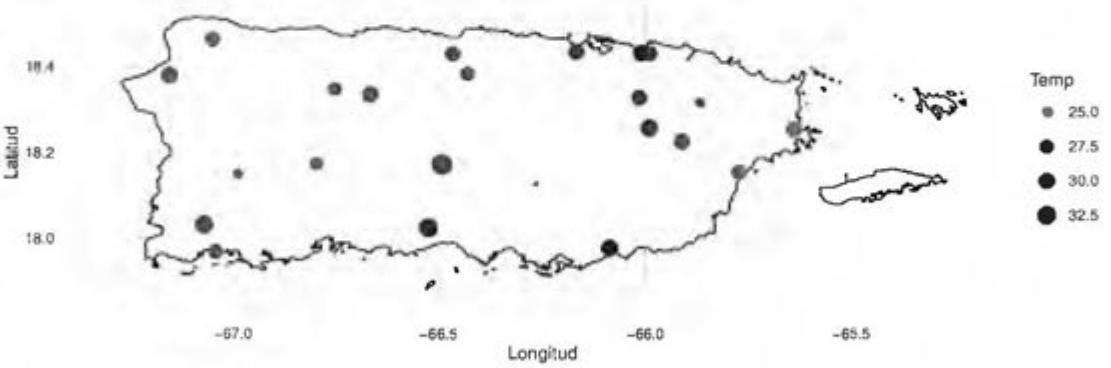


Figura 2.18: Datos de temperatura máxima en semana 1.

Si observamos el comportamiento de la temperatura máxima en la primer semana con respecto a las coordenadas, nos damos cuenta que no existe una fuerte correlación, ya que en ambos casos los gráficos de dispersión son planos(horizontales). Por lo anterior podemos asumir inicialmente que el ajuste del variograma no tomará en cuenta a las coordenadas en la media. La figura (2.19) muestra

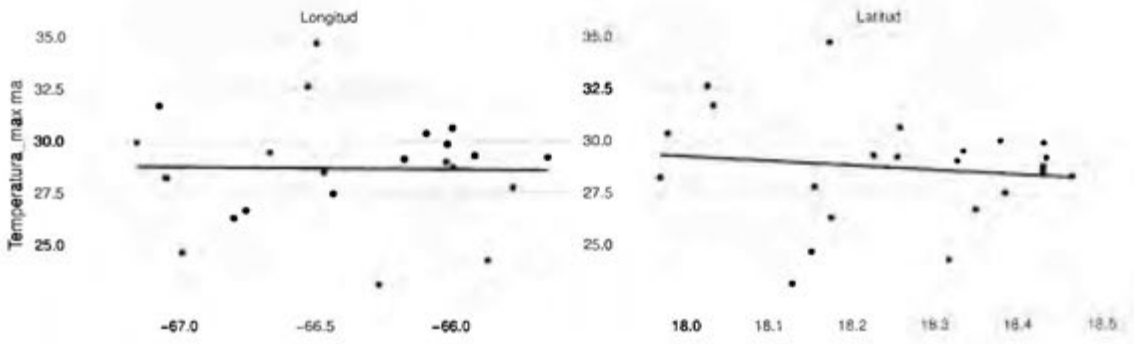


Figura 2.19: Temperatura máxima respecto a las coordenadas, semana 1.

mas claramente la afirmación anterior.

En el cuadro (2.7) se muestran los resultados obtenidos al ajustar distintos modelos para la media en el variograma. Para este caso, tanto el criterio de información de Akaike y el criterio de información bayesiano coinciden en que el mejor modelo es con media constante. Aún con la información

Modelo de Media	# de parámetros	AIC	BIC	$\log(L)$
Constante	4	116.6	121.2	-54.31
Longitud	5	118.3	123.9	-54.13
Latitud	5	118.2	123.9	-54.1
Lineal	6	120.2	127	-54.1

Cuadro 2.7: Ajuste en variogramas para temperatura máxima, semana 1.

del cuadro (2.7), se realizan las pruebas de hipótesis para concluir que modelo de variograma será usado en la interpolación. Los resultados obtenidos se muestran en el cuadro (2.8).

Prueba	H_o	H_1	p-valor
MCte vs MLo	$\beta_2 = 0$	$\beta_2 \neq 0$	0.3147
MCte vs MLa	$\beta_1 = 0$	$\beta_1 \neq 0$	1

Cuadro 2.8: Pruebas de hipótesis para temperatura máxima, semana 1.

Con lo anterior se concluye, como se suponía desde el inicio que el modelo de media constante ajustaba mejor el variograma. En la tabla (2.9) se presentan los parámetros del ajuste.

Este resultado obtenido para la temperatura máxima es un motivo suficiente para no utilizar

β_0	β_1	β_2	Efecto nugget (τ^2)	σ^2	ϕ
28.67	0	0	6.84	0	0

Cuadro 2.9: Parámetros de variograma ajustado en temperatura máxima, semana 1.

dicha covariable en el modelo CAR, ya que al hacer la interpolación con estos parámetros, queda un valor constante para toda la región de Puerto Rico en lo que respecta a temperatura máxima. Los resultados obtenidos en la primer semana se repitieron para las restantes 7 semanas, es decir, al querer ajustar un variograma exponencial al variograma empírico mediante máxima verosimilitud, resultó que el sill(ϕ) y σ^2 son cero para todas las semanas, recordando la ecuación (1.1), las predicciones para temperatura máxima en los municipios sólo dependerán de $\mu(x)$ en (1.1), por lo que queda constante en cada semana, dado que el modelo con media constante es el que mejor ajusta los datos según la prueba de hipótesis realizada.

El trabajo realizado en este capítulo, se debió repetir para cada semana que se estudia en el modelo CAR, es decir; las semanas 2, 3, 4, 31, 32, 33 y 34 de 2014, sin embargo los resultados serán incluidos en los apéndices A y B de la presente tesis.

Capítulo 3

Modelo Condicional Autoregresivo para datos de dengue en Puerto Rico

En el presente capítulo se hace la comparación de 5 modelos distintos, cada uno de estos realiza un ajuste para los casos de dengue en Puerto Rico, usando regresión de Poisson, sin embargo, no todos los modelos son espaciales, de hecho, los resultados que se dieron en algunas semanas de estudio nos indican que no hay necesidad de modelos espaciales para estos datos en dichas semanas. Los modelos son:

Modelo	Tipo	Estimación de parámetros	Siglas
Lineal Generalizado [1]	No espacial	Máxima verosimilitud, Bayesiano	GLM
Independiente, GLMM [2]	No espacial	Bayesiano	Ind
Intrínseco [2]	Espacial	Bayesiano	Int
Besag [2]	Espacial	Bayesiano	BYM
Leroux [30]	Espacial	Bayesiano	Ler

Cuadro 3.1: Modelos de ajuste a comparar.

Además, utilizando el enfoque propuesto por Lee y Mitchell en 2012 [29], se hace un estudio sobre las fronteras de riesgo en el mapa de Puerto Rico, en este caso en el mapa de riesgo se puede observar la presencia de sub-regiones en la primer semana, característica que se mantiene en las demás semanas de estudio. Como es de esperarse entonces, el modelo Lee-Mitchell sugiere que la matriz binaria utilizada para los modelos espaciales cambia en algunas de sus entradas.

3.1. 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. Entre las variables que se usan en

los modelos, existen 4 que se mantienen constantes para todas las semanas del año, estas son: altitud, nivel o porcentaje de pobreza, número de habitantes (población) y superficie por municipio (en millas cuadradas), sumado a estas, se presenta la densidad poblacional, calculada como el cociente entre la población y la superficie de cada municipio. Tales datos se resumen en el cuadro (3.2), el cual muestra los percentiles de sus distribuciones.

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^2)	4.8	28.3	40.55	53.45	126
Densidad	171.5	613	877.9	1334.1	8253.2

Cuadro 3.2: Resumen de la distribución de datos fijos.

Por otro lado, están las variables que no son fijas por semana, en el modelo son 5: la precipitación medida en décimas de milímetros, la temperatura mínima medida en grados celsius (centígrados), el índice de vegetación mejorado (EVI), los casos presuntos de dengue, que se usan en lugar de los casos confirmados, ya que estos están fuertemente correlacionados y además son más sensitivos porque aproximadamente el 60 % de los casos confirmados tienen muestras poco adecuadas para un diagnóstico definitivo [26] 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 $\ln(\mu_k) = \ln(E_k) + \theta_k$, 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 término θ_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 el cuadro 3.3 se muestra un resumen para cada semana 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 dos tipos de interpolación, una espacial en el caso de las primeras dos variables, y una temporal en el índice de vegetación, por lo que se puede esperar que para los modelos CAR estas variables aporten un "error" aún mayor con respecto a si fueran variables medidas directamente (muestras).

		Percentiles				
	Variable	0 %	25 %	50 %	75 %	100 %
Semana 1	Precipitación	0.4029	2.9044	15.5941	30.2027	39.9786
	Temp 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	49.45
	Temp 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
	Temp 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

Cuadro 3.3: Resumen de la distribución de datos variables semanales.

En el capítulo 2, sección 2.1 se presentó un mapa de colores representando la variable de riesgo relativo para la semana uno, en la figura (2.1) se puede ver que el municipio de Patillas tiene un riesgo superior, también sus vecinos, esto se debe a que dichos municipios según los informes del Departamento de Salud, pertenecen a un proyecto de vigilancia aumentada [10]. Para las primeras 3 semanas el municipio de Patillas presenta el mayor número de casos de dengue, San Juan es en esas semanas el que posee el segundo lugar en el número de casos, sin embargo la población de San Juan es poco más 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 cubran toda la isla principal de Puerto Rico, y la información del Departamento de Salud sobre la posibilidad de que algunos municipios tengan valores "atípicos", se decidió usar la información completa de la isla, además de

que no se hizo ninguna valoración en el uso de herramientas para detectar valores atípicos en los datos del estudio. Este comportamiento también se presenta en las siguientes semanas de estudio. La figura (2.1) sugiere que en Puerto Rico existen sub-regiones en las cuales el comportamiento del riesgo relativo está relacionado con el comportamiento de sus vecinos, eso puede significar que el modelo de Lee-Mitchell [29] pueda modificar el sistema vecinal que se utiliza en el presente 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 filas 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 [5]; 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 [3, 4] de R-CRAN. Los resultados de dichas pruebas se presentan en la tabla 3.4.

	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 5	0,1664	0,0128	0,0162	0,3242	0,75	0,0001
Semana 6	0,2732	0,0006	0,4282	0,0001	0,7495	0,0001
Semana 7	0,3455	0,0001	0,234	0,0003	0,6958	0,0001
Semana 8	0,2058	0,0034	0,8432	0,0001	0,6009	0,0001

Cuadro 3.4: Índices de Moran asociados a 3 variables de los modelos finales.

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 3.4 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 los mismos.

3.2. Modelos no espaciales, primer ajuste

Los modelos ajustados sin tomar en cuenta una estructura espacial son dos: modelo lineal generalizado (GLM), el cual se ajustó 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 [28], ajustado con MCMC. La forma que toma cada uno viene dada por:

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

en el GLM se tiene que:

$$\log(\mu_k) = \log(E_k) + \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6, \tag{3.1}$$

mientras que el modelo Independiente se rige por la ecuación:

$$\log(\mu_k) = \log(E_k) + \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \theta_k, \tag{3.2}$$

donde E_k es el offset que controla el tamaño de la población, β_0 es el intercepto, β_1 es el coeficiente para la variable de altitud, β_2 es el coeficiente para el porcentaje de pobreza, β_3 es la proporción del índice de vegetación en el modelo, β_4 es un coeficiente relacionado a la densidad poblacional, β_5 es el coeficiente para la precipitación promedio en el municipio, β_6 es el coeficiente para la temperatura mínima y θ_k es una variable aleatoria con distribución normal de media cero y varianza τ^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 (3.1) usando máxima verosimilitud y MCMC. En la tabla 3.5 se resume dichos resultados para las primeras tres semanas, es importante destacar que la diferencia entre los estimadores es muy poca para tales semanas, sin embargo, con máxima verosimilitud no se puede hablar de de un intervalo de predicción para cada uno de los estimadores. Para realizar el ajuste con máxima verosimilitud se usó la función glm( ) del paquete

"stats" [35], mientras que el ajuste bayesiano se realizó con la función stan_glm( ), la cual pertenece al paquete "rstanarm" [18]; las previas utilizadas por dicha función para los parámetros de ajuste asociados a las covariables son distribuciones normales $\mathcal{N}(0, 100)$.

	Semana 1		Semana 2		Semana 3	
	Verosimilitud	Bayes	Verosimilitud	Bayes	Verosimilitud	Bayes
Intercepto	-6,7199	-6,6596	-4,4317	-4,3663	-6,8358	-6,7074
Altitud	-0,0024	-0,0025	-0,0023	-0,0024	-0,0018	-0,0019
Pobreza	4,7489	4,7905	3,9488	3,9146	2,74	2,6485
EVI	0,0002	0,0002	0	0	-0,0001	-0,0001
Densidad	-0,0001	-0,0001	-0,0001	-0,0001	-0,0001	-0,0001
Prec	0,0011	0,0016	0,2533	0,2549	0,0133	0,0111
Tmin	0,1975	0,1932	0,1261	0,1227	0,3478	0,3426

Cuadro 3.5: Estimadores para modelo lineal generalizado, primer ajuste.

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 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. En la tabla 3.6 se muestran los intervalos de predicción para los parámetros en el GLM para tres de las semanas de estudio.

	Semana 1		Semana 2		Semana 3	
Intervalo	2.5 %	97.5 %	2.5 %	97.5 %	2.5 %	97.5 %
Intercepto	-10,0989	-3,2518	-7,7592	-1,0353	-10,8322	-2,6293
Altitud	-0,0051	-0,0003	-0,0051	0	-0,0046	0,0003
Pobreza	1,4784	8,2097	0,8348	7,0225	-0,1215	5,5085
EVI	0	0,0005	-0,0003	0,0002	-0,0004	0,0001
Densidad	-0,0002	0	-0,0002	0	-0,0002	0
Prec	-0,0183	0,0216	0,0024	0,5186	-0,279	0,3025
Tmin	0,0406	0,338	-0,0386	0,2808	0,1468	0,541

Cuadro 3.6: Intervalos de predicción para modelo lineal generalizado, primer ajuste.

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 demás 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 3.1, pero sin dichas covariables.

Según el DIC el modelo independiente tiene mejores resultados de ajuste que el GLM, ya que para las 8 semanas de estudio el DIC es menor, sin embargo, el índice de Watanabe nos da un resultado diferente, para las semanas 2 y 3 el GLM hace un mejor ajuste y para las semanas 1, 4, 5, 6,7 y 8 el modelo Independiente se ajusta mejor a los datos. Esos resultados se pueden ver en la tabla 3.7.

	Modelo GLM		Modelo Independiente	
	DIC	WAIC	DIC	WAIC
Semana 1	264,88	274,80	229,29	232,77
Semana 2	244,37	249,80	222,88	256,51
Semana 3	247,18	251,40	240,87	259,40
Semana 4	237,09	242,60	216,66	228,52
Semana31	248,58	255,00	216,25	220,61
Semana 32	188,79	191,70	186,67	191,43
Semana 33	226,04	230,90	198,46	152,86
Semana 34	219,59	235,20	187,76	190,85

Cuadro 3.7: DIC y WAIC para modelos GLM e Independiente, primer ajuste.

3.3. Modelos no espaciales, ajuste final

Se hace un ajuste de la misma manera que en la sección anterior, la diferencia radica en que no se toman en cuenta las covariables que no son significativas según lo mencionado en la sección 3.2, así los nuevos modelos toman la forma descrita en las ecuaciones 3.3 y 3.4,

$$\text{GLM: } \log(\mu_k) = \log(E_k) + \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_5 X_5 + \beta_6 X_6, \tag{3.3}$$

mientras que el modelo Independiente es:

$$\log(\mu_k) = \log(E_k) + \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_5 X_5 + \beta_6 X_6 + \theta_k, \tag{3.4}$$

donde cada elemento en las ecuaciones anteriores representan lo mismo que en 3.1 y 3.2. En esta ocasión los intervalos de predicción al 95 % del modelo lineal generalizado, sugieren que las covariables de mayor peso en el modelo son altitud y pobreza ya que para todas las semanas ninguno

contenía 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 [28]. En la tabla 3.8 se presentan tales intervalos.

	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
	Semana31		**Semana32**		**Semana33**		**Semana34**	
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

Cuadro 3.8: Intervalos de predicción para parámetros en GLM reducido.

Con el nuevo ajuste no se observa cambio en cuanto a la decisión de los índices DIC y WAIC, al igual que en modelo original, 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 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, Gelman y Gabry en [41], 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 trabaja para modelos singulares [41].

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

Cuadro 3.9: Parámetros de ajuste para GLM.

En el GLM, los estimadores se mantienen estables en las semanas de estudio, se puede ver en la tabla 3.9 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 algunos autores [31]. Según el resultado 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 [26].

3.4. Modelos espaciales, elección del modelo y resultados

Se ajustan 3 modelos de esta clase: modelos intrínseco, Besag descrito por la ecuación (1.38) y Leroux. El primero asume que existe una completa autocorrelación espacial en la variable dependiente, el siguiente es una suma entre el modelo intrínseco y el independiente (3.4) 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:

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

para lo modelos intrínseco y Leroux se cumple [28]:

- Intrínseco:

$$\phi_k|\phi_{-k}, W, \tau^2 \sim \mathcal{N}\left(\frac{\sum_{i=1}^n w_{ki}\phi_i}{\sum_{i=1}^n w_{ki}}, \frac{\tau^2}{\sum_{i=1}^n w_{ki}}\right), \quad (3.5)$$

$$\tau^2 \sim \mathcal{IG}(1; 0{,}01), \quad (3.6)$$

- Leroux:

$$\phi_k|\phi_{-k}, W, \tau^2 \sim \mathcal{N}\left(\frac{\rho\sum_{i=1}^n w_{ki}\phi_i}{\rho\sum_{i=1}^n w_{ki} + 1 - \rho}, \frac{\tau^2}{\rho\sum_{i=1}^n w_{ki} + 1 - \rho}\right), \quad (3.7)$$

$$\tau^2 \sim \mathcal{IG}(1; 0{,}01), \quad (3.8)$$

$$\rho \sim \mathcal{U}(0, 1). \quad (3.9)$$

Semana	1			2			3			4		
Modelo	Int	BYM	Ler	Int	BYM	Ler	Int	BYM	Ler	Int	BYM	Ler
Intercepto	-1,1	-1,3	1,6	-0,2	-0,1	1,5	0,6	-0,1	-1,1	0,3	0,3	-0,9
Altitud	-0,1	1,1	-0,3	0,8	-1,1	0	0,9	0,9	-0,2	-0,5	0,9	-0,8
Pobreza	1,3	0,9	-0,4	0,4	-0,5	-1,9	-0,3	0,3	1,1	0,4	-1,9	1,2
Prec	1,9	0,2	0,9	0,2	1,5	0	-0,7	-1,5	1,1	1,4	0,6	-0,1
Tmin	0,4	1	-1,6	0	-0,5	-1	-0,5	0,2	0,8	-0,3	0,4	0,8
τ^2	-0,3	-0,5	-0,9	0,8	1,3	1,2	0,9	-0,9	-0,4	-1,3	1,3	0
ρ	NA	NA	-1,2	NA	NA	-1,5	NA	NA	0,9	NA	NA	0,3
σ^2	NA	1,1	NA	NA	-0,8	NA	NA	0,3	NA	NA	-0,7	NA

Cuadro 3.10: Valor Z en la prueba de Geweke.

Los tres modelos se ajustan con funciones del paquete CARBayes [28], presente en el programa estadístico R-CRAN [35]. 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 explicado en la sección (1.10.3), los resultados para las semanas 1 al 4 se presentan en la tabla 3.10, se puede confirmar que en todas las semanas cada parámetro de cada modelo ha convergido a su distribución estacionaria según dicha prueba, es decir que su valor z está entre $-1{,}96$ y $1{,}96$.

En la figura 3.1 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 el gráfico 3.1, el resultado del diagnóstico de Geweke y la figura 3.2, donde se aprecia el comportamiento de la media durante la simulación.

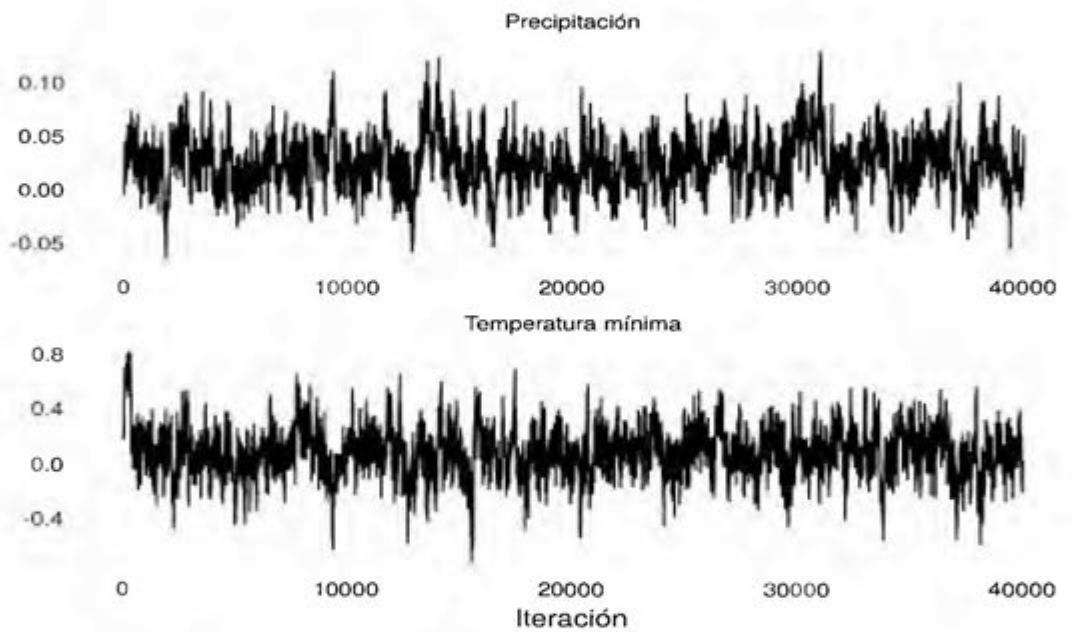


Figura 3.1: Traceplot para parámetros de Prec y Tmin en semana 1, modelo BYM.

El modelo que se toma para hacer inferencia es el Besag-York-Mollie, el criterio de información de Watanabe-Akaike(WAIC) tiene un valor inferior en 6 de las 8 semanas de estudio, en las semanas donde otro modelo ajusta mejor respecto a dicho criterio, se tiene una diferencia muy pequeña. En la tabla 3.11 se muestra las medida del criterio para los 5 modelos bayesianos.

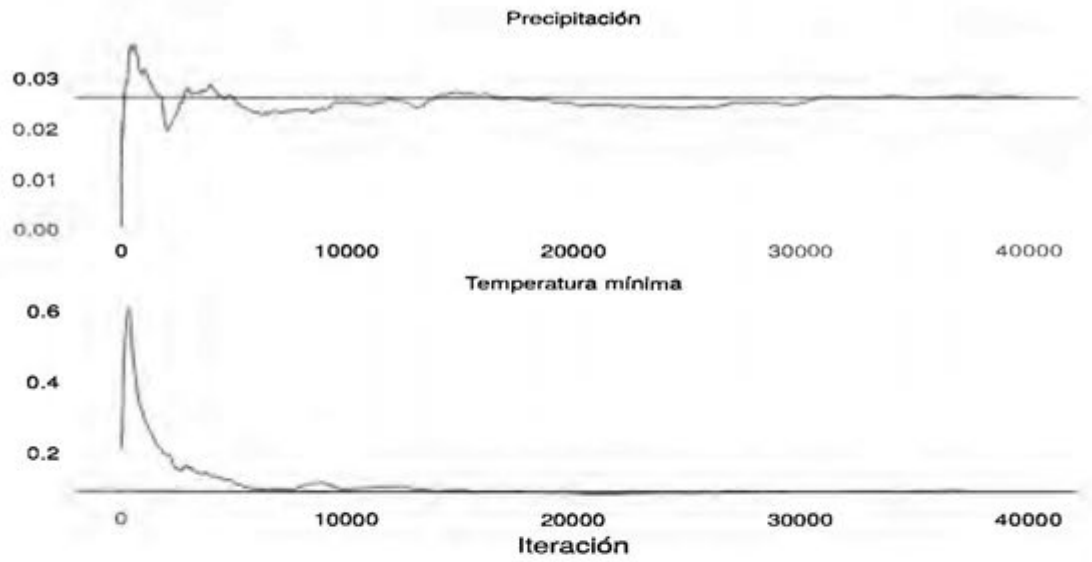


Figura 3.2: Comportamiento de la media en parámetros de Prec y Tmin, modelo BYM.

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

Cuadro 3.11: Criterio de Información de Watanabe-Akaike.

3.4.1. 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 concluyó que hace un mejor ajuste de predicción según el WAIC.

Gráficos de dispersión con lineas de regresión en la figura (3.3) representan las relaciones(lineales) crudas de las variables independientes con la dependiente para la semana 1. Dicho gráfico muestra que las tasas 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 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.

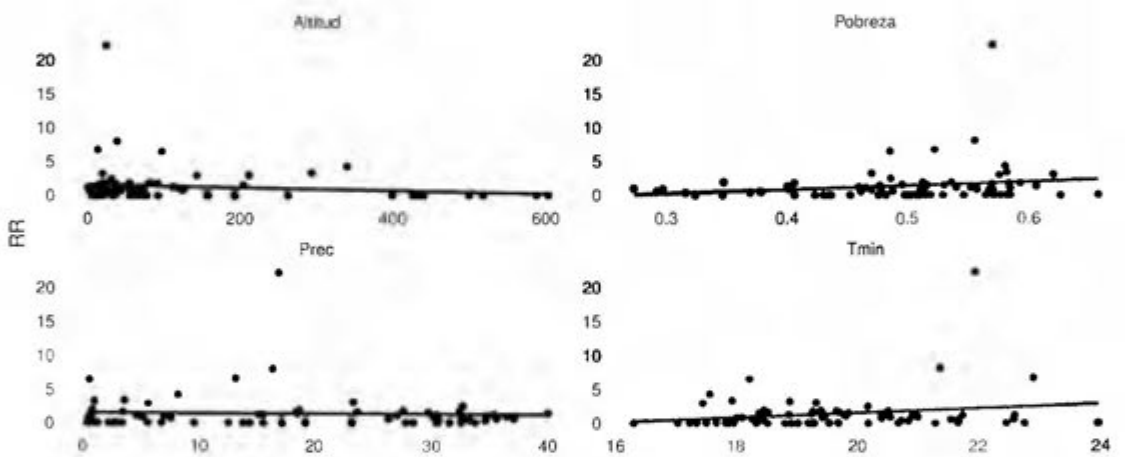


Figura 3.3: Gráficos de dispersión de variables vs Riesgo Relativo, semana 1.

El modelo BYM tiene dos efectos aleatorios, uno estructurado espacialmente, ϕ_k, donde se incorpora la información de la matriz de vecindarios W, además de que se estima un parámetro τ^2 para la varianza y otro efecto sin estructura: θ_k, independientes con distribución normal de media nula y varianza σ^2, estimada durante el ajuste con MCMC. Este modelo es el más usado en la práctica [28], sin embargo, se requiere estimar dos efectos aleatorios para cada dato puntual, pero sólo 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 [2], es por eso que el paquete CARBayes retorna la suma en las muestras del MCMC (ψ_k).

	Semana							
Parámetro	1	2	3	4	31	32	33	34
τ^2	2,0102	0,8674	0,6850	0,0231	1,9953	2,4995	2,8391	1,1925
σ^2	0,0019	0,0100	0,0018	0,3301	0,0020	0,0024	0,0021	0,0020
ψ	-0,0449	0,0020	0,0088	-0,0036	0,0037	0,0061	-0,0132	0,0040

Cuadro 3.12: Mediana de parámetros de efectos aleatorios para modelo BYM.

En la tabla 3.12 es claro que el parámetro de varianza para el efecto no estructurado es pequeña, 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 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 ψ 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 se le calcula $\psi_k = \phi_k + \theta_k$.

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 primer semana, se puede ver que no hay una estructura definida, además en el municipio de Patillas se obtiene un valor residual muy grande comparado con los demás municipios, 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 ya que al hacerlo se perdería 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 de Patillas sigue siendo superior con respecto a los demás, en el Apéndice E se presentan gráficos de dispersión para los residuos del modelo BYM, mientras que en la figura 3.4 se resumen los datos de casos, las predicciones del modelo y sus respectivos residuos.

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 figura 3.5 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 [23].

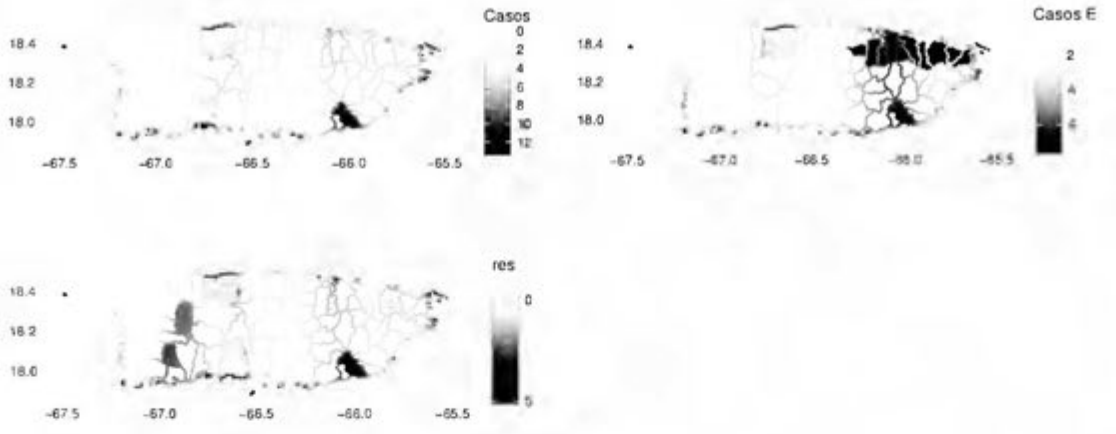


Figura 3.4: Casos, casos estimados y residuos de predicción en semana 1.

En la tabla 3.13 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 de peso para predecir el riesgo de contraer dengue en Puerto Rico.

Parámetro	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

Cuadro 3.13: Parámetros de regresión para covariables, modelo BYM.

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.

En la tabla 3.14 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

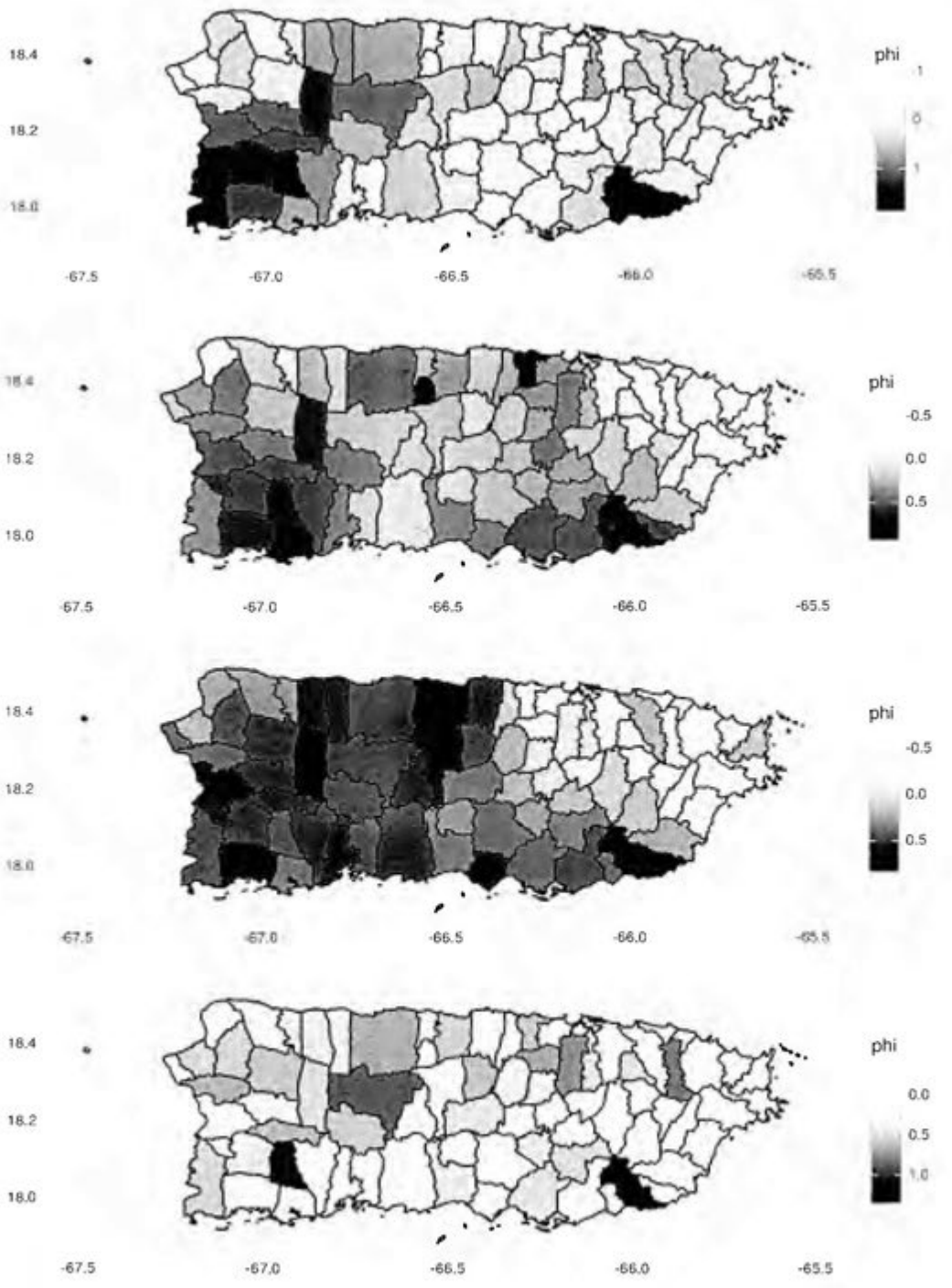


Figura 3.5: Efecto aleatorio espacial de riesgo relativo, semana 1 a semana 4.

	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

Cuadro 3.14: Intervalos de predicción en modelo BYM reducido, 95 %.

31 y 32, misma probabilidad para la tercer semana de obtener un parámetro positivo asociado a temperatura.

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 [3, 4] en R-CRAN, geary.mc() y moran.mc(), donde se utilizan 5000 permutaciones con el objetivo de obtener mejores resultados, ya que en [20] aunque sugieren el uso de dichos índices como herramientas exploratorias de asociación espacial, también sugieren usar los índice 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.

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

Cuadro 3.15: Pruebas de significancia espacial en residuos de modelo BYM.

En la tabla 3.15 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.

3.4.2. Resultados del modelo Lee-Mitchell

En el artículo de Lee [29] los autores afirman que se modela la presencia o ausencia de fronteras con métricas de disimilitud definidas en forma general como la diferencia absoluta del valor de alguna covariable entre dos regiones. En todos los modelos de la tesis, el porcentaje de pobreza resulta ser una variable explicativa de mucha importancia para el riesgo relativo, es por eso que se generó una matriz de disimilitud o métrica de disimilitud como la diferencia absoluta del porcentaje de pobreza de cada par de municipios y es una de las métricas usada en el "clustering" para riesgo relativo(sugerido por [28] y [29]). La otra métrica es generada a partir de la altitud de los municipios, usando el mismo método usado con pobreza.

Figura 3.6: Superficie de riesgo de los datos en semana 1.

La función usada para implementar el modelo CAR localizado [29] es S.CARdissimilarity() desarrollada en el paquete CARBayes, entre los argumentos de la misma está una lista de matrices de disimilitud, para este caso se usan tres matrices: las matrices definidas en el párrafo anterior y una matriz de distancias(en kilómetros) entre los centros de población de los municipios, elaborada a partir de información obtenida mediante la herramienta API de Google Maps©. Se realizan 50000 simulaciones con MCMC, de las que se eliminan las 10000 primeras, por lo que la inferencia se realiza sobre 40000 muestras restantes. El objetivo principal es detectar fronteras en la superficie de riesgo, el modelo hace la detección cuando en regiones contiguas convierte el 1 en la matriz de contigüidad binaria a 0, ya que si w_{kj} se estima como cero, el efecto aleatorio en las áreas (k, j) son condicional-

mente independientes, lo que corresponde a la presencia de una frontera en la superficie de riesgo. En contraste, si $w_{kj} = 1$ los efectos aleatorios son correlacionados, lo cual corresponde a que no hay frontera [29]. En la figura 3.6 se presenta un mapa de la superficie de riesgo de los datos.

Según el ajuste que realiza el modelo, para la semana 1 se encuentran 21 fronteras para la disimilitud basada en pobreza, 38 para la matriz generada por altitud y no hubo cambios al usar la matriz de distancias, indicando que hay municipios que comparten frontera, pero que según dos de las matrices de disimilitud son poblaciones muy distintas, motivo por el cual sus respectivos riesgos son condicionalmente independientes. El resultado se muestra en la figura 3.7 para pobreza y en la figura 3.8 para la disimilitud basada en altitud.

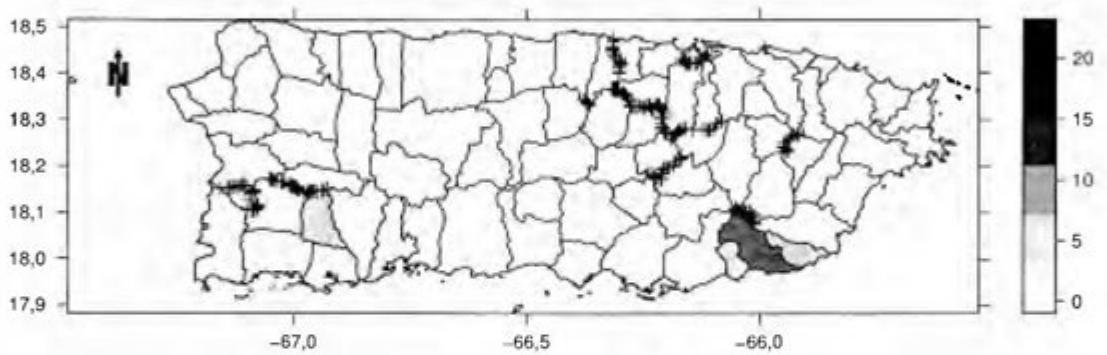


Figura 3.7: Superficie de riesgo estimado en semana 1 con disimilitud de pobreza.

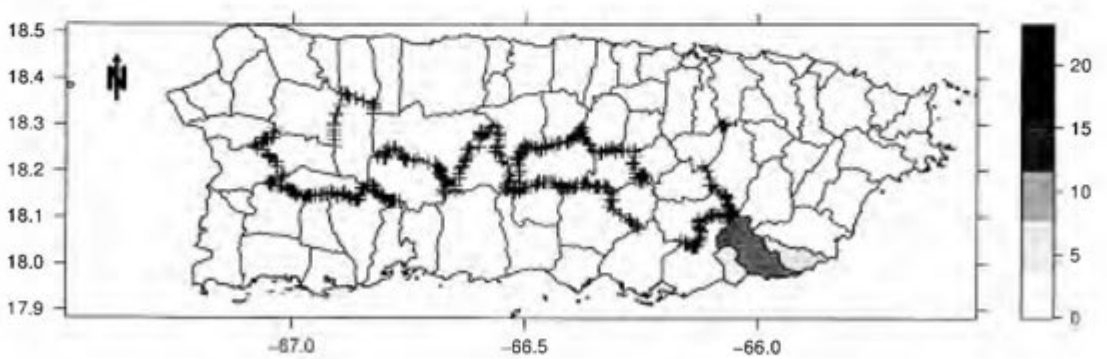


Figura 3.8: Superficie de riesgo estimado en semana 1 con disimilitud de altitud.

Siempre que se realiza simulación con MCMC se corre el riesgo de obtener divergencia en la cadena, es por eso que en todos los modelos se realizan las simulaciones repetidamente hasta obtener cadenas convergentes, para evaluar la convergencia se usa el diagnóstico de Geweke. Los resultados del modelo Lee-Mitchell respecto a las fronteras encontradas se resumen en la tabla 3.16. En el

Disimilitud	Semana	1	2	3	4	31	32	33	34
Pobreza	Cambios	21	21	21	31	12	11	5	43
	Sin Cambios	162	162	162	152	171	172	178	140
	WAIC	227.11	244.28	240.45	259.42	206.49	189.91	198.62	182.66
Altitud	Cambios	38	81	15	28	0	0	26	55
	Sin Cambios	145	102	168	155	183	183	157	128
	WAIC	227.01	251.28	245.85	254.9	210.92	187.91	204.17	185.78

Cuadro 3.16: Resultado de modelo Lee-Mitchell.

apéndice D se presentan los mapas de las demás semanas de estudio.

3.5. Conclusiones

Entre las fortalezas de un estudio como el presente es que se usa un modelo espacial sofisticado y a la vez sencillo para evaluar potenciales predictores para la incidencia de la fiebre de dengue. Se puede usar información integral y detallada sobre factores sociales, ecológicos y climáticos de los municipios para vincularla sobre toda la región en estudio, para luego integrarla en los modelos estadísticos. Aunque algunas de las variables que se desean incorporar al modelo no se tienen para todas las regiones de estudio, es posible realizar interpolación espacial para obtener una estimación de las mismas en los puntos de interés, sin embargo, en el presente trabajo no se tomó en cuenta la parte temporal, es decir, puede ser posible que se obtengan mejores resultados realizando interpolación espacio-temporal. Los resultados obtenidos en estudios como este, pueden tener implicaciones de valor en las decisiones de salud pública, haciendo la identificación de factores de riesgo así como regiones de alto riesgo con el objetivo de prevenir y controlar posibles epidemias.

Entre algunas de las limitaciones es que puede existir sesgos de medición o falta de información, por ejemplo las notificaciones por casos de dengue podrían ser menores a los casos reales, ya 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 de 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 [23] mencionan como ejemplo el uso de aire acondicionado, ya que si la tendencia del calentamiento global sigue, el uso de estos aumentará, lo que puede implicar una reducción en la probabilidad de transmisión de dengue al estar las personas menos expuestas al vector que entra a los hogares por puertas o ventanas abiertas. Otra limitación de este estudio es que no toma en cuenta la posible asociación temporal que tienen todas las variables del modelo, tanto a nivel de covariables como de variable dependiente. Como se sugiere en [11, 26, 45], las variables temperatura y precipitación se deberían relacionar con retraso, ya que un incremento en la temperatura o la lluvia de determinado mes incide directamente en un aumento en el vector [11], sin embargo estos llegarán a edad adulta meses después y es en ese momento cuando pueden transmitir el virus.

Apéndice A

Cuadros de ajuste en variogramas

Semana	Media del modelo	# de parámetros	AIC	BIC	log(L)	τ^2	σ^2	ϕ
1	Constante	4	141,7	148,7	-66,86	0,543	2,364	0,4885
	Longitud	5	143,3	152	-66,64	0,5611	1,8178	0,3632
	Latitud	5	139,9	148,6	-64,94	0,5765	1,3941	0,3266
	Lineal	6	139	149,4	-63,49	1,244	0	0
2	Constante	4	141,3	148,4	-66,65	0,4229	1,7556	0,3548
	Longitud	5	140,8	149,7	-65,39	1,44	0	0
	Latitud	5	142,9	151,8	-66,46	0,4354	1,6678	0,349
	Lineal	6	141,4	152,1	-64,72	1,11	0	0
3	Constante	4	142,1	149,3	-67,05	0	4,6845	0,1178
	Longitud	5	135,9	144,9	-62,95	0	1,1173	0,0661
	Latitud	5	142,9	151,9	-66,45	0	1,4837	0,0982
	Lineal	6	144,8	155,6	-66,39	1,119	0	0
4	Constante	4	134,5	141,8	-63,26	0	2,0608	0,2293
	Longitud	5	134,3	143,3	-62,14	0	1,5631	0,1611
	Latitud	5	135,9	145	-62,97	0	1,8942	0,2065
	Lineal	6	135,5	146,3	-61,74	0	1,4283	0,1416

Cuadro A.1: Ajuste en variogramas para precipitación, semana 1-4.

Semana	Media del modelo	# de parámetros	AIC	BIC	log(L)	τ^2	σ^2	ϕ
31	Constante	4	141,7	148,7	-66,86	0,5731	0,9297	0,0617
	Longitud	5	143,3	152	-66,64	0,5418	0,9497	0,0562
	Latitud	5	139,9	148,6	-64,94	0,4634	0,9094	0,0442
	Lineal	6	139	149,4	-63,49	0,3968	0,9702	0,0396
32	Constante	4	141,3	148,4	-66,65	0,4693	1,2937	0,1614
	Longitud	5	140,8	149,7	-65,39	0,4643	1,2674	0,1554
	Latitud	5	142,9	151,8	-66,46	0,4713	1,2789	0,1594
	Lineal	6	141,4	152,1	-64,72	0,4662	1,254	0,1537
33	Constante	4	142,1	149,3	-67,05	0	1,4485	0,0888
	Longitud	5	135,9	144,9	-62,95	0	1,3818	0,0823
	Latitud	5	142,9	151,9	-66,45	0	1,0636	0,0565
	Lineal	6	144,8	155,6	-66,39	0	0,9823	0,0478
34	Constante	4	134,5	141,8	-63,26	0,2664	0,1238	0,2647
	Longitud	5	134,3	143,3	-62,14	0,2742	0,0741	0,1522
	Latitud	5	135,9	145	-62,97	0,2679	0,1042	0,2472
	Lineal	6	135,5	146,3	-61,74	0,2804	0,0506	0,128

Cuadro A.2: Ajuste en variogramas para precipitación, semanas 31 a 34.

Semana	Media del modelo	# de parámetros	AIC	BIC	log(L)	τ^2	σ^2	ϕ
2	Constante	4	125,5	130,3	-58,73	1,2728	7,3618	0,2222
	Longitud	5	126	132,1	-58,02	1,1344	6,0591	0,1485
	Latitud	5	127	133,1	-58,5	1,5457	6,4725	0,2068
	Lineal	6	127,2	134,5	-57,6	1,3607	5,1785	0,1201
3	Constante	4	137,1	142,1	-64,53	4,4714	5,5899	0,3001
	Longitud	5	138,6	144,9	-64,29	4,6839	4,5354	0,2355
	Latitud	5	139	145,2	-64,48	4,6255	5,2331	0,2972
	Lineal	6	140,4	148	-64,21	4,9028	4,0941	0,2288
4	Constante	4	126	130,9	-59	2,3731	6,0942	0,2557
	Longitud	5	126,3	132,3	-58,13	2,012	4,8376	0,1352
	Latitud	5	127,9	134	-58,94	2,506	5,771	0,253
	Lineal	6	128	135,4	-58,02	2,1168	4,559	0,1264
31	Constante	4	141,7	148,7	-66,86	0	7,2619	0,1707
	Longitud	5	143,3	152	-66,64	0	5,8551	0,1155
	Latitud	5	139,9	148,6	-64,94	0	7,0768	0,1633
	Lineal	6	139	149,4	-63,49	0	5,6105	0,1043
32	Constante	4	141,3	148,4	-66,65	0	7,9087	0,1634
	Longitud	5	140,8	149,7	-65,39	5,41	0	0
	Latitud	5	142,9	151,8	-66,46	0	7,769	0,1582
	Lineal	6	141,4	152,1	-64,72	5,333	0	0
33	Constante	4	142,1	149,3	-67,05	0	7,7859	0,1272
	Longitud	5	135,9	144,9	-62,95	5,459	0	0
	Latitud	5	142,9	151,9	-66,45	6,806	0	0
	Lineal	6	144,8	155,6	-66,39	0	5,3467	0,0495
34	Constante	4	134,5	141,8	-63,26	0	5,0329	0,0878
	Longitud	5	134,3	143,3	-62,14	3,965	0	0
	Latitud	5	135,9	145	-62,97	0	4,7923	0,0732
	Lineal	6	135,5	146,3	-61,74	3,842	0	0

Cuadro A.3: Ajuste en variogramas de temperatura mínima.

Semana	Media del modelo	# de parámetros	AIC	BIC	log(L)	τ^2	σ^2	ϕ
2	Constante	4	123,7	128,6	-57,87	6	0	0
	Longitud	5	125,7	131,8	-57,87	5,99	0	0
	Latitud	5	125,6	131,7	-57,79	5,962	0	0
	Lineal	6	127,6	134,9	-57,79	5,9603	0	0
3	Constante	4	121,1	130	-58,54	6,33	0	0
	Longitud	5	126,5	132,6	-58,27	6,197	0	0
	Latitud	5	126,4	132,5	-58,22	6,173	0	0
	Lineal	6	128	135,4	-58,02	6,0717	0	0
4	Constante	4	122	126,8	-56,98	5,585	0	0
	Longitud	5	123,7	129,8	-56,85	5,5278	0	0
	Latitud	5	123,9	130	-56,93	5,5652	0	0
	Lineal	6	125,6	132,9	-56,82	5,5149	0	0
31	Constante	4	141,7	148,7	-66,86	3,954	0	0
	Longitud	5	143,3	152	-66,64	3,9493	0	0
	Latitud	5	139,9	148,6	-64,94	3,901	0	0
	Lineal	6	139	149,4	-63,49	3,8881	0	0
32	Constante	4	141,3	148,4	-66,65	3,807	0	0
	Longitud	5	140,8	149,7	-65,39	3,7934	0	0
	Latitud	5	142,9	151,8	-66,46	3,8061	0	0
	Lineal	6	141,4	152,1	-64,72	3,792	0	0
33	Constante	4	142,1	149,3	-67,05	0	4,439	0
	Longitud	5	135,9	144,9	-62,95	0	4,423	0,0002
	Latitud	5	142,9	151,9	-66,45	0	4,2854	0
	Lineal	6	144,8	155,6	-66,39	0	4,28	0
34	Constante	4	134,5	141,8	-63,26	6,332	0	0
	Longitud	5	134,3	143,3	-62,14	6,301	0	0
	Latitud	5	135,9	145	-62,97	6,318	0	0
	Lineal	6	135,5	146,3	-61,74	6,2793	0	0

Cuadro A.4: Ajuste en variogramas de temperatura máxima.

Apéndice B

Mapas de interpolación espacial

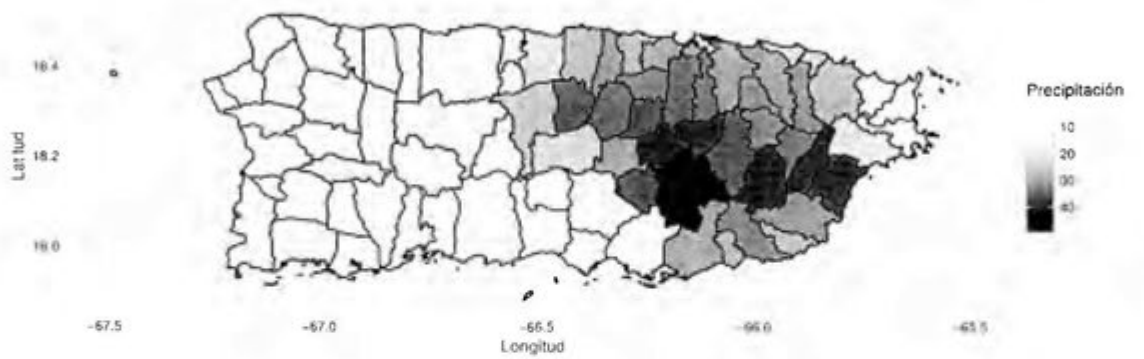


Figura B.1: Precipitación en semana 2.

Figura B.2: Precipitación en semana 3.

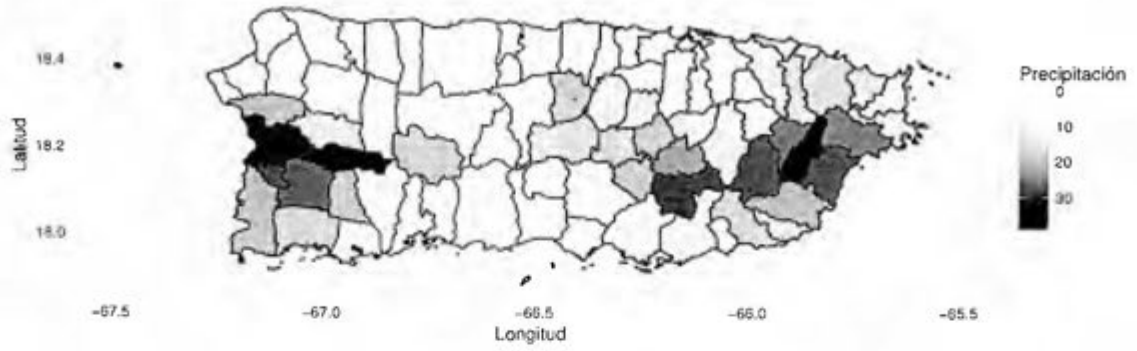


Figura B.3: Precipitación en semana 4.

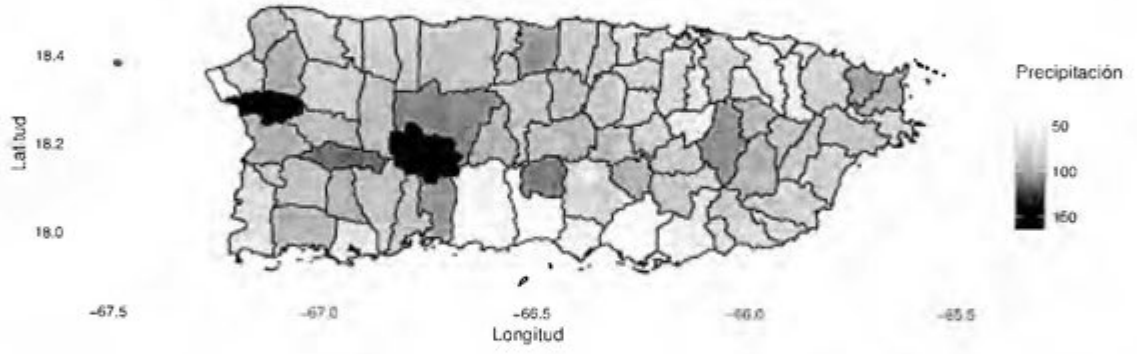


Figura B.4: Precipitación en semana 31.

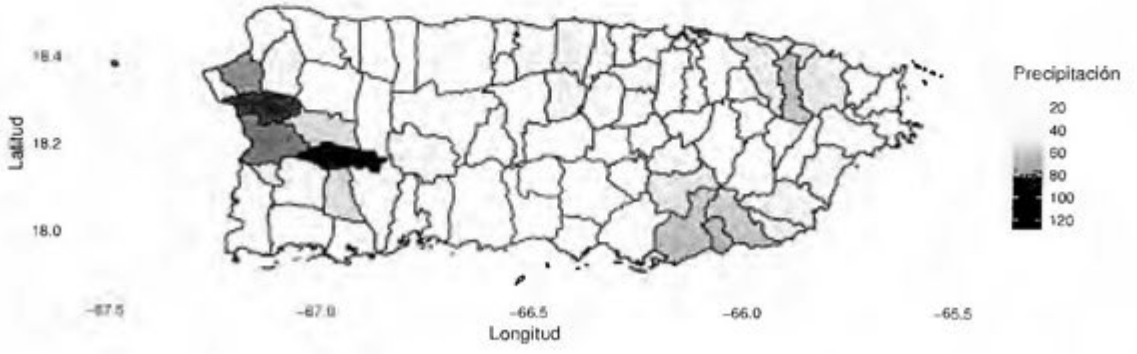


Figura B.5: Precipitación en semana 32.

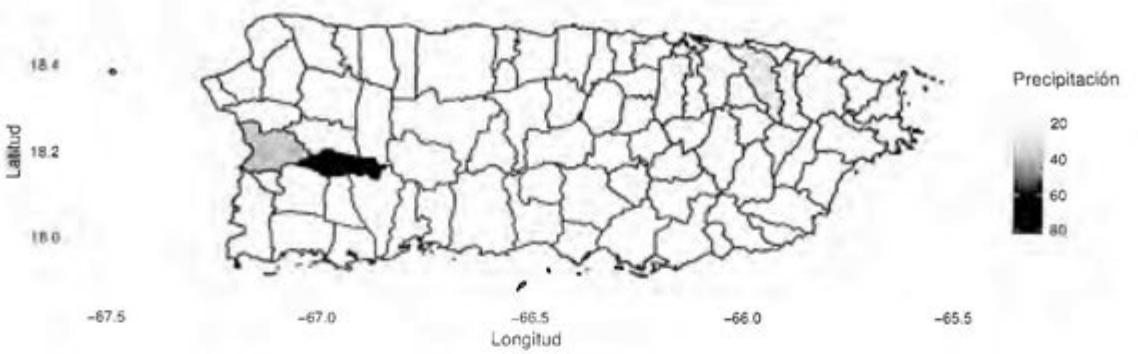


Figura B.6: Precipitación en semana 33.

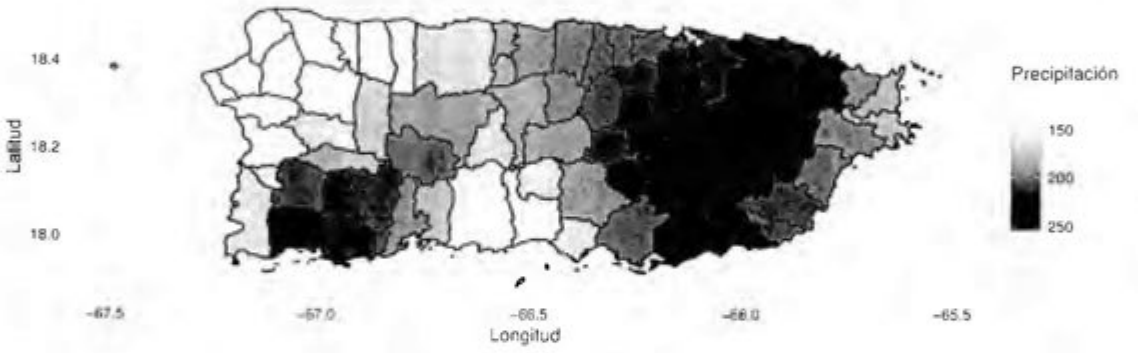


Figura B.7: Precipitación en semana 34.

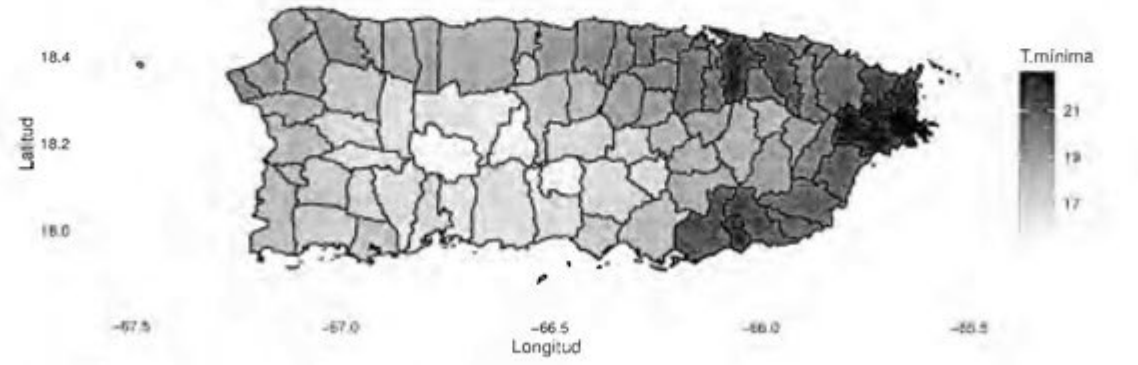


Figura B.8: Temperatura mínima en semana 2.

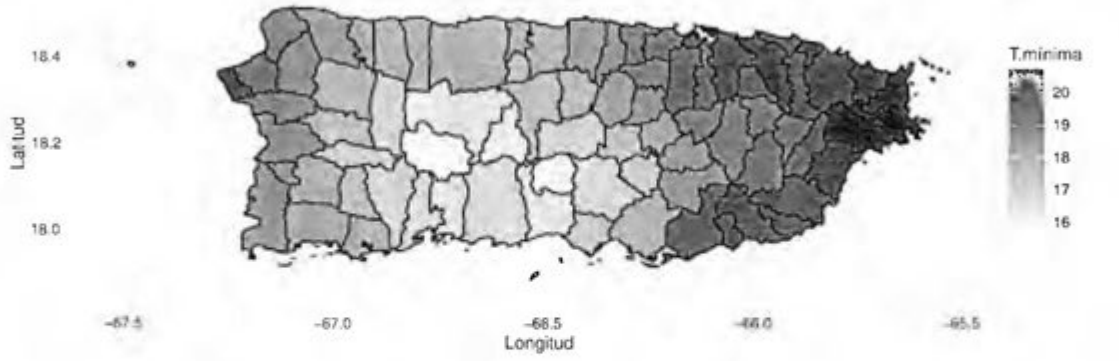


Figura B.9: Temperatura mínima en semana 3.

Figura B.10: Temperatura mínima en semana 4.

Figura B.11: Temperatura mínima en semana 31.

Figura B.12: Temperatura mínima en semana 32.

Figura B.13: Temperatura mínima en semana 33.

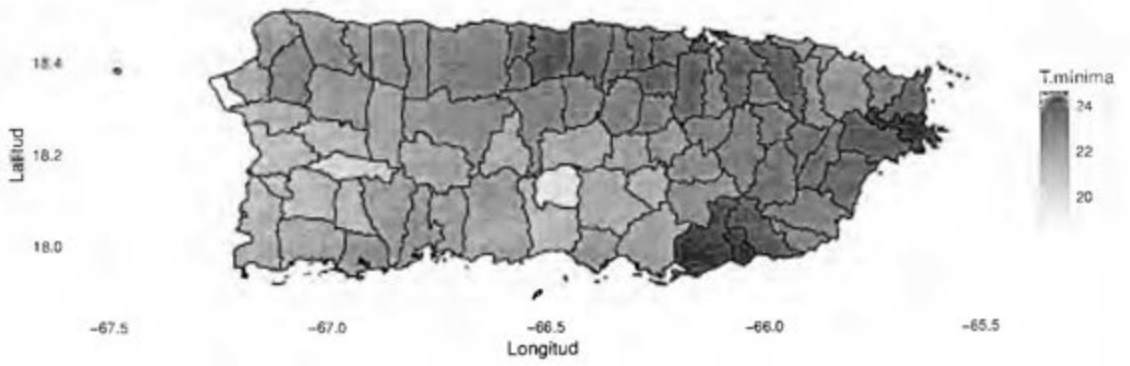


Figura B.14: Temperatura mínima en semana 34.

Apéndice C

Intervalos de predicción

C.1. Intervalos de modelos iniciales

	Semana1		Semana2		Semana3		Semana4	
	2,50 %	97,50 %	2,50 %	97,50 %	2,50 %	97,50 %	2,50 %	97,50 %
Intercepto	-10,0989	-3,2518	-7,7592	-1,0353	-10,8322	-2,6293	-8,6424	-0,1582
Altitud	-0,0051	-0,0003	-0,0051	0	-0,0046	0,0003	-0,0021	0,0023
Pobreza	1,4784	8,2097	0,8348	7,0225	-0,1215	5,5085	-3,0895	2,6279
EVI	0	0,0005	-0,0003	0,0002	-0,0004	0,0001	-0,0001	0,0004
Densidad	-0,0002	0	-0,0002	0	-0,0002	0	-0,0002	0
Prec	-0,0183	0,0216	0,0024	0,5186	-0,279	0,3025	-0,0319	0,021
Tmin	0,0406	0,338	-0,0386	0,2808	0,1468	0,541	0,0164	0,423
	Semana31		Semana32		Semana33		Semana34	
Intercepto	-1,4622	6,6812	0,0118	10,131	-7,182	1,0055	-6,9024	7,4729
Altitud	-0,0047	0,0003	-0,0058	0,0007	-0,0059	0,0003	-0,0039	0,002
Pobreza	-4,2865	0,1401	-6,7175	-1,4342	-5,0969	-0,2714	-8,1836	-1,0361
EVI	-0,0004	0,0001	-0,0004	0,0001	-0,0001	0,0004	-0,0002	0,0004
Densidad	0,0001	0,0002	0,0002	0,0003	0,0002	0,0003	-0,0001	0,0001
Prec	-0,0093	0,0126	-0,0043	0,0233	-0,0195	0,0409	-0,0148	0,0005
Tmin	-0,2519	0,0864	-0,394	0,0377	-0,0408	0,2804	-0,197	0,433

Cuadro C.1: Intervalo de predicción para parámetros en GLM.

	Semana1		Semana2		Semana3		Semana4	
	2,50 %	97,50 %	2,50 %	97,50 %	2,50 %	97,50 %	2,50 %	97,50 %
Intercepto	-10,9117	0,5535	-8,4017	0,2858	-11,7362	-1,8451	-11,5172	1,5848
Altitud	-0,0073	0,0004	-0,0055	0,0005	-0,0049	0,0006	-0,0028	0,0035
Pobreza	-0,4044	10,2457	-0,1969	7,4970	-0,5185	6,1172	-4,7786	3,9927
EVI	-0,0002	0,0006	-0,0004	0,0003	-0,0004	0,0002	-0,0002	0,0006
Densidad	-0,0003	0,0003	-0,0003	0,0001	-0,0003	0,0000	-0,0003	0,0002
Prec	-0,0297	0,0382	-0,0859	0,5477	-0,3320	0,3366	-0,0426	0,0350
Tmin	-0,1817	0,3762	-0,0854	0,3111	0,0988	0,5708	-0,0919	0,5492
τ^2	0,2088	1,3194	0,0005	0,7847	0,0004	0,4817	0,0296	1,0493
	Semana31		Semana32		Semana33		Semana34	
Intercepto	-3,4767	11,3937	-0,0536	11,2212	-14,6914	11,6541	-12,2621	8,4010
Altitud	-0,0083	0,0005	-0,0059	0,0006	-0,0088	0,0135	-0,0066	0,0020
Pobreza	-5,6356	3,9189	-7,1221	-1,4569	-7,4926	22,6203	-6,7210	4,3293
EVI	-0,0005	0,0004	-0,0005	0,0002	-0,0010	0,0007	-0,0003	0,0007
Densidad	0,0000	0,0006	0,0002	0,0003	0,0000	0,0009	0,0000	0,0005
Prec	-0,0139	0,0232	-0,0044	0,0245	-0,0773	0,0817	-0,0145	0,0076
Tmin	-0,5304	0,1044	-0,4389	0,0342	-0,7271	0,4779	-0,3723	0,5370
τ^2	0,2116	1,5946	0,0004	0,2255	0,2289	176,4934	0,1535	1,3767

Cuadro C.2: Intervalo de predicción para parámetros en modelo independiente.

	Semana1		Semana2		Semana3		Semana4	
	2,50 %	97,50 %	2,50 %	97,50 %	2,50 %	97,50 %	2,50 %	97,50 %
Intercepto	-14,653	0,3785	-10,908	0,3571	-17,1366	-1,7838	-11,888	0,566
Altitud	-0,006	0,0008	-0,0055	0,0007	-0,0044	0,0014	-0,0023	0,003
Pobreza	-0,7998	10,4813	-0,7082	7,4276	-2,5222	5,6455	-4,7648	2,963
EVI	-0,0002	0,0006	-0,0003	0,0003	-0,0004	0,0002	-0,0001	0,0005
Densidad	-0,0004	0,0002	-0,0003	0,0001	-0,0003	0,0001	-0,0003	0,0001
Prec	-0,022	0,0813	-0,339	0,8424	-0,5074	0,5147	-0,0362	0,0345
Tmin	-0,1939	0,5742	-0,1043	0,4588	0,118	0,9101	-0,0247	0,6133
τ^2	0,7228	4,4516	0,0193	2,4701	0,0088	2,1398	0,0024	2,583
	Semana31		Semana32		Semana33		Semana34	
Intercepto	-5,5367	15,9474	-2,1666	15,8808	-20,6194	9,6972	-16,9277	8,2096
Altitud	-0,0058	0,0025	-0,0054	0,0025	-0,0071	0,005	-0,0066	0,0021
Pobreza	-10,629	0,4103	-10,8327	-0,5747	-14,8388	1,5854	-8,6663	2,1956
EVI	-0,0005	0,0003	-0,0006	0,0002	-0,0006	0,0007	-0,0002	0,0007
Densidad	0,0001	0,0006	0,0001	0,0006	-0,0001	0,0012	-0,0001	0,0004
Prec	-0,016	0,0164	-0,0102	0,0276	-0,0692	0,0781	-0,0156	0,0175
Tmin	-0,6637	0,2881	-0,6406	0,1482	-0,3687	0,9316	-0,3711	0,7274
τ^2	0,516	4,7979	0,0013	4,7011	0,7437	12,7684	0,4183	3,9885

Cuadro C.3: Intervalo de predicción para parámetros en modelo intrínseco.

	Semana1		Semana2		Semana3		Semana4	
	2,50 %	97,50 %	2,50 %	97,50 %	2,50 %	97,50 %	2,50 %	97,50 %
Intercepto	-14,5259	0,2319	-10,4656	-0,1093	-17,8645	-1,4688	-12,5376	0,8338
Altitud	-0,0058	0,0008	-0,0054	0,0007	-0,0044	0,0016	-0,0024	0,0034
Pobreza	-1,117	9,9188	-0,6491	7,6412	-2,3816	5,6002	-4,6925	3,8986
EVI	-0,0001	0,0006	-0,0004	0,0003	-0,0004	0,0002	-0,0002	0,0005
Densidad	-0,0004	0,0002	-0,0003	0,0001	-0,0003	0,0001	-0,0003	0,0001
Prec	-0,0239	0,0774	-0,2529	0,7686	-0,5383	0,5009	-0,0405	0,0358
Tmin	-0,1978	0,565	-0,0824	0,431	0,1031	0,9506	-0,0566	0,631
τ^2	0,5458	4,3089	0,0008	2,2732	0,0018	2,0845	0,001	2,4497
σ^2	0,0003	0,2058	0,0003	0,4754	0,0003	0,1025	0,0004	0,8752
	Semana31		Semana32		Semana33		Semana34	
Intercepto	-5,1928	13,6796	-2,1254	13,6376	-14,9813	8,2921	-17,1536	8,2114
Altitud	-0,0066	0,0022	-0,0049	0,0029	-0,007	0,0025	-0,0065	0,0018
Pobreza	-10,0602	0,4787	-9,4844	-0,6707	-9,3281	1,8152	-8,5894	2,3256
EVI	-0,0005	0,0004	-0,0005	0,0003	-0,0006	0,0006	-0,0002	0,0007
Densidad	0,0001	0,0006	0,0002	0,0005	-0,0002	0,0007	-0,0002	0,0004
Prec	-0,0162	0,017	-0,0113	0,026	-0,0516	0,0614	-0,0135	0,0159
Tmin	-0,5639	0,2823	-0,5542	0,1306	-0,3473	0,5761	-0,3752	0,7629
τ^2	0,4172	4,2	0,001	3,6526	0,6798	6,5065	0,4831	4,1532
σ^2	0,0003	0,1103	0,0004	0,1418	0,0003	0,053	0,0003	0,0336

Cuadro C.4: Intervalo de predicción para parámetros en modelo Besag-York-Mollie.

	Semana1		Semana2		Semana3		Semana4	
	2,50 %	97,50 %	2,50 %	97,50 %	2,50 %	97,50 %	2,50 %	97,50 %
Intercepto	-12,0738	0,7763	-8,3069	0,0698	-13,0577	-1,7242	-10,5511	1,3661
Altitud	-0,0065	0,0005	-0,0056	0,0002	-0,0048	0,0009	-0,0026	0,003
Pobreza	-0,5902	10,4576	0,1471	7,5891	-0,9812	5,7868	-4,2649	3,5283
EVI	-0,0002	0,0006	-0,0003	0,0003	-0,0004	0,0002	-0,0002	0,0006
Densidad	-0,0004	0,0002	-0,0003	0,0001	-0,0003	0	-0,0003	0,0001
Prec	-0,0281	0,0545	-0,0927	0,5929	-0,4052	0,3894	-0,0412	0,0334
Tmin	-0,1987	0,4298	-0,1004	0,3105	0,0955	0,6695	-0,0791	0,5091
τ^2	0,4086	3,0959	0,0005	1,4712	0,0005	1,264	0,001	1,5564
ρ	0,0354	0,8858	0,0155	0,8825	0,0265	0,8906	0,0065	0,7749
	Semana31		Semana32		Semana33		Semana34	
Intercepto	-4,7869	14,2037	-0,1333	11,611	-14,2085	10,5933	-14,3232	9,4888
Altitud	-0,0063	0,0015	-0,0058	0,0007	-0,0112	0,0032	-0,0065	0,0019
Pobreza	-8,4779	2,0135	-7,0445	-1,4901	-11,7084	1,5242	-7,4387	3,5362
EVI	-0,0004	0,0004	-0,0004	0,0002	-0,0005	0,0008	-0,0002	0,0007
Densidad	0	0,0006	0,0002	0,0004	-0,0001	0,0008	-0,0001	0,0004
Prec	-0,0158	0,0187	-0,0054	0,0247	-0,0538	0,0866	-0,0143	0,0108
Tmin	-0,6084	0,2085	-0,4621	0,0359	-0,4505	0,5884	-0,4237	0,6288
τ^2	0,4752	3,7279	0,0004	1,0762	0,596	6,9425	0,3252	2,9188
ρ	0,1207	0,9487	0,023	0,9008	0,1447	0,9486	0,0349	0,9008

Cuadro C.5: Intervalo de predicción para parámetros en modelo Leroux.

C.2. Intervalos en modelos finales

	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	-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
	Semana 31		**Semana 32**		**Semana 33**		**Semana 34**	
	2,50 %	97,50 %	2,50 %	97,50 %	2,50 %	97,50 %	2,50 %	97,50 %
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

Cuadro C.6: Intervalo de predicción para parámetros en GLM reducido.

	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	-10,958	-0,1066	-8,2644	-0,2217	-11,9997	-2,5991	-10,4046	1,0291
Altitud	-0,0069	0,0003	-0,0056	0	-0,0051	0,0005	-0,0025	0,0031
Pobreza	1,8033	10,78	1,7751	7,8987	1,4983	6,4212	-1,8978	4,6964
Prec	-0,0253	0,0384	-0,0556	0,5281	-0,2624	0,3766	-0,0335	0,0359
Tmin	-0,1331	0,373	-0,1066	0,2566	0,0687	0,5286	-0,0638	0,4727
τ^2	0,2589	1,2819	0,0003	0,6758	0,0004	0,4432	0,0007	0,9145
	Semana 31		Semana 32		Semana 33		Semana 34	
Intercepto	-3,3834	11,6653	0,0615	16,7897	-10,9479	9,5786	-10,6384	7,9238
Altitud	-0,0088	-0,0001	-0,0071	0,0012	-0,0132	0,0007	-0,0063	0,0018
Pobreza	-9,436	-0,0662	-13,3214	-3,1813	-12,5418	0,6829	-7,198	1,3725
Prec	-0,0161	0,0232	-0,0154	0,0286	-0,0259	0,0972	-0,0148	0,0059
Tmin	-0,4168	0,166	-0,5891	0,085	-0,3335	0,4819	-0,2531	0,5619
τ^2	0,2885	1,5697	0,1858	1,7256	0,4374	2,7332	0,0875	1,1343

Cuadro C.7: Intervalos de predicción para parámetros en modelo independiente reducido.

	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	-12,9958	0,829	-10,2673	0,5997	-18,0191	-1,1252	-13,5811	1,2227
Altitud	-0,0059	0,0009	-0,0055	0,0006	-0,0044	0,0015	-0,0021	0,0036
Pobreza	1,9654	11,6581	0,3024	7,6809	-2,3355	5,5141	-3,3861	4,3637
Prec	-0,0174	0,0796	-0,4225	0,8416	-0,4734	0,5758	-0,0287	0,0476
Tmin	-0,2037	0,4405	-0,1341	0,3737	0,0459	0,8927	-0,0532	0,6685
τ^2	0,8001	4,7227	0,0245	2,638	0,0369	2,161	0,0182	3,1624
	Semana 31		**Semana 32**		**Semana 33**		**Semana 34**	
Intercepto	-7,8117	10,1146	-6,9498	15,4692	-15,4082	7,2757	-15,6577	6,8051
Altitud	-0,0067	0,0017	-0,0069	0,0031	-0,0084	0,0025	-0,0057	0,0021
Pobreza	-10,5215	-0,7565	-13,7946	-0,9381	-12,2648	2,3118	-7,1395	2,2426
Prec	-0,0144	0,019	-0,0219	0,0275	-0,0361	0,0794	-0,0156	0,014
Tmin	-0,3697	0,4073	-0,5647	0,4069	-0,296	0,7192	-0,2614	0,7467
τ^2	0,8539	4,491	0,788	6,8856	1,2222	6,9587	0,3543	3,428

Cuadro C.8: Intervalo de predicción para parámetros en modelo intrínseco reducido.

	Semana 1		Semana 2		Semana 3		Semana 4	
	2,50 %	97,50 %	2,50 %	97,50 %	2,50 %	97,50 %	2,50 %	97,50 %
Intercept)	-11,3945	1,3191	-8,5339	0,4018	-13,1138	-1,9793	-10,2749	0,3815
Altitud	-0,0061	0,0005	-0,0054	0,0002	-0,0048	0,0007	-0,0022	0,003
Pobreza	1,417	11,2491	0,9958	7,7803	0,1697	6,3277	-1,5315	4,4043
Prec	-0,0254	0,0539	-0,1327	0,5692	-0,3783	0,4037	-0,03	0,0336
Tmin	-0,2194	0,3794	-0,1271	0,2797	0,0538	0,5925	-0,0211	0,4685
τ^2	0,4515	3,1598	0,0006	1,5152	0,0009	1,2282	0,0004	1,295
ρ	0,0232	0,86	0,0134	0,8777	0,0161	0,8787	0,0087	0,8542
	Semana 31		Semana 32		Semana 33		Semana 34	
Intercepto	-5,3076	10,7033	-2,4809	16,659	-13,2677	7,7223	-12,2611	8,3253
Altitud	-0,007	0,0015	-0,0067	0,0019	-0,0098	0,0026	-0,0059	0,0021
Pobreza	-9,7276	-0,54	-13,7676	-2,3638	-13,9972	2,7232	-6,8376	2,0686
Prec	-0,0162	0,0208	-0,0182	0,0307	-0,0477	0,0852	-0,0149	0,0093
Tmin	-0,3754	0,2885	-0,5636	0,1838	-0,2647	0,5984	-0,2983	0,6051
τ^2	0,645	3,6124	0,3248	3,9454	0,9416	6,2186	0,2126	2,6114
ρ	0,1185	0,9283	0,0154	0,7906	0,1583	0,9437	0,069	0,9125

Cuadro C.9: Intervalos de predicción para parámetros en modelo Leroux reducido.

Apéndice D

Gráficos para riesgo relativo y riesgo estimado por modelo Lee-Mitchel

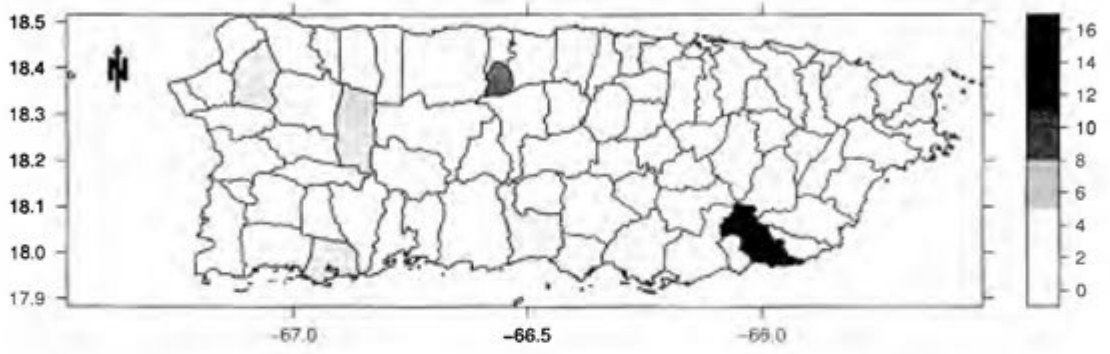


Figura D.1: Riesgo relativo en semana 2.

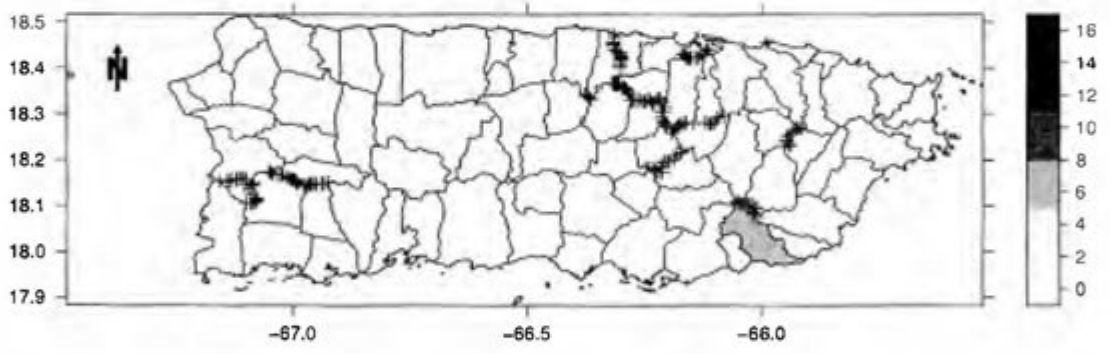


Figura D.2: Riesgo relativo estimado y fronteras en semana 2.

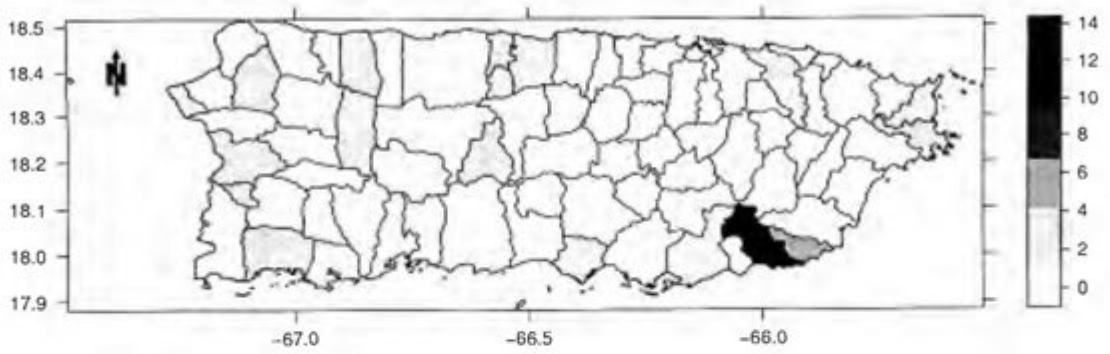


Figura D.3: Riesgo relativo en semana 3.

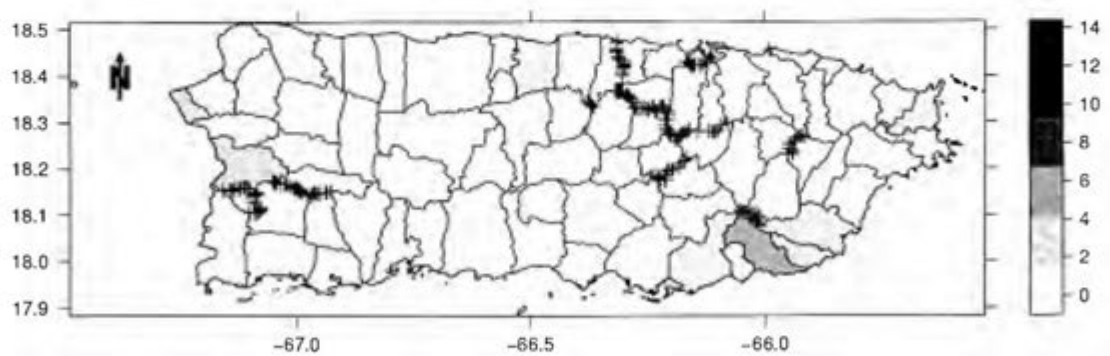


Figura D.4: Riesgo relativo estimado y fronteras en semana 3.

Figura D.5: Riesgo relativo en semana 4.

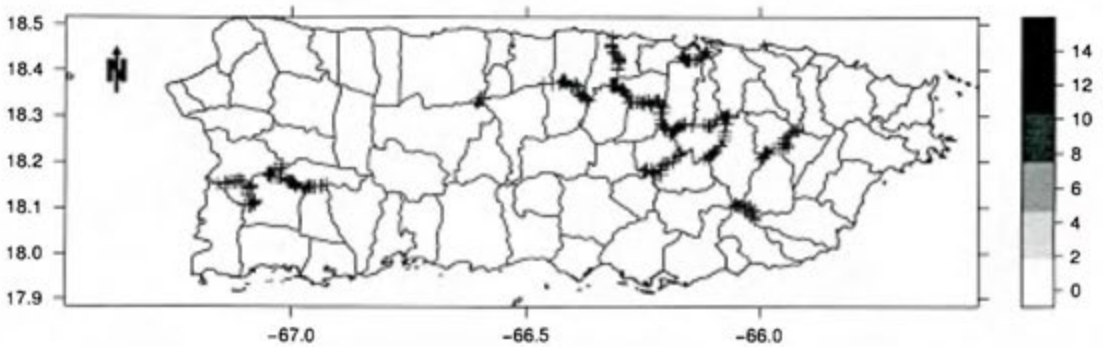


Figura D.6: Riesgo relativo estimado y fronteras en semana 4.

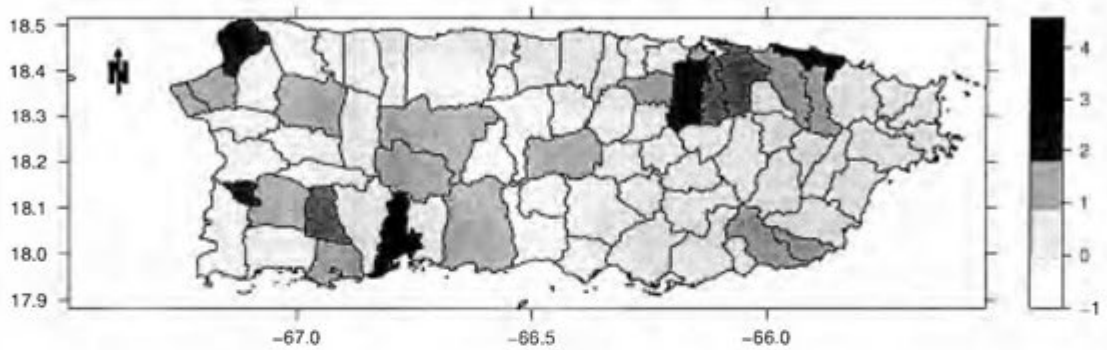


Figura D.7: Riesgo relativo en semana 31.

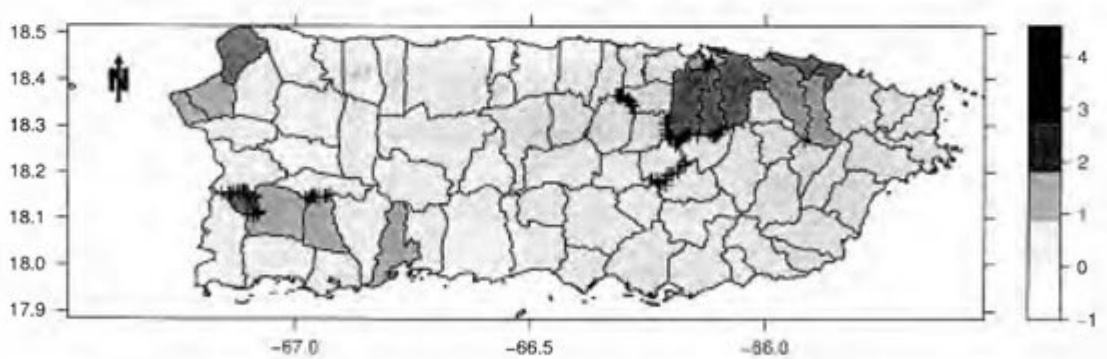


Figura D.8: Riesgo relativo estimado y fronteras en semana 31.

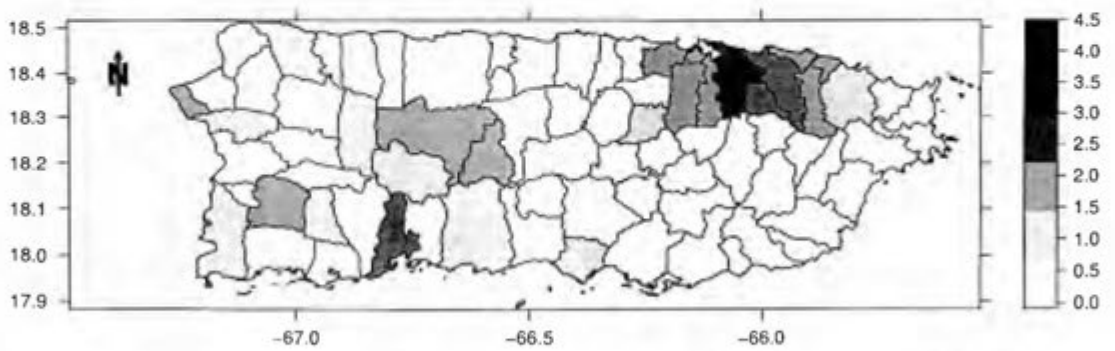


Figura D.9: Riesgo relativo en semana 32.

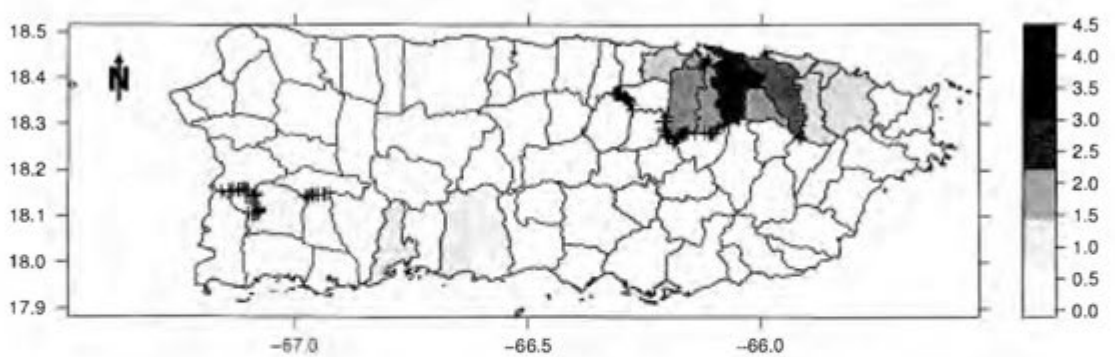


Figura D.10: Riesgo relativo estimado y fronteras en semana 32.

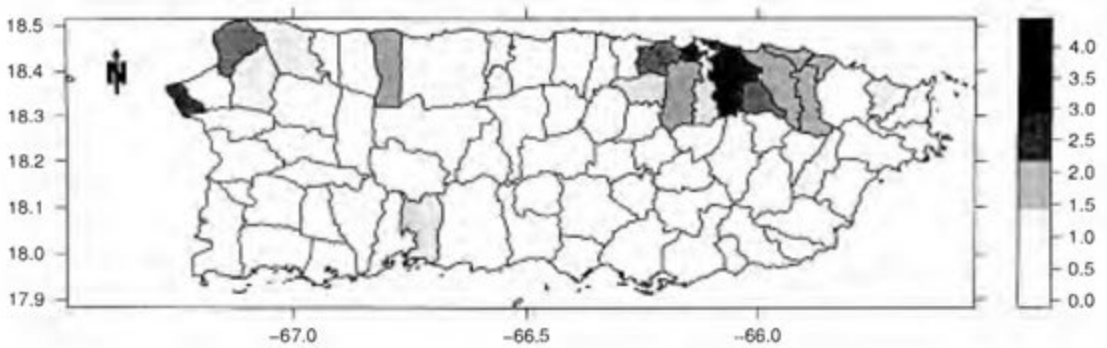


Figura D.11: Riesgo relativo en semana 33.

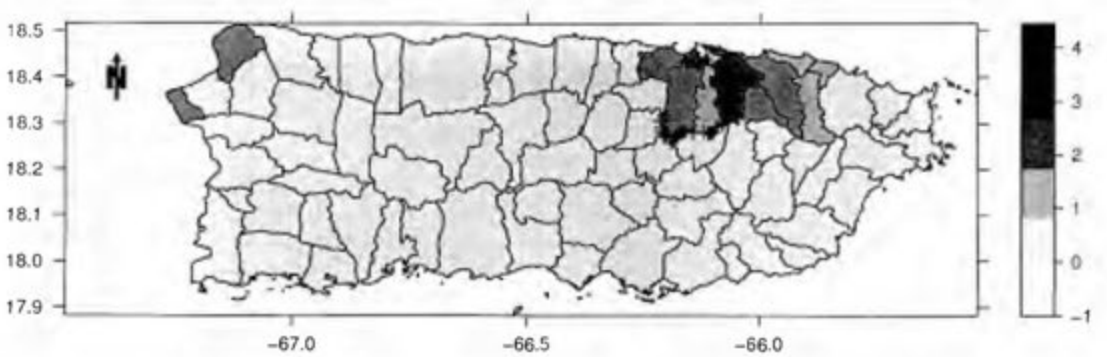


Figura D.12: Riesgo relativo estimado y fronteras en semana 33.

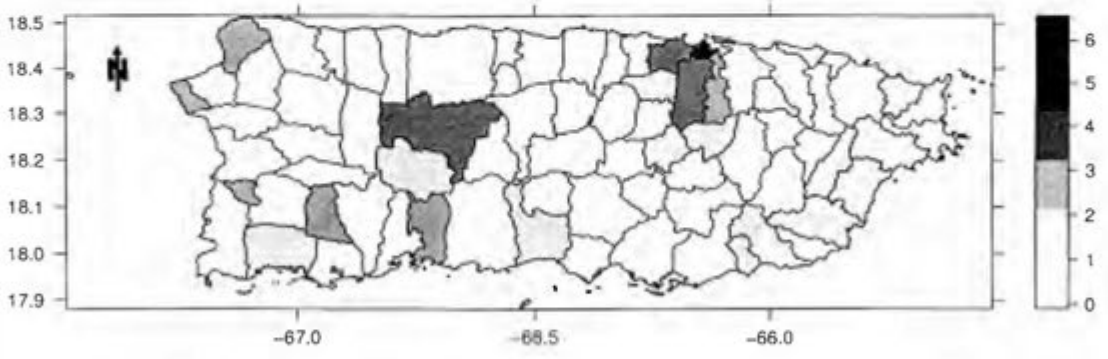


Figura D.13: Riesgo relativo en semana 34.

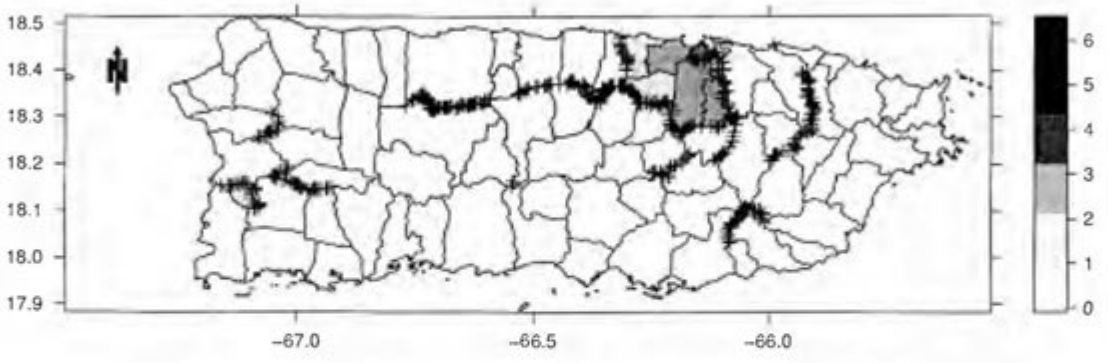


Figura D.14: Riesgo relativo estimado y fronteras en semana 34.

Apéndice E

Residuos de modelo BYM

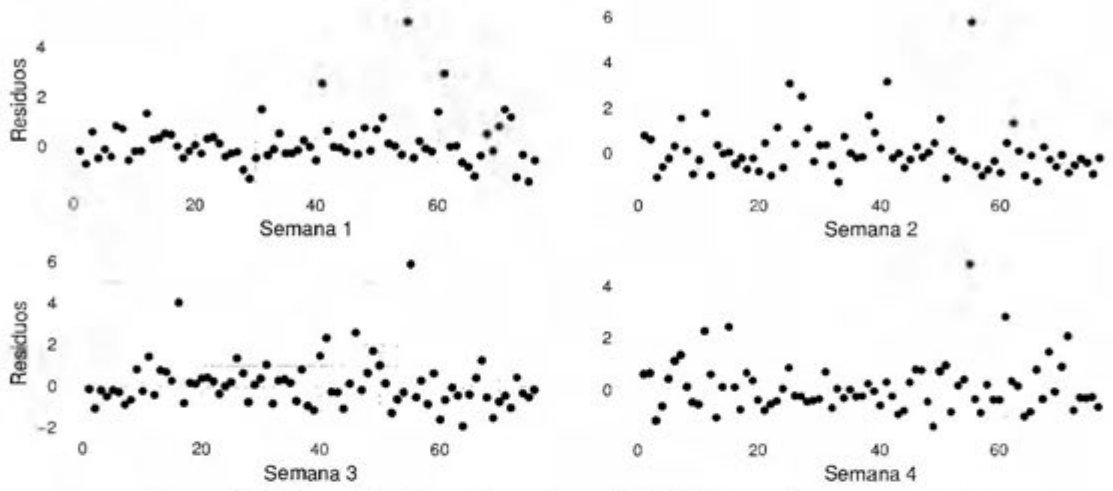


Figura E.1: Dispersión de residuos de modelo BYM para las semanas 1 a 4.

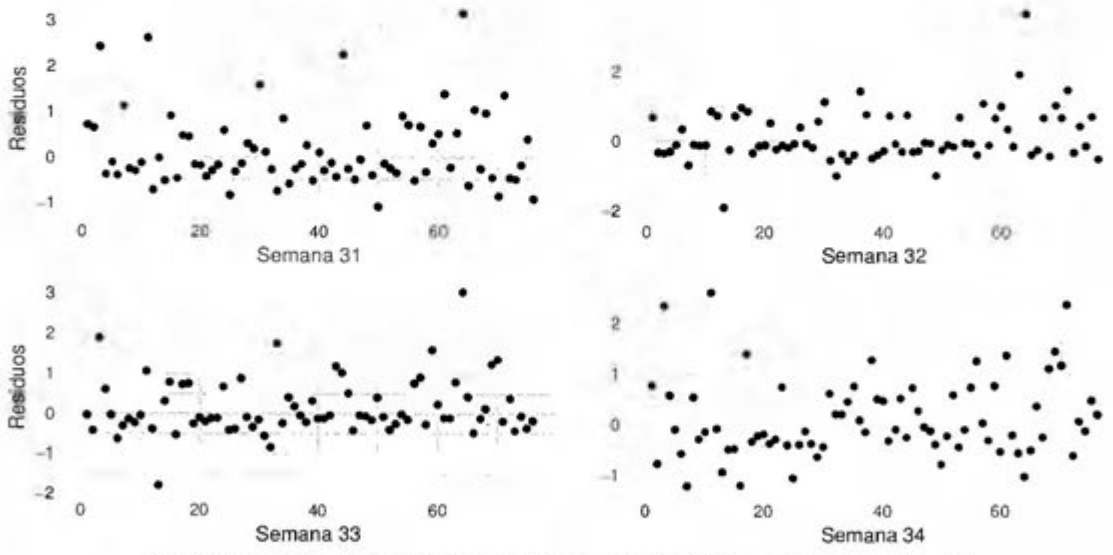


Figura E.2: Dispersión de residuos de modelo BYM para semanas 31 a 34.

Bibliografía

[1] Agresti, Allan: *Categorical Data Analysis.* John Wiley & Sons, New Jersey, 3ª edición, 2013.

[2] Besag, Julian., York, Jeremy. y Mollié, Annie.: *Bayesian image restoration, with two applications in spatial statistics.* Annals of the Institute of Statistical Mathematics, 43: 1–20, 1991.

[3] Bivand, Roger., Hauke, Jan. y Kossowski, Tomasz.: *Computing the Jacobian in Gaussian spatial autoregressive models: An illustrated comparison of available methods.* Geographical Analysis, 45: 150–179, 2013.

[4] Bivand, Roger y Piras, Gianfranco: *Comparing Implementations of Estimation Methods for Spatial Econometrics.* Journal of Statistical Software, 63: 1–36, 2015.

[5] Bivand, Roger S., Pebesma, Edzer. y Gomez-Rubio, Virgilio.: *Applied Spatial Data Analysis with R.* Springer, New York, 2ª edición, 2013.

[6] Bureau, U.S Census: *Datos del Censo 2010 de Puerto Rico,* 2011. `http://www.jp.gobierno.pr/Portal_JP/Default.aspx?tabid=120`, [Web; accedido el 02-02-2015].

[7] Carlin, John B., Gelman, Andrew., Stern, Hal S. y Rubin, Donald B.: *Bayesian Data Analisys.* Chapman & Hall, New York, 2ª edición, 2004.

[8] Carroll, Raymond J., Liang, Faming. y Liu, Chuanhai.: *Advanced Markov Chain Monte Carlo Methods: Learning from Past Samples* John Wiley & Sons Ltd, United Kingdom, 1ª edición, 2010.

[9] Casella, George. y Robert, Christian P.: *Monte Carlo Statistical Methods.* Springer-Verlag, New York, 2ª edición, 2004.

[10] CDC, Subdivisión de Dengue y PR Departamento de Salud: *Informe Semanal de Vigilancia del Dengue,* 2014. `http://www.salud.gov.pr/Estadisticas-Registros-y-Publicaciones/Pages/Dengue.aspx`, [Web; accedido el 15-02-2015].

[11] Chen, Szu Chieh *y cols.*: *Lagged temperature effect with mosquito transmission potential explains dengue variability in southern Taiwan: insights from a statistical analysis.* Science of the total environment, 408: 4069–4075, 2010.

[12] Chien, Lung-Chang. y Yu, Hwa-Lung.: *Impact of meteorological factors on the spatiotemporal patterns of dengue fever incidence.* Environment International, 74: 46–56, 2014.

[13] Chiles, Jean-Paul y Delfiner, Pierre: *Geostatistics: Modeling Spatial Uncertainty.* Wiley, New Jersey, 2ª edición, 2012.

[14] Cressie, Noel A.C.: *Statistics for Spatial Data.* John Wiley & Sons, New York, 1ª edición, 1993.

[15] De Oliveira, Victor y Song, Joon Jin: *Bayesian Analysis of Simultaneous Autoregressive Models.* Sankhyā: The Indian Journal of Statistics, Series B (2008-), 70: 323–350, 2008.

[16] Diggle, Peter J. y Ribeiro, Paulo J.: *Model-based Geostatistics.* Springer Science + Business Media, LLC, 1ª edición, 2007.

[17] Durrett, Richard: *Probability: Theory and Examples.* Cambridge University Press, 4ª edición, 2010.

[18] Gabry, Jonah y Ben Goodrich: *rstanarm: Bayesian Applied Regression Modeling via Stan*, 2016. `https://CRAN.R-project.org/package=rstanarm`, R package version 2.13.1.

[19] Gaetan, Carlo y Guyon, Xavier: *Spatial Statistics and Modeling.* Springer Science+Business Media, LLC, New York, 1ª edición, 2010.

[20] Gelfand, Allan E., Carlin, Bradley P. y Banergee, Sudipto.: *Hierarchical Modeling and Analisys for Spatial Data* Chapman & Hall/ CRC, Boca Raton, Florida, 1ª edición, 2004.

[21] Gelman, Andrew., John B. Carlin, Hal S. Stern, David B. Dunson, Aki. Vehtari y Donald B. Rubin: *Bayesian Data Analysis.* Chapman & Hall/CRC Texts in Statistical Science. Chapman and Hall/CRC, 3ª edición, 2014.

[22] Geweke, John *y cols.*: *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments*, volumen 196. Federal Reserve Bank of Minneapolis, Research Department, Minneapolis, USA, 1992.

[23] Hu, Wenbiao., Clements, Archie., Williams, Gail., Tong, Shilu. y Mengersen, Kerrie.: *Spatial Patterns and Socioecological Drivers of Dengue Fever Transmission in Queensland, Australia.* Environmental Health Perspectives, 120: 260–266, 2012.

[24] Jiang, Zhangyan., Huete, Alfredo R., Didan, Kamel. y Miura, Tomoaki.: *Development of a two-band enhanced vegetation index without a blue band.* Remote Sensing of Environment, 112: 3833–3845, 2008.

[25] Jin, Xiaoping., Carlin, Bradley P. y Banerjee, Sudipto.: *Generalized Hierarchical Multivariate CAR Models for Areal Data* . Biometrics, 2005.

[26] Johansson, Michael A., Dominici, Francesca. y Glass, Gregory E.: *Local and Global Effects of Climate on Dengue Transmission in Puerto Rico* . PLoS Negleted Tropical Diseases, 3, 2009.

[27] Kathryn-Cowles, Mary y Bradley P, Carlin: *Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review.* Journal of the American Statistical Association, 91: 883–904, 1996.

[28] Lee, Duncan: *CARBayes: An R Package for Bayesian Spatial Modeling with Conditional Autoregressive Priors.* Journal of Statistical Software, 55: 1–24, 2013.

[29] Lee, Duncan y Mitchell, Richard: *Boundary detection in disease mapping studies.* Biostatistics, 13: 415–426, 2012.

[30] Leroux, Brian G., Lei, Xingye. y Breslow, Norman.: *Estimation of disease rates in small areas: A new mixed model for spatial dependence.* Institute for Mathematics and Its Applications, 116: 179–191, 2000.

[31] Mena, Nelson., Troyo, Adriana., Bonilla-Carrión, Roger. y Calderón-Arguedas, Ólger.: *Factors associated with incidence of dengue in Costa Rica.* Revista Panamericana de Salud Publica, 29: 234–242, 2011.

[32] Menne, Matthew J., Durre, Imke., Vose, Russell S., Gleason, Byron E. y Houston, Tamara G.: *An Overview of the Global Historical Climatology Network-Daily Database.* Journal of Atmospheric and Oceanic Technology, 29: 897–910, 2012.

[33] Montgomery, Douglas C., Jennings, Cheryl L. y Kulahci, Murat.: *Introduction to time series analysis and forecasting.* John Wiley & Sons, 2015.

[34] Plummer, Martyn., Best, Nicky., Cowles, Kate. y Vines, Karen: *CODA: Convergence Diagnosis and Output Analysis for MCMC.* R News, 6: 7–11, 2006.

[35] R Core Team: *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2016. `https://www.R-project.org/`.

[36] Ribeiro, Paulo J. y Diggle, Peter J.: *geoR: a package for geostatistical analysis.* R-NEWS, 1: 14–18, 2001.

[37] Ripley, Brian D.: *Spatial Statistics and Modeling.* John Wiley & Sons Ltd, New Jersey, 1ª edición, 1981.

[38] Spiegelhalter, David J., Best, Nicola G., Carlin, Bradley P. y Van-Der-Linde, Angelika.: *Bayesian measures of model complexity and fit.* Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64: 583–639, 2002.

[39] Stein, Michael L.: *Interpolation of spatial data: some theory for kriging.* Springer-Verlag, New York, 1ª edición, 1999.

[40] Tuck, Sean y Hellen Phillips: *MODISTools: MODIS Subsetting Tools*, 2015. `http://cran.r-project.org/package=MODISTools`, R package version 0.94.6.

[41] Vehtari, Aki., Gelman, Andrew. y Gabry, Jonah.: *Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC.* Statistics and Computing, páginas 1–20, 2016.

[42] Wall, Melanie M.: *A close look at the spatial structure implied by the CAR and SAR models.* Journal of Statistical Planning and Inference, 121: 311–324, 2004.

[43] Watanabe, Sumio: *Algebraic Geometry and Statistical Learning Theory*, volumen 25. Cambridge University Press, 2009.

[44] Watanabe, Sumio: *Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory.* Journal of Machine Learning Research, 11: 3571–3594, 2010.

[45] Wu, Pei-Chih., Guo, How-Ran., Lung, Shih-Chun., Lin, Chuan-Yao. y Su, Huey-Jen.: *Weather as an effective predictor for occurrence of dengue fever in Taiwan.* Acta tropica, 103: 50–57, 2007.

[46] Zeileis, Achim y Grothendieck, Gabor: *zoo: S3 Infrastructure for Regular and Irregular Time Series.* Journal of Statistical Software, 14: 1–27, 2005.