



Análisis espacial de casos de Dengue basado en la variación de efectos climáticos

Alexia Nicole Bornhorst
John Sebastian Manjarres

Universidad del Valle
Facultad de Ingeniería, Escuela de Estadística
Santiago de Cali, Colombia
2023

Análisis espacial de casos de Dengue basado en la variación de efectos climáticos

Alexia Nicole Bornhorst
John Sebastian Manjarres

Trabajo de grado presentado como requisito parcial para optar al título de:
Estadístico(a)

Director(a):
Ph.D. Alvaro José Flórez Poveda
Codirector(a):
M.Sc. Delia Ortega Lenis

Universidad del Valle
Facultad de Ingeniería, Escuela de Estadística
Santiago de Cali, Colombia
2023

Agradecimientos

Agradecemos a nuestro director Alvaro José Flórez por orientarnos y apoyarnos en el desarrollo de este trabajo. A Delia Ortega por guiarnos y solucionar las dudas presentadas, al profesor José Rafael Tovar por su disponibilidad y asesoría.

A la Universidad del Valle y la escuela de estadística por su formación académica y profesional.

Resumen

El dengue es una enfermedad viral transmitida por mosquitos de la especie *Aedes Aegypti*, con variaciones locales que dependen de parámetros climáticos y de factores sociales y ambientales. En este trabajo se tiene como objetivo evaluar el efecto que tienen las variables climáticas y geográficas sobre la ocurrencia de casos de dengue en los municipios del Valle del Cauca durante el año 2019. Para ello se ajustó un modelo Poisson con efectos aleatorios, en donde se tiene en cuenta la correlación espacial a través de un modelo BYM2, el cual es utilizado para el mapeo de enfermedades en el marco de la estadística bayesiana. En el análisis se encontró que el modelo captura la correlación espacial que está presente entre los municipios y determina que la temperatura, la precipitación, el porcentaje de población urbana y la altitud tienen un impacto significativo en el riesgo de infección de dengue. La zona oriente del departamento presenta un mayor número de casos, en donde se encuentran municipios con temperatura y precipitación alta, un mayor porcentaje de población urbana y una altitud alrededor de los 1000 msnm. Estos municipios presentan un riesgo más elevado de infección respecto al promedio de todo el departamento.

Palabras clave: Mapeo de enfermedades, Dengue, factores climáticos, modelo lineal mixto generalizado, inferencia bayesiana, modelos espaciales, modelo Besag-York-Mollié, INLA

Abstract

Dengue is a viral disease transmitted by *Aedes Aegypti* mosquitoes, with local variations that depend on climatic parameters and social and environmental factors. The objective of this study is to evaluate the effect of climatic and geographic variables on the occurrence of dengue cases in the municipalities of Valle del Cauca during 2019. For this purpose, a Poisson model with random effects was fitted, where the spatial correlation is taken into account through a BYM2 model, which is used for disease mapping in the framework of Bayesian statistics. In the analysis it was found that the model captures the spatial correlation that is present between municipalities and determines that temperature, precipitation, percentage of urban population, and altitude have a significant impact on the risk of dengue infection. The eastern part of the department has a higher number of cases, where there are municipalities with high temperature and precipitation, a higher percentage of urban population and an altitude of around 1000 meters above sea level. These municipalities present a higher risk of infection compared to the average for the entire department.

Keywords: Disease mapping, Dengue, climatic factors, generalized linear mixed model, bayesian inference, spatial models, Besag-York-Mollié model, INLA

Contenido

Resumen	v
Lista de Figuras	ix
Lista de Tablas	1
1. Introducción	2
1.1. Justificación	4
1.2. Pregunta de investigación	5
1.3. Objetivos	5
1.3.1. Objetivo general	5
1.3.2. Objetivos específicos	5
2. Antecedentes	6
3. Marco Teórico	9
3.1. Marco Conceptual	9
3.1.1. Dengue	9
3.1.2. Temperatura	10
3.1.3. Precipitación	10
3.1.4. Humedad Relativa	11
3.1.5. Clima del departamento del Valle del Cauca	11
3.2. Marco Teórico Estadístico	12
3.2.1. Modelo Lineal Generalizado	12
3.2.2. Modelo Poisson	13
3.2.3. Modelo Mixto Lineal Generalizado	13
3.3. Modelación espacial de datos de área	14
3.3.1. Datos de área	15
3.3.2. Modelos Autorregresivos Condicionales (CAR)	17
3.3.3. Modelo Besag	17
3.3.4. Modelo Besag-York-Mollié (BYM)	18
3.3.5. Modelo BYM2	18
3.4. Enfoque Bayesiano	19
3.4.1. Cadenas de Markov Monte Carlo (MCMC)	21

3.4.2. Aproximación de Laplace Anidada Integrada (INLA)	22
4. Metodología	25
4.1. Datos del estudio	25
4.2. Variables de estudio	27
4.2.1. Preparación de la base de datos	28
4.3. Análisis Estadístico	29
4.3.1. Análisis exploratorio	29
4.3.2. Análisis espacial	29
5. Resultados	36
5.1. Análisis exploratorio	36
5.2. Análisis espacial	41
6. Conclusiones, limitaciones y recomendaciones	45
6.1. Conclusiones	45
6.2. Limitaciones	46
6.3. Recomendaciones	46
A. Anexo: Municipios del Valle del Cauca por subregión	48
B. Anexo: Análisis espacial	50
Bibliografía	52

Lista de Figuras

4-1. Ubicación de estaciones meteorológicas	26
4-2. Representación gráfica de la matriz de adyacencia (Los cuadros azules identifican los vecinos. Las filas y columnas identifican los municipios)	32
5-1. Tasa de dengue por 100,000 habitantes por municipio en el año 2019	37
5-2. Factores geográficos. (A) Altitud. (B) Porcentaje de población urbana en los municipios del Valle del Cauca	38
5-3. Variables climáticas. (A) Temperatura máxima. (B) Precipitación. (C) Humedad relativa en los municipios del Valle del Cauca	39
5-4. Correlación entre la tasa de dengue y las variables climáticas y geográficas .	40
5-5. Razón de incidencia estandarizada de dengue en los municipios del Valle del Cauca	41
5-6. Estimación de la distribución a posteriori marginal de los parámetros (A-D) e hiperparámetro (E) del modelo sin la variable humedad (media a posteriori en paréntesis)	43
5-7. (A) Riesgo relativo de dengue para los municipios del Valle del Cauca (B) Probabilidad del riesgo relativo mayor a 1	44
A-1. Municipios del Valle del Cauca por subregión. Fuente: Gobernación del Valle del Cauca	48

Lista de Tablas

4-1.	Descripción de las Estaciones Meteorológicas	26
4-2.	Información recopilada de las bases de datos	27
4-3.	Variables de estudio	28
4-4.	Vecinos de cada municipio del Valle del Cauca (Los números de la columna vecinos corresponden al orden de los municipios)	31
5-1.	Estimación efectos fijos y aleatorios del modelo BYM2	42
5-2.	Estimación de los efectos fijos y aleatorios del modelo BYM2 (Modelo sin la variable humedad)	43
A-1.	Municipios del Valle del Cauca y las estaciones meteorológicas con las que fueron relacionados (Estaciones crecanas en un rango de 50-100 km)	49
B-1.	Tasa de infección de dengue para los municipios del Valle del Cauca	50
B-2.	Resultados del modelo BYM2	51

1. Introducción

El dengue es una enfermedad viral transmitida por mosquitos hembra principalmente de la especie *Aedes aegypti*. En los últimos años esta enfermedad se ha propagado en los trópicos, con variaciones locales que dependen de parámetros climáticos, de factores sociales y ambientales (OMS 2022). Principalmente el dengue se presenta en regiones que tienen un clima tropical, y un clima templado con veranos cálidos y húmedos. Esta enfermedad puede afectar a personas de cualquier edad, pero especialmente a niños y adultos mayores.

Se ha evidenciado que las variables climáticas tienen un efecto en la ocurrencia de casos de Dengue. Esta relación se ha evaluado en estudios de países como Brasil (Lowe et al. 2011), Etiopía (Nigussie et al. 2022) y Tailandia (Phanitchat et al. 2019) en donde se han tenido en cuenta variables como la temperatura, precipitación y humedad relativa. En particular, con la ocurrencia de fenómenos climáticos extremos como el evento El Niño, al afectar las condiciones del clima regional, influyen en la incidencia de la enfermedad en diferentes países de la zona tropical. Este fenómeno climático, cuyo nombre completo es El Niño/Oscilación del Sur (ENSO), se ha considerado como el principal mecanismo modulador de la variabilidad climática global y nacional (Poveda et al. 2001).

Márquez-Benítez et al. (2019) sugieren que la temperatura puede ser más relevante que las demás variables en la transmisión del dengue, ya que afecta diferentes aspectos relacionados directa e indirectamente con la transmisión. Reportan que la temperatura afecta los ciclos de vida de los mosquitos *Aedes* en algunos de sus procesos biológicos como la maduración sexual, la cópula y la oviposición. Sin embargo, esta relación no es lineal, dado que temperaturas muy altas ocasionan una caída dramática en el desarrollo del mosquito, lo que a su vez, puede influenciar la transmisión.

En Colombia, esta enfermedad es endémica desde principios de 1970. Se presentan brotes periódicos en lugares geográficamente por debajo de los 1.800 metros sobre el nivel del mar y el 80 % del territorio nacional puede verse afectado por la presencia de la enfermedad. Debido a las altas tasas de mortalidad que presenta, se considera como la principal enfermedad transmitida por vectores en el país (Ortiz et al. 2013). Los departamentos que históricamente han tenido mayor transmisión de dengue han sido: Atlántico, Santander, Norte de Santander, Valle del Cauca, Tolima, Huila, Casanare, Cundinamarca y Antioquia.

De acuerdo con el Instituto Nacional de Salud, los picos de epidemias de dengue se presentan aproximadamente cada tres años, en donde la última fue en el año 2019. Según el Ministerio de Salud (2021), de acuerdo con la transmisión histórica del dengue y la situación observada durante las epidemias de 2010, 2013, 2016 y 2019 el incremento de los casos en esos periodos se inicio en la región Caribe e Insular y se desplazó hasta el centro oriente, centro, seguido por Orinoquía y Amazonia, terminando en el Pacífico.

La correlación espacial tiene un efecto significativo en la estimación de los factores asociados al incremento de los casos de dengue o cualquier enfermedad en estudios epidemiológicos. La presencia de correlación espacial implica que las observaciones de municipios cercanos tienden a ser más similares entre si que los municipios distantes. La correlación espacial puede introducir sesgo en las estimaciones de los efectos de variables explicativas, distorsionando la variabilidad de los datos y llevando a estimaciones incorrectas. Además, aumenta la incertidumbre en las estimaciones e infla los errores estándar de los coeficientes en el modelo (Diniz-Filho et al. 2003, Segurado et al. 2006). Para abordar adecuadamente la correlación espacial en el análisis epidemiológico, es importante utilizar modelos estadísticos apropiados que incorporen términos espaciales para capturar y ajustar la correlación espacial.

El propósito de este estudio es estimar el riesgo de infección de dengue en el departamento del Valle del Cauca, teniendo en cuenta la variación espacial entre sus 42 municipios. Para ello se propone construir un modelo lineal mixto generalizado, con efectos aleatorios para cuantificar la correlación espacial. Como covariables se consideran características climáticas como temperatura, precipitación y humedad para determinar su efecto sobre el riesgo de infección. De igual forma, se tienen en cuenta características geográficas y sociales de los municipios. En la modelación se prestó especial atención a la correlación espacial entre los municipios, lo que permitirá una mejor comprensión de cómo se propagan los casos de dengue en el departamento.

Este trabajo está constituido por los objetivos propuestos y la revisión de literatura que se realizó, en donde se presentan los antecedentes que abordan este problema de investigación. Luego se presenta el marco teórico, metodología, resultados y conclusiones, limitaciones y recomendaciones. Se establecen las definiciones y los modelos teóricos que se van a desarrollar a lo largo del trabajo, su implementación en el problema, los resultados más relevantes y por último conclusiones sobre los resultados obtenidos.

1.1. Justificación

Según la Organización Mundial de la Salud, el dengue es una enfermedad tropical desatendida. En América Latina, específicamente Colombia está en alerta con respecto a esta arbovirosis ya que se presentó un repunte en el número de casos de dengue reportados a principios de 2019. A pesar de que a nivel mundial se dio una disminución en el número de casos de dengue reportados, Colombia ha mostrado una tendencia creciente en los últimos años (Ordoñez-Sierra et al. 2021).

Por lo tanto, es necesario llevar un seguimiento de la enfermedad para así poder entender su comportamiento. De acuerdo con Filho et al. (2021), se cree que la evaluación de las interacciones temporales de los registros de casos puede revelar pistas relevantes sobre el proceso de propagación de la enfermedad en zonas extensas. Igualmente el análisis espacial es fundamental para entender la dinámica del virus y construir escenarios e identificar patrones de infección por dengue. Una identificación temprana de una epidemia de enfermedad infecciosa es un primer paso importante hacia la implementación de una intervención eficaz para controlar la enfermedad y reducir la mortalidad y la morbilidad en las poblaciones humanas (Kuhn et al. 2005).

Históricamente el departamento del Valle del Cauca es uno de los departamentos que ha tenido mayor transmisión de dengue debido a sus características geográficas, ecológicas y sociales. Por ejemplo, la altitud puede influir en las condiciones climáticas como la temperatura y humedad, que son factores clave para la propagación del mosquito. Las áreas de mayor altitud suelen tener temperaturas menos favorables para la reproducción y supervivencia del mosquito, lo que puede disminuir el riesgo de transmisión en comparación con las áreas de baja altitud. Además, la población urbana también es un factor relevante, debido a que las áreas urbanas muy pobladas pueden proporcionar un entorno propicio para la propagación de dengue. La falta de servicios de saneamiento adecuados, el almacenamiento de agua y la acumulación de residuos pueden aumentar la presencia de mosquitos.

Se considera relevante en este estudio identificar las zonas o municipios que presentan un mayor riesgo de contagio de dengue para implementar una alerta temprana al detectar aumentos de casos de dengue en municipios específicos. Adicionalmente, al identificar estas zonas se pueden realizar esfuerzos de prevención y programas de control de manera más efectiva.

Los datos de área por lo general presentan correlación espacial, esto significa que las observaciones de municipios cercanos geográficamente tienden a ser similares entre sí. Esta correlación espacial debe tenerse en cuenta al analizar los datos, sin embargo los modelos estándar (por ejemplo un modelo lineal generalizado) que asumen independencia pueden

ser inadecuados y podrían llevar a conclusiones erróneas o inválidas. Para solucionar esto, es necesario considerar la inclusión de efectos aleatorios que tienen como objetivo modelar esa correlación, para ello se utiliza un modelo lineal mixto generalizado. Al cuantificar correctamente la correlación espacial entre los municipios, se pueden obtener estimaciones más precisas e inferencias válidas.

Por la complejidad del modelo utilizado y sus características, se prefiere una estimación por métodos bayesianos. Los efectos aleatorios no se estiman fácilmente por máxima verosimilitud pero en el marco del paradigma bayesiano los inconvenientes de estimación que se presentan pueden ser solucionados gracias a su flexibilidad. Por lo tanto, la propuesta metodológica de este proyecto permite la modelación espacial y la visualización de los patrones de infección de dengue.

1.2. Pregunta de investigación

¿Cuál es el efecto de las variables climáticas en la ocurrencia de casos de dengue en los municipios del Valle del Cauca para el año 2019?

1.3. Objetivos

1.3.1. Objetivo general

Evaluar el efecto que tienen las variables climáticas sobre la ocurrencia de casos de dengue en los municipios del Valle del Cauca durante el año 2019.

1.3.2. Objetivos específicos

- Estimar el riesgo de infección de dengue en relación a factores climáticos en los municipios del Valle del Cauca.
- Identificar los factores climáticos y geográficos que influyen en el riesgo de infección de dengue.
- Cuantificar la correlación espacial de los casos de dengue entre los municipios del Valle del Cauca.

2. Antecedentes

Con el fin de conocer las investigaciones que se han realizado previamente sobre la modelación espacial de casos de dengue y su relación con las variables climáticas, se realizó una revisión bibliográfica. A continuación se presentan algunas investigaciones realizadas a nivel nacional e internacional.

Muñoz et al. (2020) en el trabajo titulado **“Spatiotemporal dynamics of dengue in Colombia in relation to the combined effects of local climate and ENSO”**, tienen como objetivo principal caracterizar la distribución temporal y espacial de la incidencia de dengue y su asociación con variables climáticas y El Niño/Oscilación del Sur (ENSO) a diferentes escalas temporales y espaciales en Colombia durante el periodo 2007-2017. El fenómeno natural ENSO se caracteriza por las fluctuaciones en la temperatura de la superficie del mar y la presión atmosférica en el Pacífico ecuatorial. En particular los autores tienen como objetivo confirmar el patrón de viaje en forma de onda de la incidencia de dengue en los diferentes departamentos de Colombia. Además mencionan que eso les ayudará a avanzar en la comprensión de la endemividad estacional y los brotes epidémicos de dengue.

El artículo también analiza la relación entre la incidencia de dengue y la ocurrencia de ambas fases del ENSO, El niño (fase cálida) y La niña (fase fría), a través de las posibles variables climáticas que se ven afectadas por estos fenómenos macro climáticos. Para la parte de estimación, utilizaron herramientas de análisis lineal, incluidas correlaciones cruzadas retrasadas (prueba de Pearson), wavelet cruzada (espectro cruzado de wavelet y coherencia de wavelet), así como un método no lineal de causalidad, que permite identificar impulsores y vínculos causales comunes entre variables simultáneas y retardadas en el tiempo de alta dimensión.

Como resultado encontraron la fuerte asociación de los casos de dengue en Colombia con los índices ENSO, con la temperatura y precipitación local. Además el fenómeno El Niño (La Niña) está relacionado con un aumento (disminución) de casos de dengue a nivel nacional y en la mayoría de las regiones y departamentos. Además mencionan que la influencia de ENSO no es estacionaria ya que hay una reducción de casos de dengue desde 2005 y las variables climáticas locales varían en el espacio y el tiempo, lo que impide extrapolar los resultados de un sitio a otro. Con los análisis que se realizaron por departamentos ellos

concluyen que la relación entre los casos de ONI (El Índice Niño Oceánico) y dengue en cada región está determinada por unos pocos departamentos, Antioquia tiene correlaciones más altas y es seguida por el Valle del Cauca. Estos son los departamentos que presentan un mayor número de casos a nivel histórico.

Ortiz et al. (2013) tenían como objetivo analizar la distribución espacial de los casos y la incidencia de dengue para la ciudad de Medellín. En la investigación que realizaron **“Distribución espacial de casos e incidencia de dengue: Análisis de la situación para Medellín, Colombia”** ubicaron espacialmente los casos de dengue notificados en el periodo 2007 a 2011 a través de las direcciones de las viviendas de los pacientes. Los autores calcularon esas incidencias y las ubicaron geoespacialmente con el software ArcGis 9,3, categorizandolas por niveles de riesgo según la Organización Panamericana de la Salud. Con el conocimiento de la distribución espacial de los casos y la incidencia se definen áreas que requieren mayor esfuerzo en las estrategias de control.

Como resultado de esta investigación encontraron que la enfermedad del dengue se distribuye heterogéneamente en los diferentes barrios de la ciudad. También evidenciaron que en los años 2007 y 2010 se presentaron el mayor número de casos y corresponden a años epidémicos, mientras que los otros años presentaban un comportamiento endémico. Mencionan que en el año 2007 ningún barrio se ubicó en la categoría riesgo muy alto, pero sí 49 localidades para el año 2010. Para los años endémicos según la incidencia de dengue, de los 249 barrios de la ciudad, en los años 2008 y 2009, sólo dos se encontraron en riesgo alto y ninguno para 2011.

Ordoñez-Sierra et al. (2021) en la investigación de la Universidad El Bosque titulada **“Multilevel analysis of social, climatic and entomological factors that influenced dengue occurrence in three municipalities in Colombia”** realizó un análisis multinivel Poisson con efectos mixtos para evaluar factores sociodemográficos, climáticos y entomológicos que pueden influir en la ocurrencia de dengue en tres municipios para el período 2010-2015. Los tres municipios estudiados fueron La Mesa, Anapoima y Apulo, ubicados en la zona suroeste del departamento de Cundinamarca, en donde se han reportado un alto número de casos de dengue.

Según los datos analizados la mayoría de casos tienden a concentrarse en áreas urbanas, los individuos más afectados son los adultos y no se encontraron diferencias estadísticas en el número de casos entre sexos. Las variables que presentaron el mejor ajuste al modelo multinivel Poisson fueron temperatura, humedad relativa, precipitación, y otro tipo de variables como infancia, hecho de vivir en zona urbana y el sistema sanitario contributivo.

Aunque el dengue es una enfermedad multifactorial, este estudio pudo concluir que en la región colombiana estudiada, la variable que mejor explica los casos de dengue es la temperatura. Actualmente se sabe que la temperatura adecuada para la reproducción y supervivencia de este mosquito oscila entre los 22 °C y los 30 °C. La temperatura no sólo afecta a las características biológicas del vector, sino que también permite que el virus sea más contagioso. Los autores añaden que es importante continuar implementando la prevención del dengue mediante programas de control, así como el seguimiento de esta arbovirosis debido a que sigue siendo un problema de salud pública que afecta principalmente a la población en edad de trabajar, impactando en la economía familiar y nacional.

Lowe et al. (2011) en su artículo **“Spatio-temporal modelling of climate-sensitive disease risk: Towards an early warning system for dengue in Brazil”** estudian la posibilidad de utilizar las previsiones climáticas estacionales para desarrollar un sistema de alerta temprana de epidemias de dengue en Brasil. Para ello los autores utilizan un modelo lineal generalizado para seleccionar las variables climáticas. El alcance del estudio es para todo Brasil para el período comprendido entre enero de 2001 y diciembre de 2008 a nivel de microrregión. Las covariables que analizaron fueron temperatura y precipitación con rezagos de tiempo relevantes para la transmisión del dengue, un índice de El Niño Oscilación del Sur y otras variables socioeconómicas y ambientales relevantes.

Los autores ajustan un modelo lineal generalizado Poisson en donde al aumentarse los casos de dengue, la varianza aumenta mucho más. Un aspecto importante a tener en cuenta es que los modelos Poisson son muy usados en el análisis de conteos, pero en el caso de enfermedades por lo general presentan sobre dispersión. Por lo tanto, utilizan un modelo binomial negativo para permitir una sobre dispersión en los casos de dengue observados que son causados por factores de confusión desconocidos o no observados. También se debe tener en cuenta una posible correlación en esos efectos tanto en el tiempo como en el espacio. Es decir, al usar un GLM se asume independencia en tiempo y espacio y esas suposiciones no son válidas. Debido a eso se utiliza un modelo lineal mixto generalizado (GLMM) que incluye efectos aleatorios en el predictor lineal. Este modelo espacio-temporal se implementa a través de un marco bayesiano usando cadenas de Markov Monte Carlo (MCMC).

Los autores concluyen que las covariables climáticas juegan un papel estadísticamente significativo en la transmisión del dengue y añaden que el modelo podría extenderse a otras regiones del mundo donde las enfermedades infecciosas sensibles al clima (por ejemplo, cólera, malaria, leptospirosis, peste) representan una carga para la infraestructura de salud pública, particularmente en los países en desarrollo.

3. Marco Teórico

En este capítulo se presenta el marco teórico de la investigación. La primera parte hace referencia al marco conceptual, se compone de las definiciones sobre la enfermedad del dengue y las variables climáticas que se utilizan en la investigación. Luego se presenta el marco teórico estadístico el cual está relacionado con la metodología que se emplea en el desarrollo de esta investigación, esta parte abarca los conceptos relacionados a modelos lineales mixtos generalizados, modelos de áreas y estadística bayesiana.

3.1. Marco Conceptual

3.1.1. Dengue

El dengue es una enfermedad vírica que se transmite de los mosquitos a las personas. En las últimas décadas ha aumentado la incidencia del dengue en el mundo, es más frecuente en las regiones de climas tropicales y subtropicales. El virus del dengue se transmite a los seres humanos a través de la picadura de mosquitos hembra infectados, principalmente de la especie *Aedes aegypti* y, en menor grado, de la especie *Aedes albopictus*. Estos mosquitos también son vectores de los virus de la fiebre chikungunya, la fiebre amarilla y el Zika. La enfermedad está muy extendida en los trópicos, con variaciones locales en el riesgo que dependen de los parámetros climáticos y de los factores sociales y ambientales (OMS 2022).

Es una enfermedad viral aguda que puede afectar a personas de cualquier edad, especialmente niños y adultos mayores. Los mosquitos del dengue se presentan en zonas urbanas con altitudes inferiores a 2200 metros sobre el nivel del mar, ponen sus huevos en depósitos de agua limpia como albercas, floreros de plantas acuáticas, llantas, baldes de agua y cualquier recipiente que está a la intemperie y que puede almacenar agua (Ministerio de Salud 2021). La urbanización está asociada a la transmisión del dengue en función de varios factores sociales y ambientales: densidad de población, desplazamientos de personas, acceso a fuentes de agua fiables, prácticas de almacenamiento de agua, entre otras. Los riesgos de contraer dengue también dependen de los conocimientos y las prácticas de la población frente a la enfermedad, así como de la ejecución de actividades sistemáticas de control vectorial de forma sostenible fuera de los establecimientos de salud. Por lo tanto, los riesgos pueden cambiar y extenderse geográficamente a causa del cambio climático en las zonas tropicales y subtropicales, y los vectores pueden adaptarse a nuevos lugares y climas (OMS 2022).

3.1.2. Temperatura

La temperatura atmosférica se refiere al nivel de calor que tiene el aire en un sitio determinado y en un momento específico. Este valor forma parte de los elementos que componen el clima, al igual que las precipitaciones, la humedad, el viento y la presión. La temperatura se mide con el termómetro de mercurio, instalado de tal forma que adopte la temperatura del aire, por lo que es necesario protegerlo de la radiación, tanto solar como de la reflejada por el suelo (Pérez Porto & Gardey 2014).

■ Temperatura máxima

La temperatura máxima media del aire (TMAX), corresponde al promedio aritmético de los valores de temperatura máxima registrados en el lapso de tiempo i , en una estación j , durante el período de tiempo t . La temperatura máxima es la mayor temperatura registrada durante un período de tiempo dado. Es registrada mediante el uso de termómetros de mercurio, en los que una contracción en el tubo capilar, sólo permite el ascenso de la columna mercurial (DANE 2005).

■ Temperatura mínima

La temperatura mínima media del aire (TMIN), corresponde al promedio aritmético de los valores de temperatura mínima registrados en el lapso de tiempo i , en una estación j , durante el período de tiempo t . Es el valor más bajo de la temperatura del aire registrada durante un período de observación (cada 24 horas). La temperatura mínima es uno de los elementos más importantes del tiempo, que ejerce gran influencia en la vida humana, en la de los animales y de las plantas. Es un elemento determinante de las condiciones de vida y productividad en las diversas regiones del país. La temperatura mínima alcanzada en el día, es registrada diariamente, mediante el uso de termómetros de alcohol, en los que un menisco en el tubo capilar es desplazado hasta alcanzar el valor mínimo (DANE 2005).

3.1.3. Precipitación

La precipitación corresponde a la cantidad de agua que llega al suelo en forma de lluvia, rocío, granizo o nieve; es contabilizada durante i períodos, en la estación j durante el tiempo t . La precipitación ocurre cuando el vapor de agua se condensa en el aire y cae como líquido o sólido a la superficie del suelo. La cantidad de precipitación se mide con un pluviómetro, dispositivo que cuenta con una probeta graduada (la cantidad se expresa en l/m^2 o en milímetros de altura, que representan la altura de la capa de agua caída), o con el pluviógrafo, cuya medición es por medio de una banda registradora (DANE 2005).

3.1.4. Humedad Relativa

La humedad relativa media mensual del aire (HR_m) del mes *h*, registrada en una estación de monitoreo *j*, corresponde al promedio aritmético de los valores de humedad relativa diaria del aire de los días *i* del mismo mes. La humedad relativa es el porcentaje de humedad que contiene el aire con respecto al total que es capaz de contener como función de su temperatura y su presión. En otras palabras, humedad relativa es la relación porcentual entre la cantidad de vapor de agua que tiene el aire y el máximo que podría contener a una temperatura y presión determinada. El registro de la humedad relativa diaria se realiza con los instrumentos de medición: psicrómetro (fórmula psicrométrica) y el Hidrógrafo que van en valores de 0 al 100 % (DANE 2005). Los niveles de humedad relativa para la comodidad humana se encuentran típicamente entre 30-70 %. Valores superiores al 70 % se consideran húmedos. Los valores inferiores al 30 % se consideran secos.

3.1.5. Clima del departamento del Valle del Cauca

El Valle del Cauca es uno de los 32 departamentos de Colombia situado en el sur occidente del país, entre la Región Andina y la Región Pacífica. Gran parte del departamento está en el valle geográfico del Río Cauca, entre las cordilleras occidental y central. La llanura del Pacífico comprendida entre el nivel del mar y los 1.000 msnm, es considerada como una de las zonas más lluviosas, no solo del país sino del mundo, con una precipitación media entre 5.000 y 6.500 mm/año. Es una región con dos épocas lluviosas y dos secas al año. La primera época seca entre diciembre y febrero, la primera época lluviosa va de marzo a mayo, la segunda época seca de junio a septiembre y la segunda época lluviosa de octubre a noviembre (IDEAM 2018). Sobre la franja del litoral Pacífico, no existe una temporada seca definida y prácticamente llueve durante todo el año. En algunos sitios, se observa una disminución de lluvias hacia el inicio del año.

En el departamento del Valle del Cauca el régimen térmico presenta amplios rangos de variación. Hacia el fondo del valle geográfico y sobre la franja costera del Pacífico, las temperaturas medias anuales alcanzan valores entre los 24 y los 28 °C, siendo las mayores del departamento. Sobre las estribaciones cordilleranas, las temperaturas disminuyen linealmente en función de la elevación, dando lugar a valores extremadamente bajos en los picos más altos de la cordillera central, al extremo oriental del departamento. La humedad relativa fluctúa en el rango 65 %-75 %.

El clima de la mayor parte del valle geográfico del río Cauca, oscila entre cálido semi húmedo y templado semihúmedo. Al extremo oriental del departamento y también al centro, en las estribaciones de la cordillera occidental, se distribuyen los climas fríos. Sobre la franja litoral, al occidente, el clima es cálido y superhúmedo (IDEAM 2018).

3.2. Marco Teórico Estadístico

En esta sección se presenta la teoría estadística relacionada con el modelo lineal mixto generalizado, modelos de áreas y modelos autorregresivos condicionales. Luego se presentan definiciones sobre el enfoque bayesiano y la aproximación de Laplace anidada integrada (INLA).

3.2.1. Modelo Lineal Generalizado

Un modelo lineal generalizado (GLM) es una unificación de modelos de regresión lineal y no lineal que también permite la incorporación de distribuciones de respuesta no normales. Es decir, son una clase de modelos que permite modelar variables aleatorias con distribución de probabilidad diferentes a la normal. En un GLM, la distribución de la variable de respuesta solo debe ser miembro de la familia exponencial, que incluye las distribuciones normal, poisson, binomial, exponencial y gamma (Montgomery et al. 2012).

Un GLM está definido por tres componentes:

- **Componente aleatorio**

Variable respuesta Y y su distribución de probabilidad. Se asume que se cuenta con n observaciones independientes $(y_i, \dots, y_n)$ de una variable aleatoria Y cuya distribución de probabilidad pertenece a la familia exponencial.

- **Predictor lineal**

Para cada observación i , se tiene un conjunto de covariables observadas. Esto es:

$$\mathbf{x}_i = (1, x_{i1}, \dots, x_{i,p-1})' \quad (3-1)$$

donde x_{ij} es la j -ésima covariable asociada al individuo i . El predictor lineal está definido como:

$$\eta_i = \beta_0 + \sum_{j=1}^{p-1} \beta_j x_{ij} = \mathbf{x}_i' \boldsymbol{\beta} \quad (3-2)$$

- **Función de enlace**

La función de enlace conecta μ_i con el predictor lineal η_i a través de una función $g(\cdot)$:

$$\mu_i = g^{-1}(\eta_i) = g^{-1}(\mathbf{x}_i' \boldsymbol{\beta}) \quad (3-3)$$

donde la función $g(\cdot)$ es monótona y diferenciable. Generalmente, $g(\cdot)$ esta determinada por la distribución que se asume para $\mathbf{Y}$.

3.2.2. Modelo Poisson

Se supone que la variable de respuesta y_i es un conteo, tal que la observación $y_i = 0, 1, \dots$. Un modelo de probabilidad razonable para los datos de conteo suele ser la distribución de Poisson, cuya distribución de probabilidad es:

$$f(y; \mu) = \frac{e^{-\mu} \mu^y}{y!}, \quad y = 0, 1, \dots \quad (3-4)$$

donde $\mu > 0$. Para la distribución de Poisson se tiene que $E(y) = Var(y) = \mu$ (Montgomery et al. 2012). Si se considera que la media depende de un conjunto de covariables x_i , se obtiene el modelo Poisson:

$$y_i \sim \text{Poisson}(\mu_i), \quad i = 1, \dots, n, \quad \text{donde} \quad \mu_i = \exp(\mathbf{x}_i' \boldsymbol{\beta}) \quad (3-5)$$

El modelo Poisson presenta algunas características importantes que se deben tener en cuenta. En primer lugar, se puede utilizar como una aproximación de la binomial cuando el tamaño de muestra n_i es grande y la probabilidad de éxito π_i es pequeña. En este caso, la binomial converge hacia una Poisson con $\mu_i = n_i \pi_i$ (Agresti 2015).

Los conteos y_i en el modelo Poisson son proporcionales a un índice t_i (intervalo de tiempo, área de espacio, tamaño de población). Cuando t_i no es fijo, se debe incluir un término llamado offset. Además, en la práctica es común encontrar que la varianza es mayor que la media, lo que se conoce como sobredispersión. Esto puede indicar la presencia de factores no considerados en el modelo que contribuyen a la variabilidad observada. Por último, es importante mencionar que la frecuencia de ceros puede ser mayor de lo esperado en el modelo ajustado, a esto se le conoce como exceso de ceros. Para mayor detalle sobre estos temas, ver Agresti (2015) capítulo 7.

3.2.3. Modelo Mixto Lineal Generalizado

Un modelo mixto lineal generalizado (GLMM) es una extensión del GLM que contiene efectos fijos y aleatorios en el predictor lineal. Es importante tener en cuenta que el predictor lineal describe la relación lineal entre una variable respuesta y una o más variables predictoras. Los GLMM también heredan de los GLM la idea de extender modelos mixtos lineales a datos no normales. Los efectos fijos se ven como constantes en la población, mientras que los efectos aleatorios se consideran estocásticos o variables (Anderson et al. 2013).

Los GLMM proporcionan una amplia gama de modelos para el análisis de datos agrupados, ya que las diferencias entre grupos se pueden modelar como un efecto aleatorio. Estos modelos son útiles cuando la variable dependiente es binaria, ordinal, de conteo o

cuantitativa. También son útiles cuando se tienen datos longitudinales y cuando la variable dependiente implica medidas repetidas, ya que los GLMM pueden modelar las correlaciones.

Se tiene que Y_{ij} es el j -ésimo resultado medido para el grupo o sujeto i , $i = 1, \dots, N$, $j = 1, \dots, n_i$ y $\mathbf{Y}_i = (y_{i1}, \dots, y_{in_i})$ es el vector n_i -dimensional de todas las medidas disponibles para el grupo i . Se asume que, condicionalmente en q -dimensiones los efectos aleatorios $\mathbf{b}_i$ se distribuyen independientemente como $N(\mathbf{0}, \mathbf{D})$, los resultados Y_{ij} son independientes con densidades de la siguiente forma

$$f_i(y_{ij}|\mathbf{b}_i, \boldsymbol{\beta}, \phi) = \exp \{ \phi^{-1} [y_{ij}\theta_{ij} - \psi(\theta_{ij})] + c(y_{ij}, \phi) \} \quad (3-6)$$

con $\eta(\mu_{ij}) = \eta[E(\mathbf{Y}_{ij}|\mathbf{b}_i)] = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i$ para una función de enlace conocida $\eta(\cdot)$, con $\mathbf{x}_{ij}$ y $\mathbf{z}_{ij}$ vectores p -dimensional y q -dimensional de covariables conocidas, con $\boldsymbol{\beta}$ un vector p -dimensional de coeficientes de regresión fijos desconocidos y con ϕ un parámetro de escala. Finalmente, sea $f(\mathbf{b}_i|\mathbf{D})$ la densidad de la distribución $N(\mathbf{0}, \mathbf{D})$ para los efectos aleatorios $\mathbf{b}_i$ (Molenberghs & Verbeke 2005).

La distribución marginal de $\mathbf{Y}_i$ está dada por:

$$f_i(y_{ij}|\boldsymbol{\beta}, \mathbf{D}, \phi) = \int f_{ij}(y_{ij}|\mathbf{b}_i, \boldsymbol{\beta}, \phi) f(\mathbf{b}_i|\mathbf{D}) d\mathbf{b}_i \quad (3-7)$$

La función de verosimilitud para el vector de parámetros $\boldsymbol{\Theta} = (\boldsymbol{\beta}, \mathbf{D}, \phi)'$ se puede escribir como:

$$L(\boldsymbol{\Theta}) = \prod_{i=1}^N \int \prod_{j=1}^{n_i} f_{ij}(y_{ij}|\mathbf{b}_i, \boldsymbol{\beta}, \phi) f(\mathbf{b}_i|\mathbf{D}) d\mathbf{b}_i \quad (3-8)$$

3.3. Modelación espacial de datos de área

En muchas aplicaciones, se analizan datos que están referenciados geográficamente, por ejemplo, se conoce su ubicación en el espacio. Estos también se denominan datos espaciales y se definen como realizaciones de un proceso estocástico indexado por el espacio

$$Y(s) \equiv \{y(s), s \in \mathcal{D}\}$$

donde $\mathcal{D}$ es un subconjunto (fijo) de $\mathbb{R}^d$

3.3.1. Datos de área

Los datos de área o lattice, en donde $y(s)$ es un valor agregado aleatorio sobre una unidad de área s con límites bien definidos en $\mathcal{D}$, que se define como una colección contable de unidades espaciales d -dimensionales. La diferencia entre área y lattice es que la primera suele ser irregular y se basa en límites administrativos (código postal, condado, departamento, municipio, etc.), mientras que el último es regular (Banerjee et al. 2014).

Cuando se trabaja con datos a nivel de área, la dependencia espacial se tiene en cuenta a través de la estructura de vecindad. El (i, j) -ésimo elemento de una matriz de vecindad espacial W , denotada por w_{ij} , conecta espacialmente las áreas i y j de alguna manera, $i, j \in 1, \dots, n$. W define una estructura de vecindad sobre toda la región de estudio, y sus elementos pueden verse como pesos. Se asocia más peso a los j más cercanos a i que a los más alejados de i . La definición de vecindad más simple la proporciona la matriz binaria donde $w_{ij} = 1$ si las regiones i y j comparten algún límite común, tal vez un vértice, y $w_{ij} = 0$ en caso contrario. Habitualmente, w_{ii} se establece en 0 para $i = 1, \dots, n$. Esta elección de definición de vecindad da como resultado una matriz de vecindad espacial simétrica (Moraga 2019).

Matriz de adyacencia

La matriz de adyacencia o de proximidad es útil para una exploración inicial de los datos de área. Sea Y_i las observaciones de la i -ésima área, $i = 1, \dots, n$, los elementos w_{ij} de $\mathbf{W}$ conectan espacialmente las áreas i y j

$$\mathbf{W} = \begin{pmatrix} w_{11} & w_{12} & \dots & w_{1n} \\ w_{21} & w_{22} & \dots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \dots & w_{nn} \end{pmatrix}$$

Se pueden considerar varias posibilidades para definir los elementos w_{ij} . En este caso se utiliza una opción binaria, en donde $w_{ij} = 1$ si las áreas i y j comparten un límite en común y se establece $w_{ii} = 1$. En general $\mathbf{W}$ es simétrica y los elementos de $\mathbf{W}$ pueden ser vistos como pesos. Los elementos w_{ij} pueden ser estandarizados por $\sum_j w_{ij} = w_{j+}$. Si $\widetilde{\mathbf{W}}$ tiene los elementos $\widetilde{w}_{ij} = w_{ij}/w_{j+}$, entonces $\widetilde{\mathbf{W}}$ es estocástica por fila, es decir, $\widetilde{\mathbf{W}}\mathbf{1} = \mathbf{1}$, pero ahora $\widetilde{\mathbf{W}}$ no necesita ser simétrica. En este contexto exploratorio $\mathbf{W}$ proporciona el mecanismo para introducir la estructura espacial (Banerjee et al. 2014).

Uno de los problemas con esta definición de la matriz $\mathbf{W}$ es que los bordes en común de las áreas i y j varían en longitud. Algunas áreas solo están conectadas por un borde corto,

mientras que otras comparten un borde extenso.

Otras opciones, especialmente para áreas irregulares, para definir $\mathbf{W}$ son:

- $w_{ij} = \|A_i - A_j\|^{-\gamma}$
- $w_{ij} = \exp(\|A_i - A_j\|^{-\gamma})$
- $w_{ij} = (l_{ij}/l_i)^\gamma$
- $w_{ij} = (l_{ij}/l_i) / \|A_i - A_j\|^{-\gamma}$

Donde $\gamma \geq 0$. $\|A_i - A_j\|$ define la separación espacial entre áreas. Cuanto más cercanas en el espacio sean las áreas A_i y A_j , mayor será el factor de ponderación w . Para las dos últimas definiciones, l_{ij} es la longitud del borde en común entre el área A_i y A_j , y l_i es el perímetro del área A_i . De esta manera se define la asociación espacial de las áreas como una función de distancias físicas entre cada área. Dado que la distancia entre regiones no se puede definir de una única manera, el centro o cualquier otro punto significativo puede ser utilizado (Zuur et al. 2007).

Razón de incidencia estandarizada (SIR)

Para cada área i , $i = 1, \dots, n$, el SIR está definido como la razón entre el número de casos observados Y_i y el número de casos esperados E_i

$$SIR_i = Y_i / E_i \quad (3-9)$$

Los casos esperados E_i representan el número total de casos que uno esperaría si la población del área i se comportara de la misma forma que la población estándar (o regional). E_i se puede calcular usando estandarización indirecta como

$$E_i = r^{(s)} n^{(i)}, \quad (3-10)$$

donde $r^{(s)}$ es el ratio de la población estándar (el número total de casos dividido por la población de todas las áreas), y $n^{(i)}$ es la población en el área i . El SIR_i indica si un área i tiene mayor ($SIR_i > 1$), igual ($SIR_i = 1$) o menor ($SIR_i < 1$) riesgo que el esperado por la población estándar (Moraga 2019).

Al identificar una población estándar, comúnmente se selecciona una de las siguientes posibilidades:

- Una de las poblaciones de estudio

- La entidad de los grupos de población de estudio combinados
- Una población “externa”, es decir, la población de un área local o país
- Una población hipotética

La elección es algo arbitraria y no hay un estándar correcto o incorrecto. Pero debe tenerse en cuenta que la elección tiene un efecto sobre los ratios estandarizados. Por lo tanto, el estándar debe ser preferiblemente representativo de las poblaciones de estudio que se comparan (Van den Broeck et al. 2013).

3.3.2. Modelos Autorregresivos Condicionales (CAR)

Considerando n áreas, cada una caracterizada por un conjunto de vecinos $\mathcal{N}(i)$, supongamos que u_i es la siguiente variable aleatoria:

$$u_i | \mathbf{u}_{-i} \sim \text{Normal} \left(\mu_i + \sum_{j=1}^n r_{ij} (u_j - \mu_j), s_i^2 \right),$$

donde $\mathbf{u}_{-i} = \{u_1, \dots, u_{i-1}, u_{i+1}, \dots, u_n\}$, μ_i es la media para el área i y $s_i^2 = \sigma_u^2 / \mathcal{N}_i$ es la varianza para la misma área, siendo $\mathcal{N}_i = \#\mathcal{N}(i)$ el número de vecino del área i . Si un área tiene muchos vecinos entonces su varianza será pequeña. La cantidad r_{ij} indica la proximidad espacial y puede ser calculado como $\phi \times W_{ij}$, donde $W_{ij} = a_{ij} / \mathcal{N}_i$, a_{ij} es 1 si el área i y j son vecinos y 0 de lo contrario, el parámetro ϕ controla que tan adecuada es la distribución.

Si se tiene $\phi = 1$, $\mu_i = 0$ para cada i y se fija una restricción tal que $\sum_{i=1}^n u_i = 0$:

$$u_i | \mathbf{u}_{-i} \sim \text{Normal} \left(\frac{1}{\mathcal{N}_i} \sum_{j=1}^n a_{ij} u_j, s_i^2 \right) \quad (3-11)$$

La distribución condicional ahora tiene la forma típicamente presentada en la literatura sobre mapeo de enfermedades (Blangiardo & Cameletti 2015).

3.3.3. Modelo Besag

Riebler et al. (2016) mencionan que uno de los enfoques más populares para modelar la correlación espacial es un modelo intrínseco de campo aleatorio Gaussiano de Markov (GMRF), denominado modelo Besag. Una suposición común es considerar las áreas i y j como vecinos si comparten una frontera común, denotada aquí como $i \sim j$. La distribución condicional para u_i es como (3-11). La distribución conjunta para $\mathbf{u}$ es:

$$f(\mathbf{u}|\tau_u) \propto \exp\left(-\frac{\tau_u}{2} \sum_{i \sim j} (u_i - u_j)^2\right) \propto \exp\left(-\frac{\tau_u}{2} \mathbf{u}' \mathbf{Q} \mathbf{u}\right)$$

donde la matriz de precisión $\mathbf{Q}$ es:

$$Q_{ij} = \begin{cases} \mathcal{N}_i & i = j \\ -1 & i \sim j \\ 0 & \text{resto.} \end{cases}$$

El modelo es intrínseco en el sentido de que $\mathbf{Q}$ es singular, es decir tiene un espacio nulo $\mathbf{V}$ no vacío. La deficiencia de rango del modelo Besag es igual al número de subgrafos conectados. La deficiencia de rango es igual a uno si todas las regiones están conectadas. Por lo tanto, la densidad del modelo Besag es

$$f(\mathbf{u}|\tau_u) = K \tau_u^{(n-I)/2} \exp\left(-\frac{\tau_u}{2} \mathbf{u}' \mathbf{Q} \mathbf{u}\right),$$

donde I denota el número subgrafos conectados y K es una constante.

3.3.4. Modelo Besag-York-Mollié (BYM)

Riebler et al. (2016) menciona que el modelo de Besag solo asume un componente espacialmente estructurado y no puede tomar la forma limitante que permite la variabilidad espacialmente no estructurada. Por lo tanto, el error aleatorio no estructurado o la sobredispersión pura dentro del área i se modelarán como correlación espacial, dando estimaciones erróneas de los parámetros.

Para resolver este problema, el modelo BYM descompone el efecto espacial regional $\mathbf{b}$ en una suma de un componente espacial estructurado y no estructurado, de modo que $\mathbf{b} = \mathbf{v} + \mathbf{u}$. Aquí $\mathbf{v} \sim N(\mathbf{0}, \tau_v^{-1} \mathbf{I})$ representa la sobredispersión pura, mientras $\mathbf{u} \sim N(\mathbf{0}, \tau_u^{-1} \mathbf{Q}^-)$ es el modelo Besag, donde $\mathbf{Q}^-$ denota la inversa generalizada de $\mathbf{Q}$. La matriz de covarianza resultante de $\mathbf{b}$ es

$$\text{Var}(\mathbf{b}|\tau_u, \tau_v) = \tau_v^{-1} \mathbf{I} + \tau_u^{-1} \mathbf{Q}^-.$$

3.3.5. Modelo BYM2

Simpson et al. (2015) proponen una modificación del modelo que aborda tanto el problema de la identificabilidad como el del escalado del modelo BYM. El modelo utiliza un componente

estructurado escalado $\mathbf{u}_*$, donde $\mathbf{Q}_*$ es la matriz de precisión del modelo de Besag escalada. Esto da una versión modificada del efecto aleatorio

$$\mathbf{b} = \frac{1}{\sqrt{\tau_b}}(\sqrt{1-\phi}\mathbf{v} + \sqrt{\phi}\mathbf{u}_*),$$

con una matriz de covarianza

$$\text{Var}(\mathbf{b}|\tau_b, \phi) = \tau_b^{-1}((1-\phi)\mathbf{I} + \phi\mathbf{Q}_*^{-1}),$$

donde $\mathbf{u}_* \sim N(\mathbf{0}, \mathbf{Q}_*^{-1})$, $\mathbf{v} \sim N(\mathbf{0}, \mathbf{I})$ y $0 \leq \phi \leq 1$ mide la proporción de la varianza marginal explicada por el efecto estructurado.

Usando esta parametrización $\text{Var}(\mathbf{b}|\tau_b, \phi)$ tiene una clara e intuitiva interpretación, y las interpretaciones en términos de la distribución condicional de b_i dado $b_j, j \in \mathcal{N}_i$, son evitadas. Se tiene una relación entre la sobredispersión pura y la correlación espacial estructurada. El modelo se reduce a pura sobredispersión para $\phi = 0$ y solo riesgo espacialmente estructurado cuando $\phi = 1$ (Riebler et al. 2016).

En el contexto de los modelos de datos de áreas, es importante destacar que en el modelo BYM el componente espacialmente estructurado no está escalado, lo que puede generar problemas en la asignación de hiperpriors y en la interpretación de los resultados. La escala es crucial para asegurar que los hiperpriors utilizados en una aplicación tengan la misma interpretación en otra aplicación. Sin una escala adecuada, los hiperparámetros no tienen un significado claro y podrían interpretarse incorrectamente. Además, la elección de una distribución previa para este parámetro puede resultar poco clara.

Por otro lado, el modelo BYM2 consta de un parámetro de precisión y un parámetro de mezcla. El parámetro de precisión representa la precisión marginal y controla la variabilidad explicada por un efecto espacial. El parámetro de mezcla distribuye la variabilidad existente entre un componente estructurado y no estructurado. Es importante destacar que el componente estructurado está escalado, lo que facilita la asignación de la a priori (Riebler et al. 2016).

3.4. Enfoque Bayesiano

Los modelos jerárquicos bayesianos se utilizan a menudo para modelar datos espaciales y espacio temporales. Estos modelos permiten una flexibilidad completa en la forma en que las estimaciones toman fuerza en el espacio y el tiempo, y mejoran la estimación y la predicción de las características subyacentes del modelo. En un enfoque bayesiano, una

distribución de probabilidad $f(\mathbf{y}|\boldsymbol{\theta})$, llamada verosimilitud, se especifica para los datos observados $\mathbf{y} = (y_1, \dots, y_n)$ dado un vector de parámetros desconocidos $\boldsymbol{\theta}$ (Moraga 2019).

Es importante destacar que en el enfoque bayesiano se consideran los parámetros como variables aleatorias, es decir que $\boldsymbol{\theta}$ es una variable aleatoria con una distribución de probabilidad asociada. Por el contrario, en el enfoque frecuentista, los parámetros del modelo se consideran constantes fijas desconocidas y deben ser estimadas a partir de los datos.

Una distribución previa $f(\boldsymbol{\theta}|\boldsymbol{\eta})$ es asignada a $\boldsymbol{\theta}$ donde $\boldsymbol{\eta}$ es un vector de hiperparámetros. La distribución previa para $\boldsymbol{\theta}$ representa el conocimiento previo antes de obtener los datos (Moraga 2019). La inferencia concerniente a $\boldsymbol{\theta}$ se basa en la distribución posterior de $\boldsymbol{\theta}$, la cual se define a partir del teorema de Bayes como se muestra en la siguiente ecuación.

$$f(\boldsymbol{\theta}|\mathbf{y}) = \frac{f(\mathbf{y}, \boldsymbol{\theta})}{f(\mathbf{y})} = \frac{f(\mathbf{y}|\boldsymbol{\theta})f(\boldsymbol{\theta})}{\int f(\mathbf{y}|\boldsymbol{\theta})f(\boldsymbol{\theta})d\boldsymbol{\theta}} \quad (3-12)$$

El denominador $f(\mathbf{y}) = \int f(\mathbf{y}|\boldsymbol{\theta})f(\boldsymbol{\theta})d\boldsymbol{\theta}$ define la verosimilitud marginal de los datos. Esto está libre de $\boldsymbol{\theta}$ y puede establecerse en una constante de escala que no afecte la forma de la distribución posterior. Por lo tanto, la distribución posterior se expresa como

$$f(\boldsymbol{\theta}|\mathbf{y}) \propto f(\mathbf{y}|\boldsymbol{\theta})f(\boldsymbol{\theta}) \quad (3-13)$$

Una dificultad principal en la aplicación de métodos bayesianos es el cálculo de la distribución posterior $f(\boldsymbol{\theta}|\mathbf{y})$. Usualmente implica una integración de alta dimensión que por lo general no se puede obtener de forma analítica. Aun cuando la probabilidad y la distribución previa tienen expresiones de forma analítica, la distribución posterior puede que no. Los métodos de cadena de Markov Monte Carlo (MCMC) se han utilizado tradicionalmente para resolver este problema (Moraga 2019).

Los métodos bayesianos son útiles cuando se quiere incorporar creencias previas al modelo o cuando los métodos clásicos de estimación son complejos, por ejemplo datos espaciales, medidas repetidas, datos perdidos y datos multivariados.

Banerjee et al. (2014) define para el modelo BYM la a priori de u como

$$p(\mathbf{u}|\sigma_u) \propto \frac{1}{\sigma_u^{n/2}} \exp \left\{ -\frac{1}{2\sigma_u} \sum_i \sum_{j \in \mathcal{N}(i)} (u_i - u_j)^2 \right\}$$

La heterogeneidad no correlacionada v fue definida por Besag et al. (1991) para tener una distribución a priori gaussiana convencional de media cero:

$$p(\mathbf{v}) \propto \sigma_v^{-n/2} \exp \left\{ -\frac{1}{2\sigma_v} \sum_{i=1}^n v_i^2 \right\}$$

Tanto σ_u como σ_v fueron asumidos por Besag et al. (1991) para tener hiperprioris exponenciales inversas:

$$\text{prior}(\sigma_u, \sigma_v) \propto e^{-\epsilon/2\sigma_u} e^{-\epsilon/2\sigma_v} \quad \sigma_u, \sigma_v > 0,$$

donde ϵ es tomado como 0,001. La distribución a conjunta posterior para una distribución de probabilidad Poisson está dada por

$$\begin{aligned} P(u, v, \sigma_u, \sigma_v | y_i) = & \prod_{i=1}^n \left\{ \frac{e^{-E_i \mu_i} (E_i \mu_i)^{y_i}}{y_i!} \right\} \\ & \times \frac{1}{\sigma_u^{n/2}} \exp \left\{ -\frac{1}{2\sigma_u} \sum_i \sum_{j \in \mathcal{N}(i)} (u_i - u_j)^2 \right\} \\ & \times \sigma_v^{-n/2} \exp \left\{ -\frac{1}{2\sigma_v} \sum_{i=1}^n v_i^2 \right\} \times \text{prior}(\sigma_u, \sigma_v) \end{aligned}$$

Banerjee et al. (2014) expone la conveniencia de los modelos CAR desde el punto de vista computacional, ya que los métodos para encontrar la parte posteriori de $\mathbf{u}$ y $\mathbf{v}$ es en sí mismo un algoritmo condicional, el muestreador de Gibbs. Este algoritmo opera muestreando sucesivamente la distribución condicional conjunta de cada parámetro (es decir, la distribución de cada parámetro dados los datos y todos los demás parámetros en el modelo). Entonces, por ejemplo, la distribución condicional de u_i es

$$p(u_i | u_{i \neq j}, \mathbf{v}, \boldsymbol{\beta}, \mathbf{y}) \propto \text{Po}(y_i | E_i e^{x_i' \boldsymbol{\beta} + u_i + v_i}) \times N(u_i | \bar{u}_i, 1/(\tau_u m_i))$$

lo que significa que no se necesita trabajar con la distribución conjunta de $\mathbf{u}$ en lo absoluto. El enfoque condicional también elimina la necesidad de cualquier inversión de matriz.

3.4.1. Cadenas de Markov Monte Carlo (MCMC)

Las técnicas de cadenas de Markov Monte Carlo (MCMC) se aplican a menudo para resolver problemas de integración y optimización en espacios de grandes dimensiones. Estos dos tipos de problemas juegan un papel fundamental en el aprendizaje automático, la física, la estadística, la econometría y el análisis de decisiones. Algunos ejemplos son

la inferencia Bayesiana y aprendizaje, Mecánica estadística, optimización, selección de modelos de verosimilitud penalizada, etc. (Andrieu et al. 2003). En un contexto de estimación Bayesiana los métodos MCMC funcionan generando una muestra de valores $\{\boldsymbol{\theta}^{(g)}, g = 1, \dots, G\}$ de una cadena de Markov convergente cuya distribución estacionaria es la posterior $f(\boldsymbol{\theta}|\mathbf{y})$. A partir de estas muestras, se pueden obtener resúmenes empíricos de los valores $\boldsymbol{\theta}^{(g)}$ para resumir la distribución posterior de los parámetros de interés.

La convergencia de la cadena de Markov a la distribución estacionaria correcta puede garantizarse para una clase amplia de posteriores, lo que explica la popularidad de MCMC. Aunque esta convergencia también es la fuente de la mayor parte de la dificultad en la implementación real de los procedimientos MCMC. La primera razón es porque obliga a tomar una decisión sobre cuándo es seguro detener el algoritmo de muestreo y resumir su salida, conocido como diagnóstico de convergencia. La segunda razón es que altera la determinación de la calidad de las estimaciones producidas (ya que no se basan en extracciones i.i.d. de la posterior, sino en valores correlacionados). Esto es un problema de estimación de la varianza, ya que el objetivo es estimar las varianzas de Monte Carlo (errores estándar equivalentes) asociadas con las estimaciones posteriores basadas en MCMC. Los algoritmos más populares son Gibbs sampling y el algoritmo Metropolis-Hastings (Banerjee et al. 2014).

Los métodos MCMC se utilizan ampliamente en la inferencia bayesiana para modelos complejos. Sin embargo, estos métodos son extremadamente exigentes computacionalmente y presentan una amplia gama de problemas en términos de convergencia. La Aproximación de Laplace Anidada Integrada (INLA) es una alternativa computacional menos intensiva a MCMC diseñada para realizar inferencia bayesiana aproximada en modelos gaussianos latentes (Rue et al. 2009).

3.4.2. Aproximación de Laplace Anidada Integrada (INLA)

La Aproximación de Laplace Anidada Integrada (INLA) es un método reciente para ajustar modelos Bayesianos. El objetivo de INLA es resolver la dificultad computacional de MCMC en problemas con bastantes datos. En bastantes aplicaciones, el proceso de muestreo de la distribución posterior usando MCMC puede tardar demasiado y en muchos casos no es viable con los recursos computacionales existentes.

INLA permite realizar una inferencia bayesiana aproximada en modelos Gaussianos latentes como los modelos mixtos lineales generalizados, modelos espaciales y espacio-temporales.

Específicamente, los modelos son de la forma

$$\begin{aligned} y_i | \boldsymbol{\psi}, \boldsymbol{\theta} &\sim f(y_i | \boldsymbol{\psi}, \boldsymbol{\theta}), \quad i = 1, \dots, n, \\ \boldsymbol{\psi} | \boldsymbol{\theta} &\sim N(\mu(\boldsymbol{\theta}), Q(\boldsymbol{\theta})^{-1}), \\ \boldsymbol{\theta} &\sim f(\boldsymbol{\theta}) \end{aligned} \tag{3-14}$$

donde y_i son los datos observados, $\boldsymbol{\psi}$ representa un campo Gaussiano, y $\boldsymbol{\theta}$ son los hiperparámetros. $\mu(\boldsymbol{\theta})$ es la media y $Q(\boldsymbol{\theta})$ es la matriz de precisión (la inversa de la matriz de covarianza) del campo Gaussiano latente $\boldsymbol{\psi}$. Aquí y y $\boldsymbol{\psi}$ pueden ser de alta dimensión. Sin embargo, para realizar inferencias rápidas, la dimensión del vector de hiperparámetros $\boldsymbol{\theta}$ debe ser pequeña porque las aproximaciones se calculan usando integración numérica sobre el espacio de hiperparámetros (Moraga 2019).

En muchas ocasiones, se supone que las observaciones y_i pertenecen a una familia exponencial con media $\mu_i = g^{-1}(\eta_i)$. El predictor lineal η_i expresa los efectos de las covariables de forma aditiva

$$\eta_i = \alpha + \sum_{k=1}^{n_\beta} \beta_k x_{ki} + \sum_{j=1}^{n_f} f^{(j)}(u_{ji}) \tag{3-15}$$

donde, α es el intercepto, β_k cuantifica los efectos lineales de las covariables x_{ki} en la respuesta, y $f^{(j)}(\cdot)$ son un conjunto de efectos aleatorios definidos en términos de algunas covariables u_{ji} . Esta formulación permite acomodar una amplia gama de modelos gracias a las diferentes formas que pueden tomar las funciones $f^{(j)}$ incluyendo modelos espaciales y espacio-temporales.

INLA utiliza una combinación de aproximaciones analíticas e integración numérica para obtener distribuciones posteriores aproximadas de los parámetros. Sea $\boldsymbol{\psi} = (\alpha, \{\beta_k\}, \{f^{(j)}\}) | \boldsymbol{\theta} \sim N(\mu(\boldsymbol{\theta}), Q(\boldsymbol{\theta})^{-1})$ el vector de las variables gaussianas latentes, y sea $\boldsymbol{\theta}$ el vector de hiperparámetros que no son necesariamente gaussianos (Moraga 2019).

Las marginales posteriores de cada elemento ψ_i del campo latente $\boldsymbol{\psi}$ están dadas por

$$f(\boldsymbol{\psi} | \mathbf{y}) = \int f(\boldsymbol{\psi} | \boldsymbol{\theta}, \mathbf{y}) f(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta} \tag{3-16}$$

y las marginales posteriores de los hiperparámetros se pueden escribir como

$$f(\boldsymbol{\theta}_j | \mathbf{y}) = \int f(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta}_{-j} \tag{3-17}$$

Esta formulación anidada se utiliza para aproximar $f(\boldsymbol{\psi}_i | \mathbf{y})$ mediante la combinación de aproximaciones analíticas (a las condicionales $f(\boldsymbol{\psi}_i | \boldsymbol{\theta}, \mathbf{y})$ y $f(\boldsymbol{\theta} | \mathbf{y})$) y rutinas de integración

numérica para integrar $\boldsymbol{\theta}$. De manera similar, $f(\boldsymbol{\theta}_j|\mathbf{y})$ se calcula mediante la aproximación de $f(\boldsymbol{\theta}|\mathbf{y})$ e integrando $\boldsymbol{\theta}_{-j}$. Específicamente, la densidad a posteriori de los hiperparámetros se calcula mediante una aproximación gaussiana a posteriori del campo latente, $\tilde{f}_G(\boldsymbol{\psi}|\boldsymbol{\theta}, \mathbf{y})$, evaluada del siguiente modo, $\boldsymbol{\psi}^*(\boldsymbol{\theta}) = \arg \max_{\boldsymbol{\psi}} f_G(\boldsymbol{\psi}|\boldsymbol{\theta}, \mathbf{y})$,

$$\tilde{f}(\boldsymbol{\theta}|\mathbf{y}) \propto \frac{f(\boldsymbol{\psi}, \boldsymbol{\theta}, \mathbf{y})}{\tilde{f}_G(\boldsymbol{\psi}|\boldsymbol{\theta}, \mathbf{y})} \Big|_{\boldsymbol{\psi}=\boldsymbol{\psi}^*(\boldsymbol{\theta})} \quad (3-18)$$

Luego INLA construye la siguiente aproximación anidada:

$$\tilde{f}(\boldsymbol{\psi}_i|\mathbf{y}) = \int \tilde{f}(\boldsymbol{\psi}_i|\boldsymbol{\theta}, \mathbf{y}) \tilde{f}(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta}, \quad \tilde{f}(\boldsymbol{\theta}_j|\mathbf{y}) = \int \tilde{f}(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta}_{-j} \quad (3-19)$$

Finalmente, estas aproximaciones pueden ser integradas numéricamente respecto a $\boldsymbol{\theta}$

$$\begin{aligned} \tilde{f}(\boldsymbol{\psi}_i|\mathbf{y}) &= \sum_k \tilde{f}(\boldsymbol{\psi}_i|\boldsymbol{\theta}_k, \mathbf{y}) \tilde{f}(\boldsymbol{\theta}_k|\mathbf{y}) \times \Delta_k, \\ \tilde{f}(\boldsymbol{\theta}_j|\mathbf{y}) &= \sum_l \tilde{f}(\boldsymbol{\theta}_l^*|\mathbf{y}) \times \Delta_l^* \end{aligned} \quad (3-20)$$

donde Δ_k (Δ_l^*) indica el peso del área correspondiente a $\boldsymbol{\theta}_k$ ($\boldsymbol{\theta}_l^*$).

Las aproximaciones para las marginales a posteriori de las $\boldsymbol{\psi}_i$ condicionadas a los valores seleccionados de $\boldsymbol{\theta}_k$, $\tilde{f}(\boldsymbol{\psi}_i|\boldsymbol{\theta}_k, \mathbf{y})$, pueden obtenerse utilizando una aproximación Gaussiana, una Laplace o una aproximación simplificada de Laplace.

4. Metodología

En esta sección se presenta la metodología utilizada en este proyecto. En la primera parte se plantea la descripción de la población de estudio y sus características más importantes. Se explica el proceso de depuración y organización de la base de datos. Luego se realiza el análisis estadístico de los datos, en donde se incluye un análisis exploratorio para observar el comportamiento espacial de los datos. Se estima la correlación espacial entre los municipios del Valle del Cauca y el efecto de las covariables en el riesgo de infección de dengue, además de identificar los municipios con mayor riesgo.

4.1. Datos del estudio

En esta investigación la población de estudio son los casos confirmados de dengue en el período de tiempo de enero a diciembre del año 2019, para los 42 municipios del Valle del Cauca. Los datos sobre casos de dengue son extraídos del [Sistema Nacional de Vigilancia en Salud Pública \(Sivigila\)](#) del Instituto Nacional de Salud (INS).

Los datos de variables climáticas fueron extraídos de la plataforma [Visual Crossing Weather Data](#). Esta plataforma utiliza varias fuentes de información para los datos meteorológicos, una de ellas es la base de datos de superficie integrada (ISD) de la NOAA (Oficina Nacional de Administración Oceánica y Atmosférica). La ISD es una gran colección de observaciones meteorológicas por día y por hora recopiladas de todo el mundo. Las variables climáticas extraídas fueron la temperatura máxima y mínima, precipitación y humedad relativa.

Esta plataforma es de acceso público, sin embargo, para poder descargar la información es necesario registrarse con una cuenta. En esta plataforma se busca el municipio que se desea consultar y se selecciona el periodo de tiempo para obtener la información que se necesita. Una vez completada la búsqueda se realiza la descarga de esa información. Es importante destacar que el criterio de asociación entre la estación y el municipio lo realiza la plataforma considerando las estaciones meteorológicas cercanas en un rango de 50-100 km. La plataforma automáticamente toma los resultados de cada estación, elimina cualquier observación faltante y además agrega valores ponderados de las estaciones más cercanas, otorgándoles mayor relevancia que a las estaciones más lejanas. Este proceso se hace a nivel de observación individual para acomodar los datos faltantes a corto plazo. Este enfoque

brinda los mejores resultados para posibles datos faltantes e inconsistencias de las estaciones.

Las estaciones meteorológicas que se utilizaron para la recopilación de la información se muestran en la Tabla 4-1. Como se observa en la Figura 4-1, no todas las estaciones se encuentran ubicadas en el departamento del Valle del Cauca, de modo que para algunos municipios se utiliza la estación más cercana (Ver Anexo A-1).

Tabla 4-1.: Descripción de las Estaciones Meteorológicas

Estación	ID	Ubicación	Longitud	Latitud
Alfonso Bonilla Aragon	80259099999	Palmira	−76.38	3.54
SKCL	SKCL	Palmira	−76.38	3.55
Gerardo Tobar Lopez	80252099999	Ciudad Puerto de Buenav.	−76.99	3.82
SKBU	SKBU	Ciudad Puerto de Buenav.	−77.07	3.88
Matecaña	80210099999	Pereira	−75.73	4.81
SKPE	SKPE	Pereira	−75.73	4.82
La Nubia	80149099999	Manizales	−75.47	5.03
Perales	80214099999	Ibaguè	−75.13	4.42
SKIB	SKIB	Ibaguè	−75.76	4.43
El Eden	80211099999	Armenia	−75.76	4.45
SKAR	SKAR	Armenia	−75.72	4.5
Santa Ana	80222799999	Cartago	−75.95	4.75

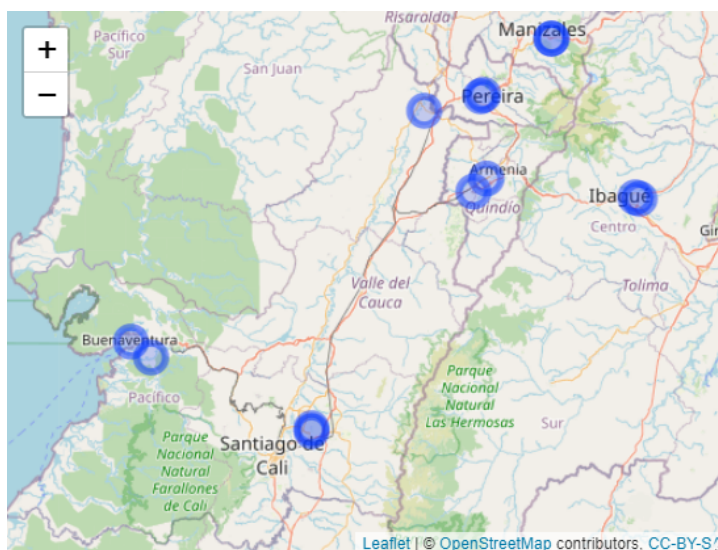


Figura 4-1.: Ubicación de estaciones meteorológicas

Adicionalmente se tienen en cuenta otras variables que se consideran importantes en el estudio, estas variables corresponden a: altitud (msnm), área de cada municipio (km^2),

población total de cada municipio y población urbana. La información de altitud y área de cada municipio fue extraída del [Geoportal Instituto Geográfico Agustín Codazzi \(IGAC\)](#), y la información relacionada a la población pertenece a las [proyecciones de población del DANE](#).

4.2. Variables de estudio

A partir de los datos recopilados de las diferentes bases de datos mencionadas anteriormente, se utilizan para el estudio las siguientes variables. Ver Tabla 4-2.

Tabla 4-2.: Información recopilada de las bases de datos

Variable	Tipo	Descripción
FEC_CON	Cualitativa nominal	Fecha registrada en la que consultó el paciente por el evento de dengue
Confirmado	Cualitativa nominal	Presencia de dengue en un individuo. (1=confirmado; 0=No confirmado)
Departamento_ocurrencia	Cualitativa nominal	Departamento de dónde proviene u ocurre el caso de dengue.
Municipio_ocurrencia	Cualitativa nominal	Municipio de donde proviene u ocurre el caso de dengue.
tempmax	Cuantitativa continua	Temperatura máxima diaria registrada en cada municipio. Escala: Celsius ($^{\circ}C$).
tempmin	Cuantitativa continua	Temperatura mínima diaria registrada en cada municipio. Escala: Celsius ($^{\circ}C$).
temp	Cuantitativa continua	Temperatura promedio diaria registrada en cada municipio. Escala: Celsius ($^{\circ}C$).
humidity	Cuantitativa continua	Humedad relativa diaria registrada en el municipio. Valor en porcentaje (%).
precip	Cuantitativa continua	Precipitación acumulada diaria en milímetros en cada municipio.
Altitud	Cuantitativa continua	Altitud promedio de la cabecera municipal, tomando como referencia la superficie del nivel del mar (msnm).
Area	Cuantitativa continua	Área geográfica del municipio.
Poblacion	Cuantitativa discreta	Número total de habitantes por cada municipio.
Pob_Urbana	Cuantitativa discreta	Número de habitantes que residen en la zona urbana de cada municipio.

4.2.1. Preparación de la base de datos

Antes de realizar el procesamiento de la información, se depuró la base de datos y se reestructuró para poder realizar el análisis. Inicialmente los datos de las variables climáticas se tenían diarios, por lo que se calculó el promedio de tempmax, tempmin, temp, humidity y la suma acumulada de la precipitación para el año 2019 para cada municipio; para los datos de dengue primero se filtraron los casos para el departamento del Valle del Cauca y posteriormente se realizó el conteo de todos los casos por cada municipio. Toda la información se encontraba en 3 bases de datos diferentes, por lo que fue necesario unirla. Finalmente la base de datos que se creó cuenta con las variables que se presentan en la Tabla 4-3.

Tabla 4-3.: Variables de estudio

Variable	Tipo	Descripción
id	Cualitativa nominal	Municipio de donde proviene u ocurre el caso de dengue.
casos	Cuantitativa discreta	Número de casos de dengue para cada municipio en un determinado periodo de tiempo.
tempmaxprom	Cuantitativa continua	Temperatura máxima promedio anual en cada municipio. Escala: Celsius ($^{\circ}C$).
tempminprom	Cuantitativa continua	Temperatura mínima promedio anual en cada municipio. Escala: Celsius ($^{\circ}C$).
temp	Cuantitativa continua	Temperatura promedio anual en cada municipio. Escala: Celsius ($^{\circ}C$).
humidity	Cuantitativa continua	Humedad relativa promedio anual registrada en el municipio. Valor en porcentaje (%)
precip	Cuantitativa continua	Precipitación acumulada en milímetros para un determinado periodo de tiempo.
Altitud	Cuantitativa continua	Altitud promedio de la cabecera municipal, tomando como referencia la superficie del nivel del mar (msnm).
Población	Cuantitativa discreta	Número total de habitantes por cada municipio.
%Pob_Urbana	Cuantitativa continua	Porcentaje de habitantes que residen en la zona urbana de cada municipio.

4.3. Análisis Estadístico

En esta sección se explica la metodología estadística empleada para desarrollar este trabajo. Primero se realiza un análisis exploratorio de los datos en donde se quiere conocer el comportamiento del dengue y de las variables climáticas por municipio para el año 2019. También se explica el procedimiento para calcular las tasas de infección de dengue para cada municipio.

4.3.1. Análisis exploratorio

Se realizó un análisis exploratorio mediante mapas para observar la distribución de casos de dengue y el comportamiento de las variables temperatura, precipitación, humedad, altitud y el porcentaje de población urbana en los diferentes municipios del Valle del Cauca.

Tasas

Las tasas expresan la dinámica del dengue en la población de cada municipio para el año 2019. Se pueden definir como la magnitud del cambio de la enfermedad por unidad de cambio del tiempo en relación con el tamaño de la población que se encuentra en riesgo de experimentar el suceso (Moreno-Altamirano et al. 2007).

Las tasas de casos de dengue se calcularon para cada municipio del Valle del Cauca, dividiendo el número de casos de dengue sobre la proyección poblacional de cada municipio para el año 2019 y luego se multiplicó este valor por 100.000 habitantes.

$$\text{Tasa} = \frac{\text{Número de casos de dengue de cada municipio en el año 2019}}{\text{Proyección poblacional de cada municipio para el año 2019}} \times 100.000 \quad (4-1)$$

4.3.2. Análisis espacial

En el análisis espacial se calcula la razón de incidencia estandarizada (SIR) y se representa gráficamente a través de un mapa con el fin de observar su distribución espacial. Luego se calcula la matriz de adyacencia para ajustar un modelo BYM2 utilizando las aproximaciones de Laplace anidadas integradas.

SIR

Se proporcionaron estimaciones del riesgo de infección de dengue en cada uno de los municipios del Valle del Cauca. Para cada municipio i , $i = 1, \dots, 42$, se calculó el SIR , que está definido como la razón entre el número de casos de dengue observados Y_i y el número de casos de dengue esperados E_i (Ecuación 3-9). Se selecciona como población estándar todo el departamento del Valle del Cauca.

Matriz de adyacencia

En la Figura 4-2 se puede ver una representación gráfica de la matriz de adyacencia para los municipios del Valle del Cauca en el orden que se presenta en la Tabla 4-4. Esta definición de matriz de adyacencia binaria presenta algunas complicaciones. Al ser las áreas irregulares y contar con características geográficas muy diferentes entre ellas, puede ocasionar que áreas alejadas pero con límites en común sean catalogadas como vecinos. Una alternativa es definir w_{ij} de manera que refleje una distancia entre áreas conservando una estructura binaria. Se podría establecer $w_{ij} = 1$ para todo i y j dentro de una distancia específica. O, para un i dado, se podría establecer $w_{ij} = 1$ si j es uno de los K vecinos más cercanos (en distancia) de i . Las opciones anteriores sugieren que W sería simétrica. Sin embargo, para unidades de área irregulares, este no tiene por qué ser el caso. Para este estudio se escoge trabajar con la opción clásica de matriz de adyacencia, que es la comúnmente utilizada en la literatura (Banerjee et al. 2014).

Para crear la matriz de vecinos en Rstudio se utilizaron las funciones `poly2nb()` y `nb2INLA()` de la librería “spdep”. Para ello, primero se utilizó la función `poly2nb()` para crear un listado de vecinos basado en fronteras contiguas. Luego, se utiliza la función `nb2INLA()` para convertir el listado en un archivo con la representación de la matriz de vecinos en el formato que requiere R-INLA. Posteriormente este archivo se usa para establecer el efecto espacial estructurado.

Tabla 4-4.: Vecinos de cada municipio del Valle del Cauca (Los números de la columna vecinos corresponden al orden de los municipios)

Orden	Municipio	Num. vecinos	Vecinos
1	El Águila	1	2
2	Ansermanuevo	5	1, 3, 4, 5, 8
3	El Cairo	3	2, 4, 9
4	Argelia	4	2, 3, 8, 9
5	Cartago	5	2, 6, 7, 8, 40
6	Alcalá	2	5, 40
7	Obando	4	5, 8, 11, 13
8	Toro	6	2, 4, 5, 7, 9, 11
9	Versalles	5	3, 4, 8, 10, 11
10	El Dovio	4	9, 11, 12, 18
11	La Unión	6	7, 8, 9, 10, 12, 13
12	Roldanillo	5	10, 11, 13, 14, 18
13	La Victoria	4	7, 11, 12, 14
14	Zarzal	5	12, 13, 16, 18, 41
15	Caicedonia	1	41
16	Bugalagrande	6	14, 17, 18, 19, 21, 41
17	Andalucía	4	16, 19, 20, 21
18	Bolívar	5	10, 12, 14, 16, 19
19	Trujillo	4	16, 17, 18, 20
20	Riofrío	5	17, 19, 21, 24, 26
21	Tuluá	7	16, 17, 20, 22, 23, 26, 41
22	San Pedro	3	21, 23, 26
23	Buga	6	21, 22, 26, 28, 29, 30
24	Calima (El Darién)	6	20, 25, 26, 36, 37, 42
25	Restrepo	4	24, 26, 27, 37
26	Yotoco	9	20, 21, 22, 23, 24, 25, 27, 28, 30
27	Vijes	6	25, 26, 30, 31, 34, 37
28	Guacarí	4	23, 26, 29, 30
29	Ginebra	3	23, 28, 30
30	El Cerrito	6	23, 26, 27, 28, 29, 31
31	Palmira	6	27, 30, 32, 34, 35, 39
32	Candelaria	4	31, 33, 35, 39
33	Florida	2	32, 39
34	Yumbo	5	27, 31, 35, 36, 37
35	Cali	6	31, 32, 34, 36, 38, 42
36	Dagua	5	24, 34, 35, 37, 42
37	La Cumbre	5	24, 25, 27, 34, 36
38	Jamundí	2	35, 42
39	Pradera	3	31, 32, 33
40	Ulloa	2	5, 6
41	Sevilla	4	14, 15, 16, 21
42	Buenaventura	4	24, 35, 36, 38

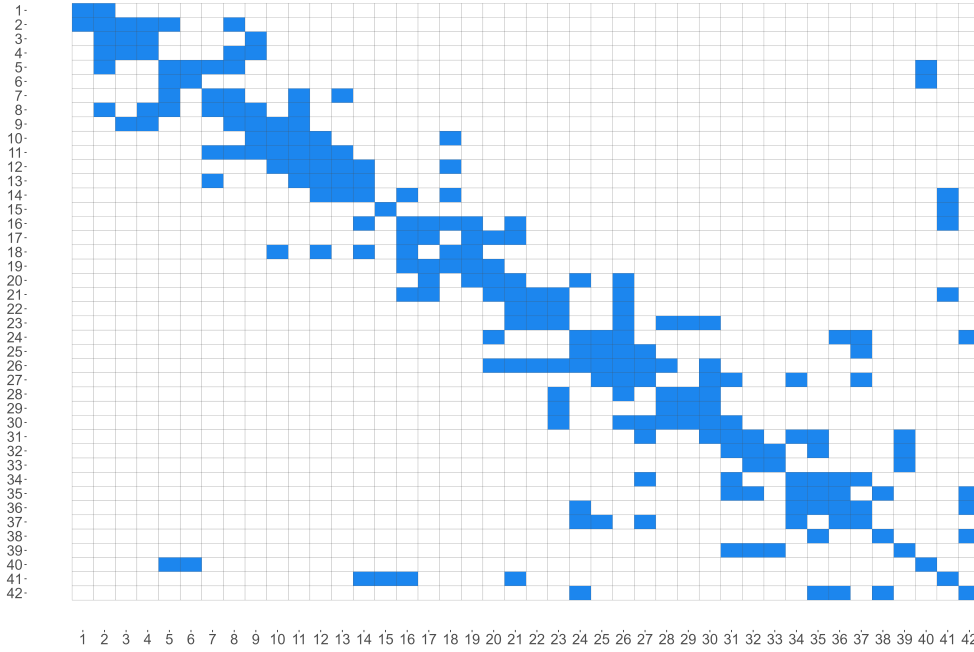


Figura 4-2.: Representación gráfica de la matriz de adyacencia (Los cuadros azules identifican los vecinos. Las filas y columnas identifican los municipios)

Modelación

El modelo Poisson se utiliza cuando la variable respuesta es un recuento de algún evento durante un período de tiempo o espacio determinado. En este caso, la variable respuesta corresponde al número de casos de dengue en los 42 municipios del Valle del Cauca en el año 2019. Sea E_i el número de personas en riesgo en el municipio i ($i = 1, \dots, 42$) y y_i el número de casos de dengue en el municipio i . Se especifica el modelo asumiendo que los conteos observados y_i son condicionalmente independientes distribuidos Poisson

$$y_i | \mu_i \sim \text{Poisson}(E_i \mu_i), \quad i = 1, \dots, 42 \quad (4-2)$$

donde E_i es el número esperado de casos como se muestra en (3-10) y μ_i es el riesgo relativo en el municipio i . Se asume que el logaritmo del riesgo $\eta_i = \log(\mu_i)$ se descompone en:

$$\eta_i = \alpha + \mathbf{x}_i' \boldsymbol{\beta} + b_i \quad (4-3)$$

En donde α denota el nivel de riesgo general, $\mathbf{x}_i' = (x_{i1}, \dots, x_{ip})'$ un conjunto de p covariables con los parámetros de regresión correspondientes $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)$ y $\mathbf{b} = (b_1, \dots, b_n)$ un efecto aleatorio. Los efectos aleatorios $\mathbf{b}$ se utilizan para tener en cuenta la variación extra-Poisson o la correlación espacial debido a factores de riesgo latentes o no medidos (Riebler et al. 2016).

■ Modelo BYM2

En términos prácticos, el modelo BYM2 ofrece beneficios en cuanto a interpretabilidad y asignación de la a priori, por lo que hacen que sea mejor que el modelo BYM. Por ello, se propone modelar la correlación espacial del riesgo de infección de dengue entre los municipios con el modelo BYM2, ya que es un modelo utilizado para el análisis de datos espaciales en el marco de la estadística bayesiana.

Cabe resaltar que en el análisis de datos espaciales de enfermedades es posible utilizar un GLMM dentro del contexto del modelo BYM2. Se puede considerar un GLMM para modelar la relación entre las covariables explicativas y la ocurrencia de la enfermedad, mientras que el componente espacial del BYM2 se utilizaría para capturar la correlación espacial no explicada por las covariables.

Es importante mencionar que la variables precipitación y altitud presentaban una gran dimensión de los datos, por lo que se dividieron por 100 para mejorar la interpretación del modelo. Es decir, que el cambio unitario para la precipitación esta dado por cada 100 milímetros y para la altitud es por cada 100 metros.

Teniendo en cuenta lo anterior, se ajustará el siguiente modelo:

$$\eta_i = \log(\mu_i) = \beta_0 + \beta_1 \text{precip}_i + \beta_2 \text{tempmax}_i + \beta_3 \text{humidity}_i + \beta_4 \text{pob_urb}_i + \beta_5 \text{altitud}_i + b_i \quad (4-4)$$

en donde β_0 es el intercepto que representa el riesgo total, los demas β_i son los coeficientes de las covariables. Teniendo en cuenta el modelo BYM2, el efecto aleatorio está dado por

$$\begin{aligned} \mathbf{b} &= \frac{1}{\sqrt{\tau_b}} (\sqrt{1 - \phi} \mathbf{v} + \sqrt{\phi} \mathbf{u}_*) \\ \mathbf{u}_* &\sim N(\mathbf{0}, \mathbf{Q}_*^-), \\ \mathbf{v} &\sim N(\mathbf{0}, \mathbf{I}) \end{aligned}$$

El escalado de la matriz $\mathbf{Q}$ es crucial para facilitar la asignación de hiperprioris y garantizar que las hiperprioris utilizadas en una aplicación tengan la misma interpretación en otra aplicación. Esto se puede hacer por medio de la función `inla.scale.model()` del paquete INLA, que toma como argumento cualquier matriz de precisión singular $\mathbf{Q}$ y las restricciones lineales que abarcan en el espacio nulo de $\mathbf{Q}$. Se obtiene la matriz de precisión escalada, donde la media geométrica de las varianzas marginales es igual a uno (Riebler et al. (2016), sección 3.2).

■ Distribución a priori para los parámetros

En los modelos de datos de áreas a menudo se utiliza una a priori no informativa para evitar introducir sesgos o suposiciones injustificadas en el análisis. Los datos espaciales pueden ser altamente complejos y variados, lo que dificulta la especificación de una a priori precisa. Además, la elección de una a priori específica puede influir en gran medida en los resultados. Utilizar una a priori no informativa permite que los datos proporcionen la mayor parte de la información para estimar los parámetros del modelo y obtener la distribución a posterior (Banerjee et al. 2014, Lawson 2013).

Cuando el límite de la a priori conjugada (normal) es en realidad una a priori uniforme o “plana”, la posterior se convierte en la verosimilitud (posiblemente renormalizada para integrar 1 como función de θ). Por supuesto, la a priori plana es impropia, ya que la uniforme no integra a ningún valor finito sobre toda la línea real. Sin embargo, la posterior sigue estando bien definida ya que la verosimilitud puede integrarse con respecto a θ (Banerjee et al. 2014).

Estimación

Se realizó la implementación del modelo BYM2 utilizando aproximaciones de Laplace anidadas integradas (INLA). Este método proporciona una inferencia bayesiana aproximada. En los últimos años se ha consolidado como una alternativa a otros métodos como, cadenas de Markov Monte Carlo, por su rapidez y facilidad de uso a través del paquete R-INLA en Rstudio. (Para mayor información y descarga del paquete acceder a <https://www.r-inla.org/>).

Las aproximaciones de las marginales a posteriori de las ψ_i condicionadas a los valores seleccionados de θ_k se pueden obtener por medio de una aproximación Gaussiana, de Laplace o aproximación simplificada de Laplace. En el paquete R-INLA la opción predeterminada para realizar las aproximaciones es la aproximación simplificada de Laplace, ya que tiene menor costo y soluciona las imprecisiones de ubicación y asimetría de la aproximación Gaussiana.

La aproximación de las marginales posteriores se calculan en tres pasos. El primer paso consiste en aproximar la marginal posterior de θ usando la aproximación de Laplace. En el segundo paso se calcula la aproximación de Laplace o la aproximación simplificada de Laplace de $f(\psi_i|\theta, \mathbf{y})$ para valores seleccionados de θ , para mejorar la aproximación gaussiana. Finalmente el tercer paso combina los dos pasos anteriores utilizando la integración numérica.

La función `inla()` realiza un análisis bayesiano completo de modelos aditivos utilizando la aproximación de Laplace anidada integrada. Es importante especificar en esta función un objeto `formula` que incluye la variable respuesta, las covariables y los efectos fijos y aleatorios. Los efectos aleatorios son establecidos usando la función `f()` con el parámetro del índice de la variable y el modelo a utilizar “bym2”. Además, se indica el objeto que contiene la matriz de adyacencia y, en caso de especificarlo, la distribución a priori. Finalmente, se ajusta el modelo con la función `inla()` indicando la `formula`, la familia (`family="poisson"`), los datos y los casos esperados.

En este caso, en la función `inla()` no se especifica una distribución a priori, por lo que se utiliza por defecto una distribución de complejidad penalizada (PC) con valores 0.5 en los parámetros. Una a priori PC permite distribuir la variación total entre los componentes del modelo BYM modificado. Para mayor detalle sobre la distribución a priori, ver Simpson et al. (2015).

Al realizar la estimación del modelo se obtienen las distribuciones marginales posteriores de los parámetros, los predictores lineales y los valores ajustados. También proporciona estimaciones a distintos criterios para evaluar y comparar modelos bayesianos. Estos incluyen el criterio de información de la desviación del modelo (DIC) y el criterio de información de Watanabe-Akaike (WAIC).

Sin embargo, se encontró una fuerte relación entre la humedad y las variables temperatura máxima y precipitación. Al ajustar el modelo con la variable humedad se evidenció la presencia de multicolinealidad. De acuerdo con Lowe et al. (2011), estas variables presentan una influencia significativa en los casos de dengue. Se decidió eliminar la variable humedad del modelo con el objetivo de mejorar los resultados.

5. Resultados

En esta sección se presentan las estadísticas descriptivas de los casos de dengue en los 42 municipios del Valle del Cauca para el año 2019. De igual manera se realiza el análisis descriptivo para las variables climáticas y geográficas utilizadas en este estudio. Finalmente se presentan los resultados obtenidos en el análisis espacial de los casos de dengue.

5.1. Análisis exploratorio

Para el año 2019 se cuenta con 3824 casos confirmados de dengue para el departamento del Valle del Cauca. El municipio que presenta más casos es Cali con 1970, seguido de Palmira y Buga con 505 y 335 casos respectivamente. Para algunos municipios no se tienen datos sobre los casos confirmados de dengue, esos municipios son Argelia, El Águila, El Cairo y el Dovio. Además municipios como Alcalá y Toro reportaron 1 caso confirmado de dengue (Ver Anexo **B-1**).

En la Figura **5-1** se observan las tasas de dengue en cada uno de los 42 municipios del Valle del Cauca, para el año 2019. Se puede identificar que el municipio de Bugalagrande es el que mayor tasa de infección de dengue presentó en este año (286.88 por 100,000 habitantes), es decir, indica una mayor ocurrencia de la enfermedad en relación con su población. El municipio de Buga presenta una tasa de 261.07 casos por 100,000 habitantes. Seguido se encuentran los municipios de Florida, Zarzal y Palmira en donde la tasa se encuentra entre 140 y 170 casos por 100,000 habitantes. Aunque Cali es el municipio con mas casos, la tasa es relativamente baja (87.88 casos por 100,000 habitantes).

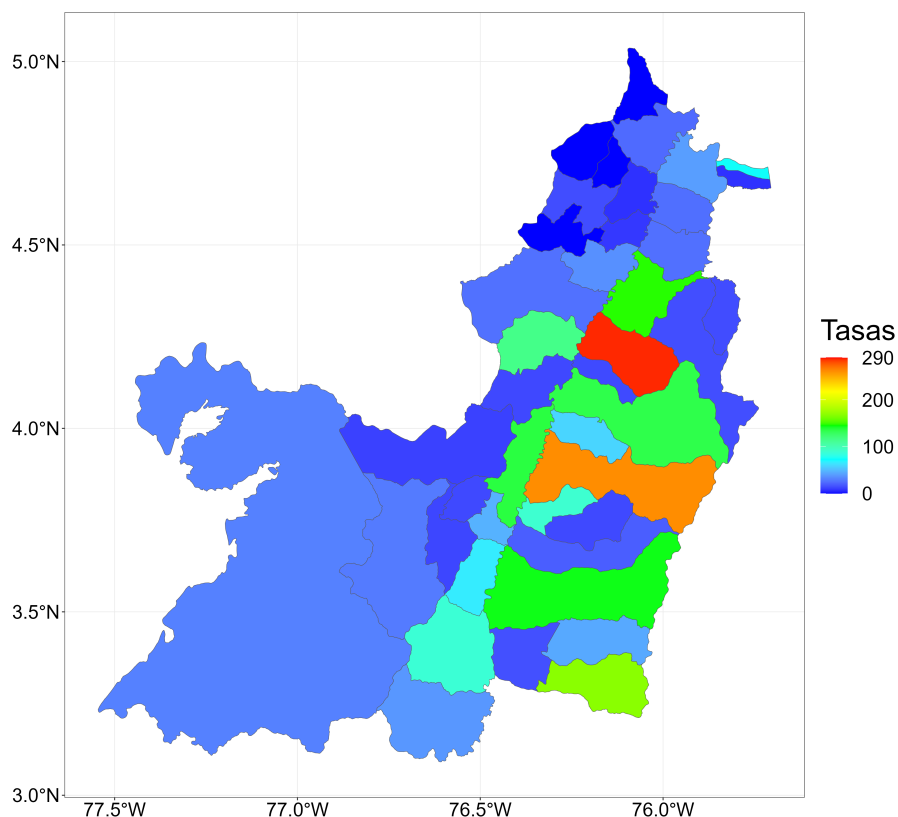


Figura 5-1.: Tasa de dengue por 100,000 habitantes por municipio en el año 2019

En la Figura 5-2 (A) se presenta la altitud de los municipios del Valle del Cauca, se puede observar que en el Valle del Cauca hay una diversidad altitudinal significativa que va desde municipios con áreas cercanas al nivel del mar hasta municipios con zonas montañosas que tienen mayor elevación. En la zona occidental del Valle del Cauca se encuentra el municipio de Buenaventura que cuenta con una altitud cercana al nivel del mar. En la zona del centro y el sur se encuentran municipios con una altitud alrededor de 1000 metros sobre el nivel del mar, en el municipio de Dagua se tiene una altitud de 827 msnm, en Cali y Palmira se tienen 991 y 1003 msnm respectivamente.

También en el centro del departamento, hacia las zonas montañosas, la altitud aumenta considerablemente. Municipios como Calima (El Darién), Restrepo y Trujillo se ubican entre los 1300 y 1500 metros sobre el nivel del mar. En la zona oriente se encuentra el municipio de Sevilla con una altitud de 1613 msnm. Finalmente en la zona norte se encuentran los municipios con una altitud entre 1500 y 2000 metros sobre el nivel del mar, los municipios de Versalles, El Cairo y el Águila tienen una altitud de 1900, 1800 y 1750 msnm respectivamente.

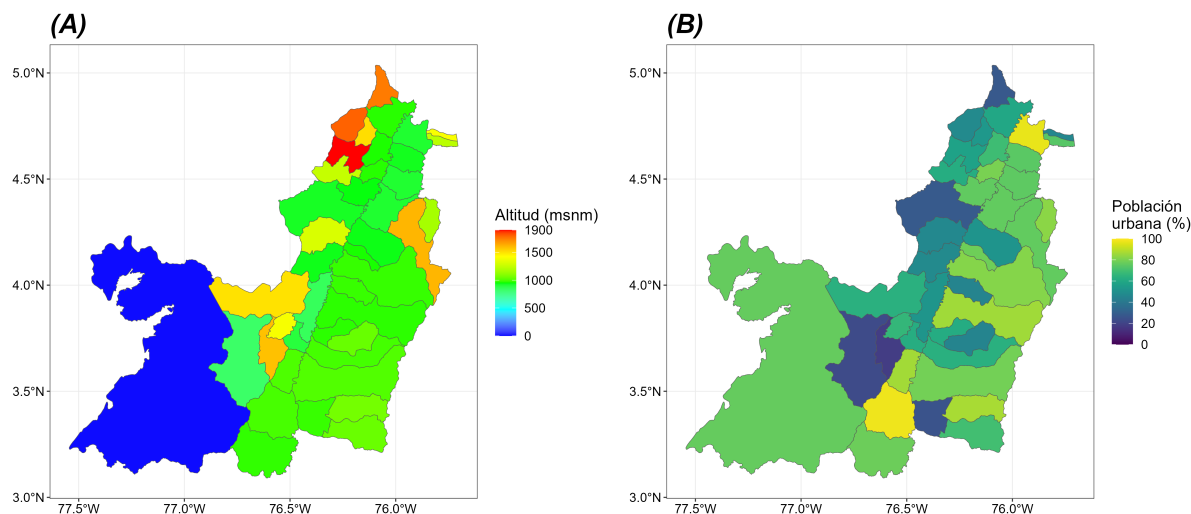


Figura 5-2.: Factores geográficos. (A) Altitud. (B) Porcentaje de población urbana en los municipios del Valle del Cauca

Igualmente en la Figura 5-2 (B) se presenta el porcentaje de población urbana en cada municipio para el año 2019, se puede observar una variación en la proporción de población urbana, sin embargo por lo general esta proporción es alta en comparación con la población rural. Cabe destacar que los municipios de La Cumbre, Dagua, Candelaria, El Águila y Bolívar tienen un porcentaje de población urbana que oscila entre el 15 y el 30 %, siendo el municipio de La cumbre el de menor porcentaje con un 17.8 %. En general la población urbana para la mayoría de municipios se encuentra entre el 40 % y 80 %. Los municipios de Cali y Cartago son los que presentan un mayor porcentaje de población urbana con el 97.7 % y 96.7 % respectivamente.

En la Figura 5-3 se muestra la representación espacial de las variables climáticas en los municipios del Valle del Cauca, incluyendo la temperatura máxima en $^{\circ}C$, precipitación en milímetros y el porcentaje de humedad relativa. En la Figura 5-3 (A) se observa que la temperatura máxima promedio anual oscila entre $26^{\circ}C$ y $31^{\circ}C$. Los municipios destacados por presentar las temperaturas más altas son Cali, Candelaria, Florida, Jamundí, Buga y Palmira con una temperatura máxima promedio anual de $31^{\circ}C$. Estos municipios se encuentran ubicados principalmente en la zona sur y centro del departamento. Por otro lado, los municipios de Buenaventura, El Águila, Ansermanuevo y Argelia registran una temperatura máxima promedio de $26^{\circ}C$. Es importante destacar que estos valores son promedios anuales y pueden existir variaciones diarias y estacionales en las temperaturas máximas de cada municipio. Además, la ubicación geográfica y la altitud también pueden influir en las diferencias de temperatura observadas entre los municipios.

La Figura 5-3 (B) muestra el comportamiento de la precipitación acumulada anual en los 42 municipios del departamento del valle del Cauca durante el año 2019. Se puede notar que el municipio de Buenaventura destaca por presentar mayor cantidad de precipitación acumulada. Además, se puede apreciar una tendencia de niveles altos de precipitación en los municipios que se encuentran ubicados en la zona norte del departamento. Por otro lado, en la zona sur y centro del departamento se encuentran los municipios que registran menor cantidad de precipitación acumulada.

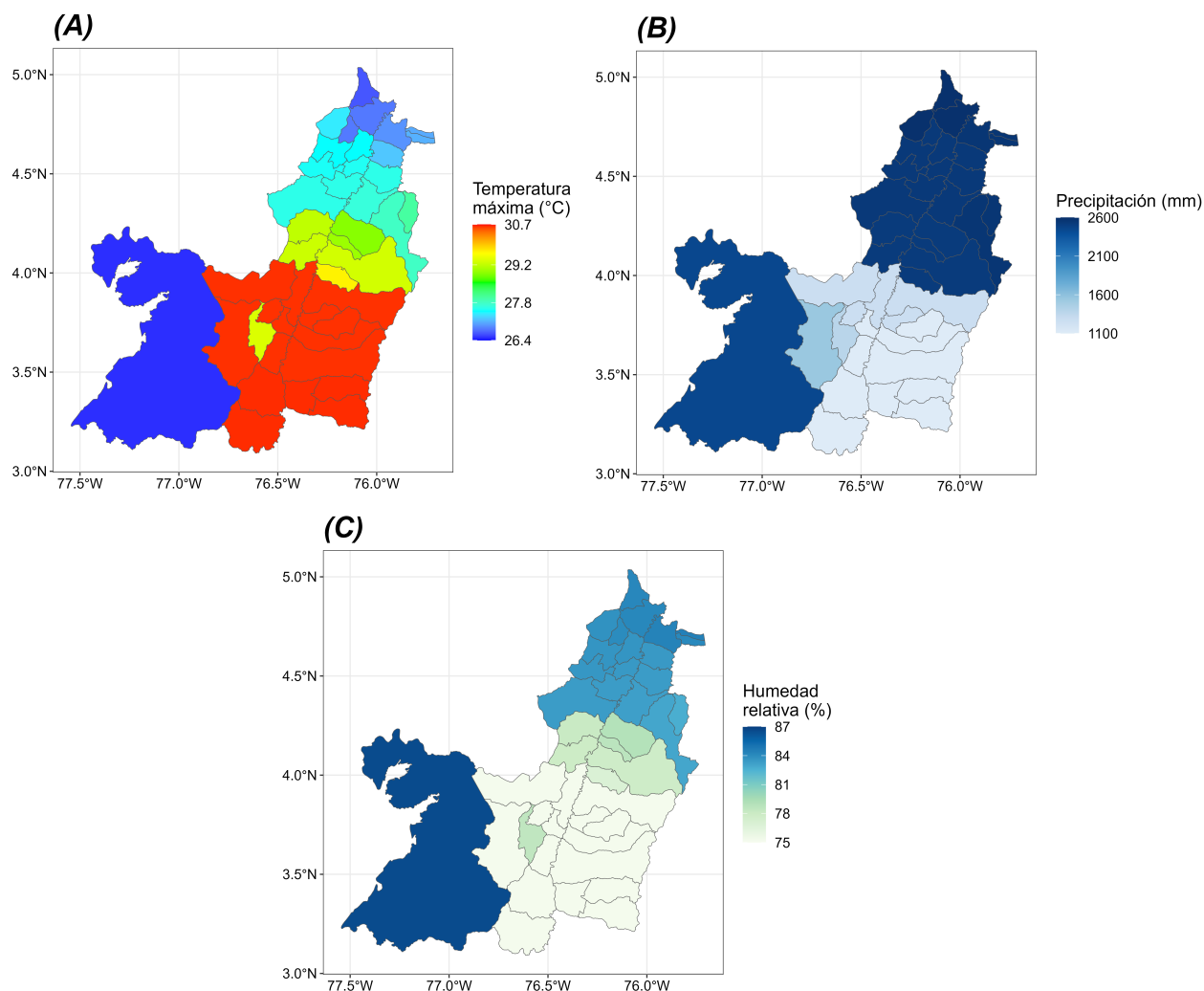


Figura 5-3.: Variables climáticas. (A) Temperatura máxima. (B) Precipitación. (C) Humedad relativa en los municipios del Valle del Cauca

Por último, en la Figura 5-3 (C) se observa el porcentaje de humedad relativa en cada municipio. De igual manera se presenta que Buenaventura registra un alto porcentaje de humedad relativa, alcanzando el 86.5 %. Se observa que en la zona sur del departamento la humedad relativa oscila entre 75 % y 77 %, en la zona centro la humedad es entre 78 % y 81 % y en la zona norte los municipios presentan valores de humedad más altos, entre 82 % y 87 %.

La correlación entre la tasa de dengue y las variables climáticas y geográficas se presenta en la Figura 5-4, en donde se puede observar que la temperatura máxima tiene una relación negativa fuerte con la humedad relativa y la precipitación. Es decir, en los municipios donde la temperatura es alta, se presenta humedad baja y poca precipitación, y viceversa. De igual forma, hay una relación positiva fuerte entre humedad y precipitación. Por el contrario, no hay relación entre el porcentaje de población urbana con ninguna de las variables climáticas. La tasa de dengue tiene una relación positiva con la temperatura máxima y el porcentaje de población urbana, por el contrario con la precipitación, humedad y altitud presenta una relación negativa.

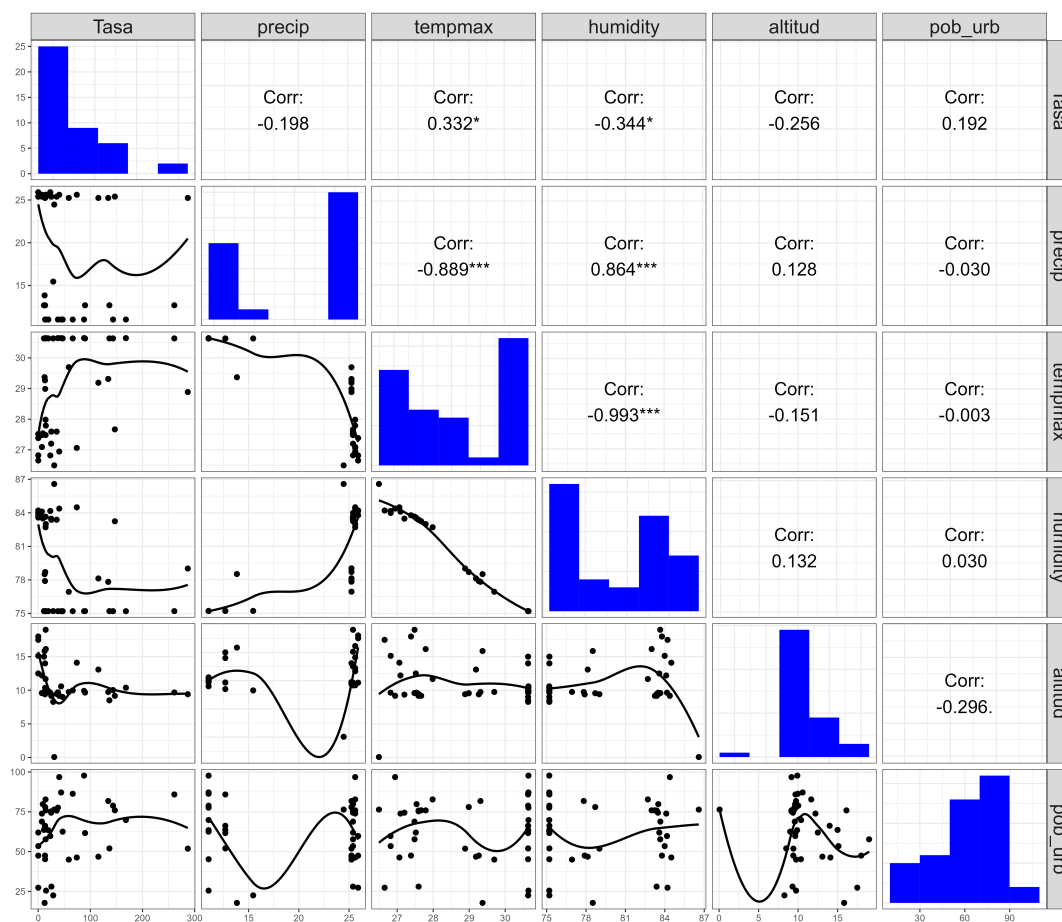


Figura 5-4.: Correlación entre la tasa de dengue y las variables climáticas y geográficas

5.2. Análisis espacial

La razón de incidencia estandarizada (SIR) se observa en la Figura 5-5 para los municipios del Valle del Cauca. Se puede notar que los municipios de Cali y Guacarí tienen un SIR cercano a 1, el cual indica que no hay diferencia entre el número de casos de dengue observados y el número de casos esperados. Para los municipios que presentan un SIR mayor a 1, el número de casos observados es mayor que los casos esperados, esta situación se presenta en los municipios de Bugalagrande, Buga, Florida, Zarzal, Palmira y Yotoco. La mayoría de los municipios presenta un SIR menor que 1, lo que indica que tienen menos casos observados de dengue de lo esperado.

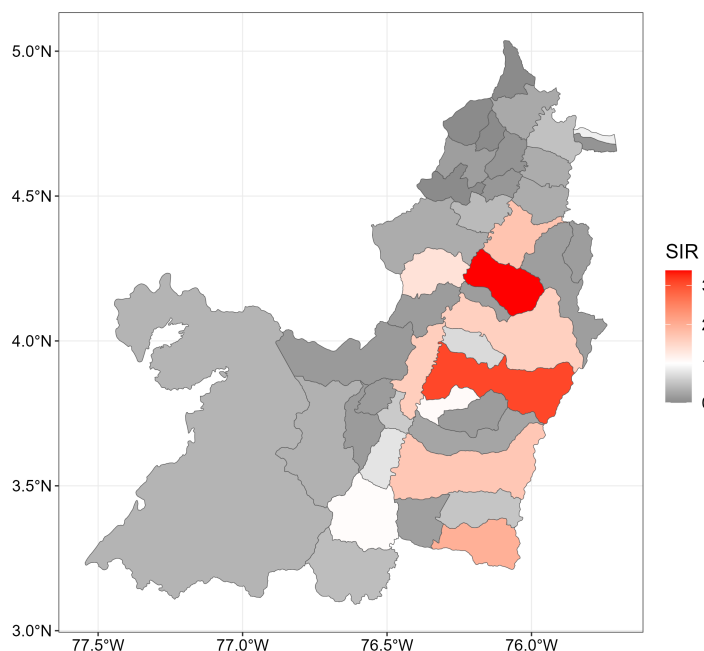


Figura 5-5.: Razón de incidencia estandarizada de dengue en los municipios del Valle del Cauca

En la Tabla 5-1 se presentan las estimaciones de los efectos fijos y aleatorios del modelo BYM2 ajustado (4-4), se observa que los intervalos de credibilidad de los efectos fijos de las variables climáticas se encuentran alrededor de cero, resalta la temperatura máxima que cuenta con una desviación estándar de 1.357 y un intervalo de credibilidad de (-2.473, 2.877). Este comportamiento se debe a la multicolinealidad entre las variables climáticas. Al eliminar la variable humedad del modelo, se presentan las estimaciones en la Tabla 5-2.

Tabla 5-1.: Estimación efectos fijos y aleatorios del modelo BYM2

Efectos Fijos	Media	Sd	Q_{2,5}	Q₅₀	Q_{97,5}
intercepto	4.870	75.956	-145.438	5.080	154.037
precip	0.069	0.066	-0.063	0.070	0.196
tempmax	0.193	1.357	-2.473	0.190	2.877
humidity	-0.155	0.456	-1.049	-0.157	0.748
pob_urb	0.014	0.008	-0.001	0.014	0.030
altitud	-0.131	0.059	-0.247	-0.130	-0.017
Efectos aleatorios					
τ_b	1.313	0.368	0.734	1.265	2.173
ϕ	0.172	0.164	0.007	0.116	0.618
Criterios de información					
Deviance Information Criterion (DIC)					247.46
Watanabe-Akaike information criterion (WAIC)					246.26

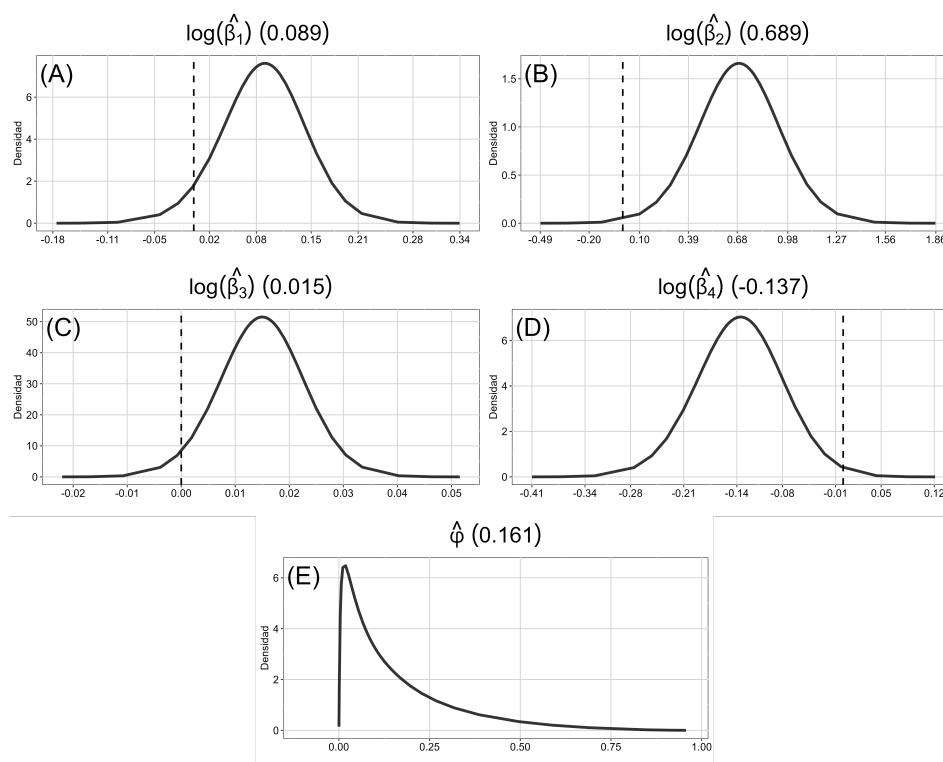
En la Tabla **5-2** se presentan los resultados del modelo ajustado sin la variable humedad. Se puede identificar que todas las covariables aportan información al modelo y se tiene una menor variabilidad en precipitación y temperatura máxima. En la Tabla **5-2** la temperatura máxima ahora cuenta con una desviación estándar de 0.247 y un intervalo de credibilidad de (0.199, 1.176). Además de esto, se observa una disminución en los criterios de información. En el modelo con la variables humedad se tiene un DIC igual a 247.46 y WAIC igual a 246.26, al eliminar esta variable los criterios ahora tienen los valores 245.98 y 244.06 respectivamente, lo que indica un mejor ajuste en el modelo.

Los parámetros estimados se encuentran en escala logarítmica, por lo que se deben transformar para realizar una correcta interpretación. Un aumento en la precipitación se asocia con un aumento en el riesgo relativo de los casos de dengue. Igualmente con la temperatura máxima y el porcentaje de población urbana se relacionan con un incremento en el riesgo relativo a medida que estas aumentan. Por último, los municipios con mayor altitud cuentan con un menor riesgo relativo. Estos resultados obtenidos son consistentes de acuerdo a los antecedentes revisados Lowe et al. (2011) y Phanitchat et al. (2019).

En la Figura **5-6** se observa que las distribuciones a posteriori no se encuentran centradas en cero, es decir que los datos han proporcionado evidencia suficiente para determinar que los factores incluidos en el modelo tienen un aporte significativo. Adicionalmente en la Figura **5-6** (E) se observa la distribución a posteriori de la proporción de varianza marginal explicada por el efecto estructurado ($\phi = 0.161$), indicando que el modelo es capaz de capturar la correlación espacial entre los municipios.

Tabla 5-2.: Estimación de los efectos fijos y aleatorios del modelo BYM2 (Modelo sin la variable humedad)

Efectos Fijos	Media	Sd	Q _{2,5}	Q ₅₀	Q _{97,5}
intercepto	-22.187	8.079	-38.038	-22.218	-6.157
precip	0.089	0.054	-0.020	0.090	0.194
tempmax	0.689	0.247	0.199	0.690	1.176
pob_urb	0.015	0.008	-0.001	0.015	0.030
altitud	-0.137	0.058	-0.252	-0.136	-0.024
Efectos aleatorios					
τ_b	1.325	0.377	0.726	1.278	2.199
ϕ	0.161	0.156	0.007	0.109	0.589
Criterios de información					
Deviance Information Criterion (DIC)					245.98
Watanabe-Akaike information criterion (WAIC)					244.06

**Figura 5-6.:** Estimación de la distribución a posteriori marginal de los parámetros (A-D) e hiperparámetro (E) del modelo sin la variable humedad (media a posteriori en paréntesis)

En la Figura 5-7 (A) se presenta el riesgo relativo estimado de dengue para los municipios del Valle del Cauca. Se observa que los municipios de Bugalagrande y Buga presentan mayor riesgo relativo, seguido de Zarzal, Tuluá, Palmira y Florida. En la Figura 5-7 (B) se muestra la estimación de la probabilidad a posteriori de que el riesgo relativo sea mayor a 1, en este caso los municipios con una probabilidad cercana a 1, es muy probable que el riesgo relativo sea mayor a 1. En los municipios con probabilidades cercanas a 0, es muy poco probable que el riesgo exceda 1. Si la probabilidad se encuentra alrededor de 0.5 se tiene mayor incertidumbre y corresponde a municipios donde el riesgo relativo esta por debajo o por encima de 1 con igual probabilidad. Los municipios con una alta probabilidad a posteriori de exceder el riesgo relativo de 1 son Zarzal, Bugalagrande, Tuluá, Buga, Yotoco, Palmira, Florida y Cali (Tabla B-2), los cuales están agrupados en su mayoría en la zona oriente del departamento.

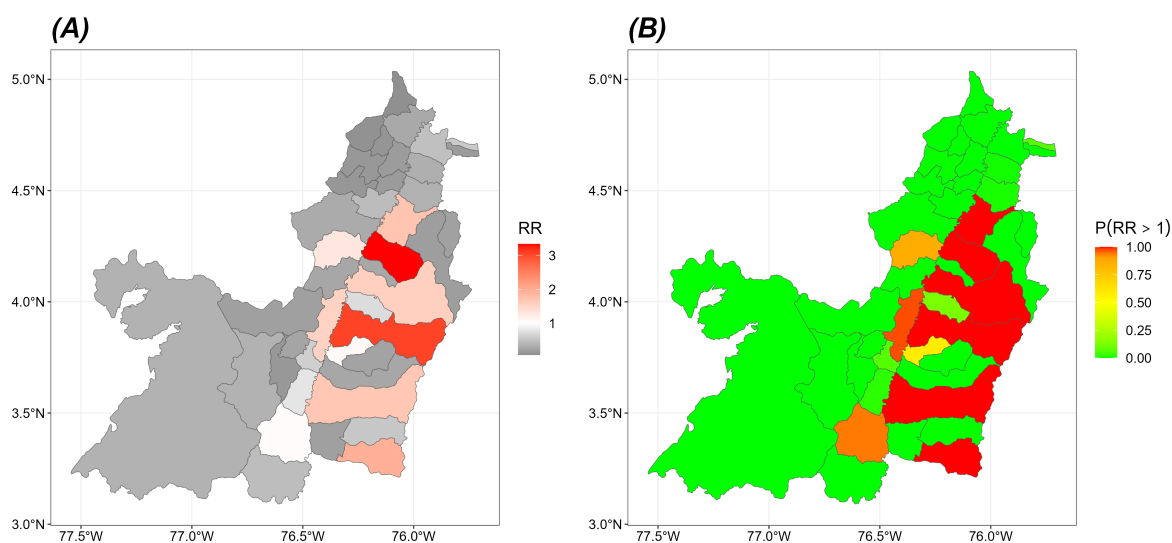


Figura 5-7.: (A) Riesgo relativo de dengue para los municipios del Valle del Cauca
(B) Probabilidad del riesgo relativo mayor a 1

6. Conclusiones, limitaciones y recomendaciones

6.1. Conclusiones

En este estudio se aborda la enfermedad del dengue entre los municipios del Valle del Cauca para el año 2019, se modela el número de casos de dengue en relación a variables climáticas y geográficas mediante un modelo Poisson con efectos aleatorios. La correlación espacial entre los municipios fue estimada a través del modelo BYM2. Esta proporción de variabilidad explicada por el componente espacial estructurado del modelo tiene un valor estimado de ϕ igual a 0.161, lo que implica que el modelo es capaz de capturar algo de la correlación espacial presente entre los municipios.

Los resultados muestran que la zona oriente del departamento presenta un mayor número de casos de dengue por 100,000 habitantes. Los municipios de esta zona se caracterizan por tener una altitud alrededor de 1000 metros sobre el nivel del mar, así como temperaturas más elevadas y una precipitación moderada. Esto se debe al efecto que tienen las condiciones climáticas sobre el riesgo de infección de dengue. Los municipios con mayor temperatura, precipitación, porcentaje de población urbana, y altitud alrededor de los 1000 msnm presentan un riesgo más elevado de infección de dengue. Mientras que, aquellos ubicados a mayor altitud, con menores temperaturas y con un porcentaje de población urbana, menor muestran riesgos bajos. Estos resultados resaltan la importancia de considerar los factores climáticos y geográficos al evaluar el riesgo de dengue en la región, lo cual permite dirigir estrategias preventivas y de control de manera más precisa y efectiva.

Lo anterior coincide con lo expresado por los autores Lowe et al. (2011), quienes mencionan que la precipitación y temperatura son estadísticamente significativas y se relacionan positivamente con el dengue. Esto puede ser el resultado de condiciones cálidas/húmedas que promueven el desarrollo del mosquito, vector de la enfermedad del dengue. También se identificaron como significativas la altitud y el porcentaje de población urbana en la ocurrencia de casos de dengue, lo que es previsto al influir en el desarrollo y transmisión de la enfermedad en las comunidades humanas.

El ajuste del modelo permitió identificar y caracterizar las zonas o municipios que presentan un mayor riesgo de infección de dengue. Esto es fundamental para implementar sistemas de alerta temprana por parte de los entes locales, como alcaldías y gobernaciones. Los mapas de riesgo estimado identifican las áreas más propensas a la transmisión de dengue y permiten a las autoridades locales tomar decisiones informadas sobre los programas de prevención y control. Esto incluye la implementación de programas de eliminación de criaderos de mosquitos, campañas de concientización, fumigación focalizada y fortalecimiento de la atención médica en las zonas de mayor riesgo. Además, al identificar los factores climáticos y utilizar el riesgo estimado, se pueden tomar medidas preventivas oportunas, protegiendo la salud de la comunidad y reduciendo la ocurrencia de esta enfermedad transmitida por mosquitos.

6.2. Limitaciones

La fuente utilizada para obtener la información de los datos climáticos presentaba ciertas limitaciones, como un número reducido de estaciones meteorológicas, que además la mayoría no se encuentran ubicadas en el Valle del Cauca. Esta limitación en la calidad de los datos climáticos podría afectar la precisión y la representatividad de los resultados obtenidos en el análisis. Igualmente se exploraron datos climáticos de imágenes satelitales, sin embargo se encontró que estos datos presentaban valores menores a los esperados en comparación con los estándares climáticos establecidos. Debido a esto, se tomó la decisión de descartar los datos de imágenes satelitales y enfocarse en los datos provenientes de las estaciones climáticas disponibles. A pesar de que los datos de las estaciones climáticas presentaban una limitación en la cobertura geográfica, se consideró que esta fuente era más confiable y representativa en relación con las condiciones climáticas locales.

6.3. Recomendaciones

Se recomienda para futuras investigaciones trabajar con una fuente de datos climáticos más confiable y con datos completos en términos de cobertura espacial que permita una comprensión más completa de los patrones climáticos. Adicionalmente, se podrían incluir en el análisis otras variables geográficas y sociodemográficas que tengan influencia en la ocurrencia de los casos de dengue.

El dengue es una enfermedad que muestra un comportamiento estacional y presenta fluctuaciones en la incidencia a lo largo del tiempo. Por lo tanto, se recomienda estudiar el componente temporal para mejorar la comprensión de la dinámica del dengue, identificar patrones estacionales, cambios a largo plazo y posibles asociaciones con los factores climáticos. Además, el análisis temporal puede ayudar a predecir periodos de mayor riesgo,

lo que permitiría una mejor planificación y respuesta por parte de las autoridades de salud. Esto se puede realizar por medio de una extensión del modelo BYM en un contexto espacio-temporal, incorporando un efecto aleatorio que modele el componente temporal de los datos.

Riebler et al. (2016) realiza un estudio de simulación concluyendo que los parámetros estimados por el modelo BYM2 son bastante cercanos a los reales bajo ciertos escenarios y considera que es un modelo robusto a la elección de la distribución a priori para el hiperparámetro mixto ϕ . Por lo tanto, se recomienda realizar en futuros estudios simulaciones que permitan explorar diversas situaciones y escenarios con el objetivo de evaluar la robustez de las estimaciones. Si bien los resultados obtenidos desde un enfoque descriptivo son relevantes, es fundamental analizar las estimaciones desde una perspectiva epidemiológica, considerando la variación de variables y covariables. Esto puede lograrse mediante la modificación de estas variables para observar su influencia en las estimaciones. Estas simulaciones enriquecerán el estudio al proporcionar una comprensión más completa de la dinámica de la enfermedad, mejorando así su aplicabilidad en la toma de decisiones y políticas de salud pública.

A. Anexo: Municipios del Valle del Cauca por subregión

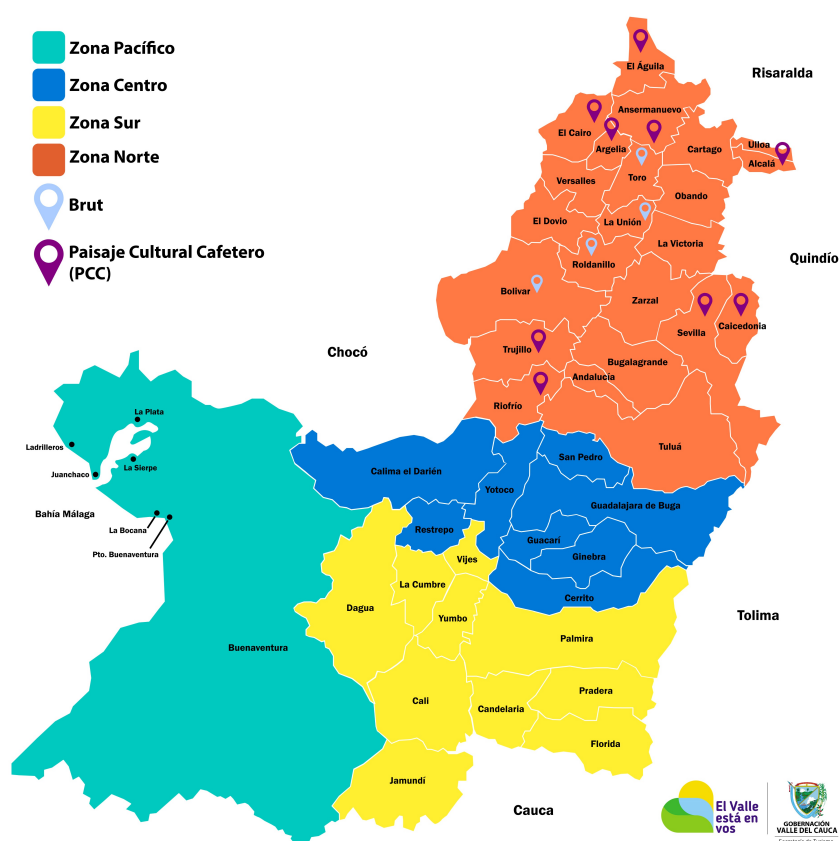


Figura A-1.: Municipios del Valle del Cauca por subregión. Fuente: Gobernación del Valle del Cauca

Tabla A-1.: Municipios del Valle del Cauca y las estaciones meteorológicas con las que fueron relacionados (Estaciones crecanas en un rango de 50-100 km)

Municipio	ID Estaciones
El Águila	80210099999, 80211099999, SKPE, SKAR, SKMZ
Ansermanuevo	80222799999, 80210099999, 80211099999, SKPE, SKAR
El Cairo	80222799999, 80210099999, 80211099999,SKPE
Argelia	80222799999, 80210099999, 80211099999, SKPE, SKAR
Cartago	80222799999, 80210099999, 80211099999, SKPE, SKAR
Alcalá	80210099999, 80222799999, 80211099999, SKPE, SKAR
Obando	80222799999, 80211099999, SKAR ,80210099999, SKPE
Toro	80222799999, 80210099999, 80211099999, SKPE, SKAR
Versalles	80222799999, 80210099999, 80211099999, SKPE, SKAR
El Dovio	80222799999, 80210099999, 80211099999, SKPE, SKAR
La Unión	80222799999, 80210099999, 80211099999, SKPE, SKAR
Roldanillo	80210099999, 80211099999, 80222799999, SKPE, SKAR
La Victoria	80222799999, 80210099999, 80211099999, SKPE, SKAR
Zarzal	80210099999, 80211099999, 80222799999, SKPE, SKAR
Caicedonia	80211099999, 80222799999, 80210099999, SKPE, SKAR
Bugalagrande	80259099999, 80222799999, SKCL, 80211099999, SKAR
Andalucía	80222799999, 80259099999, SKCL, 80211099999, SKAR
Bolívar	80210099999, 80211099999, 80222799999, SKPE, SKAR
Trujillo	80259099999, 80222799999, 80211099999, SKCL, SKAR
Riofrío	80211099999, 80259099999, 80222799999, SKCL, SKAR
Tuluá	80211099999, 80259099999, SKCL, SKAR
San Pedro	80211099999, 80259099999, SKCL, SKAR
Buga	80259099999, 80252099999, SKCL
Calima (El Darién)	80259099999, 80252099999, SKCL, SKBU
Restrepo	80259099999, 80252099999, SKCL, SKBU
Yotoco	80259099999, 80252099999, SKCL, SKBU
Vijes	80259099999, 80252099999, SKCL, SKBU
Guacarí	80259099999, 80252099999, SKCL, SKBU
Ginebra	80259099999, SKCL
El Cerrito	80259099999, 80252099999, SKCL
Palmira	80259099999, SKCL
Candelaria	80259099999, SKCL
Florida	80259099999, SKCL
Yumbo	80259099999, 80252099999, SKCL, SKBU
Cali	80259099999, 80252099999, SKCL, SKBU
Dagua	80259099999, 80252099999, SKCL, SKBU
La Cumbre	80259099999, 80252099999, SKCL, SKBU
Jamundí	80259099999, 80252099999, SKCL
Pradera	80259099999, SKCL
Ulloa	80210099999, 80222799999, 80211099999, 80149099999, SKPE
Sevilla	80210099999, 80211099999, 80222799999, SKPE, SKAR
Buenaventura	80252099999, SKBU

B. Anexo: Análisis espacial

Tabla B-1.: Tasa de infección de dengue para los municipios del Valle del Cauca

Municipio	Número de casos	Tasa de infección
Alcalá	1	7.04
Andalucía	3	13.32
Ansermanuevo	4	23.05
Argelia	0	0
Bolívar	4	25.27
Buenaventura	95	30.62
Buga	335	261.07
Bugalagrande	70	286.88
Caicedonia	4	13.98
Cali	1970	87.88
Calima (El Darién)	2	11.006
Candelaria	14	15.02
Cartago	55	40.26
Dagua	14	28.71
El Águila	0	0
El Cairo	0	0
El Cerrito	11	19.35
El Dovio	0	0
Florida	97	168.12
Ginebra	3	13.09
Guacarí	30	89.80
Jamundí	62	37.76
La Cumbre	2	12.17
La Unión	3	8.76
La Victoria	3	25.03
Obando	3	24.71
Palmira	505	143.45
Pradera	21	43.82
Restrepo	2	13.07
Riofrío	2	12.91
Roldanillo	13	35.35
San Pedro	10	58.54
Sevilla	6	14.45
Toro	1	6.91
Trujillo	22	115.34
Tuluá	290	134.27
Ulloa	4	73.90
Versalles	1	14.14
Vijes	6	46.81
Yotoco	22	136.50
Yumbo	72	66.12
Zarzal	62	146.95

Tabla B-2.: Resultados del modelo BYM2

Municipio	Media	Q _{2,5}	Q ₅₀	Q _{97,5}	$P(\hat{\mu}_i > 1)$
El Águila	0.044	0.0034	0.0266	0.1901	0.0002
Ansermanuevo	0.249	0.0913	0.2240	0.5509	0.0006
El Cairo	0.0759	0.0087	0.0519	0.2865	0.0004
Argelia	0.0762	0.0092	0.0528	0.2829	0.0003
Cartago	0.4752	0.3623	0.4710	0.6123	0
Alcalá	0.1476	0.0366	0.1231	0.401	0.0003
Obando	0.3192	0.1097	0.2838	0.734	0.0047
Toro	0.169	0.0467	0.1452	0.429	0.0002
Versalles	0.124	0.0185	0.0897	0.4333	0.0014
El Dovio	0.1247	0.0232	0.0985	0.38	0.0004
La Unión	0.1612	0.0632	0.1477	0.337	0
Roldanillo	0.4247	0.2439	0.4097	0.6896	0.0004
La Victoria	0.357	0.1275	0.3203	0.7989	0.0073
Zarzal	1.724	1.3303	1.7088	2.1978	1
Caicedonia	0.2166	0.0875	0.1991	0.447	0
Bugalagrande	3.389	2.6539	3.3635	4.2643	1
Andalucía	0.2797	0.1125	0.2579	0.5721	0.0004
Bolívar	0.2825	0.1037	0.2544	0.6242	0.0014
Trujillo	1.2916	0.8273	1.2619	1.9249	0.8595
Riofrío	0.279	0.0968	0.2508	0.6246	0.0014
Tuluá	1.5829	1.4085	1.5802	1.7728	1
San Pedro	0.7433	0.3968	0.7111	1.2734	0.1253
Buga	3.0658	2.7499	3.0612	3.4077	1
Calima (El Darién)	0.1954	0.0642	0.1731	0.4554	0.0002
Restrepo	0.2459	0.0812	0.2181	0.5722	0.001
Yotoco	1.549	0.9973	1.5152	2.2995	0.9741
Vijes	0.605	0.2765	0.5662	1.1576	0.0592
Guacarí	1.0575	0.7303	1.0406	1.4822	0.5859
Ginebra	0.2268	0.0863	0.2064	0.4856	0.0001
El Cerrito	0.2594	0.1463	0.2502	0.4258	0
Palmira	1.6879	1.5453	1.6862	1.84	1
Candelaria	0.1945	0.1154	0.1887	0.3077	0
Florida	1.9588	1.596	1.9487	2.3793	1
Yumbo	0.787	0.6216	0.7816	0.9829	0.0174
Cali	1.036	0.9909	1.0357	1.0825	0.9401
Dagua	0.3654	0.2156	0.3544	0.5806	0
La Cumbre	0.1233	0.0279	0.0998	0.3572	0,0002
Jamundí	0.4546	0.3528	0.4511	0.5767	0
Pradera	0.5425	0.3517	0.5307	0.8003	0.0012
Ulloa	0.4839	0.1315	0.4094	1.2685	0.0608
Sevilla	0.184	0.0827	0.1717	0.3562	0
Buenaventura	0.3605	0.2935	0.3589	0.4381	0

Bibliografía

Agresti, A. (2015), *Foundations of Linear and Generalized Linear Models*, John Wiley & Sons.

Anderson, C. J., Verkuilen, J. & Johnson, T. R. (2013), *Applied Generalized Linear Mixed Models: Continuous and Discrete Data*.

Andrieu, C., De Freitas, N., Doucet, A. & Jordan, M. I. (2003), ‘An introduction to mcmc for machine learning’, *Machine learning* **50**, 5–43.

Banerjee, S., Carlin, B. P. & Gelfand, A. E. (2014), *Hierarchical Modeling and Analysis for Spatial Data*, CRC press.

Besag, J., York, J. & Mollié, A. (1991), ‘Bayesian image restoration, with two applications in spatial statistics’, *Annals of the institute of statistical mathematics* **43**, 1–20.

Blangiardo, M. & Cameletti, M. (2015), *Spatial and spatio-temporal Bayesian models with R-INLA*, John Wiley & Sons.

DANE (2005), ‘Ficha técnica Sistema de información del medio ambiente’, *Departamento Administrativo Nacional de Estadística (DANE)* .

URL: <https://www.dane.gov.co/index.php/estadisticas-por-tema/ambientales/indicadores-ambientales-iaii/iniciativa-sima>

Diniz-Filho, J. A. F., Bini, L. M. & Hawkins, B. A. (2003), ‘Spatial autocorrelation and red herrings in geographical ecology’, *Global ecology and Biogeography* **12**(1), 53–64.

Filho, A. S., Murari, T. B., Ferreira, P., Saba, H. & Moret, M. A. (2021), ‘A spatio-temporal analysis of dengue spread in a Brazilian dry climate region’, *Scientific Reports* **11**.

Gobernación del Valle del Cauca (2020), ‘Mapas y Guías Turísticas’.

URL: www.valledelcauca.gov.co/turismo/publicaciones/60157/mapas-y-guias-turisticas/

IDEAM (2018), ‘Instituto de Hidrología, Meteorología y Estudios Ambientales. Atlas climatológico de Colombia - IDEAM’.

URL: <http://www.ideam.gov.co/web/tiempo-y-clima/tiempo-clima>

Kuhn, K., Campbell-Lendrum, D., Haines, A., Cox, J., Corvalán, C. & Anker, M. (2005), ‘Using climate to predict infectious disease epidemics’, *World Health Organization* .

- Lawson, A. (2013), *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology, Second Edition*, Chapman & Hall/CRC Interdisciplinary Statistics, CRC Press.
- Lowe, R., Bailey, T. C., Stephenson, D. B., Graham, R. J., Coelho, C. A., Carvalho, M. S. & Barcellos, C. (2011), ‘Spatio-temporal modelling of climate-sensitive disease risk: Towards an early warning system for dengue in Brazil’, *Computers and Geosciences* **37**, 371–381.
- Márquez-Benítez, Y., Monroy-Cortés, K. J., Martínez-Montenegro, E. G., Peña-García, V. H. & Monroy-Díaz, A. L. (2019), ‘Influencia de la temperatura ambiental en el mosquito *Aedes* spp y la transmisión del virus del dengue’, *Rev CES Med* **33**(1), 42–50.
- Ministerio de Salud (2021), ‘Colombia, en alerta por repunte de dengue’, *Unidad de salud. El Tiempo*.
- Molenberghs, G. & Verbeke, G. (2005), *Models for Discrete Longitudinal Data*, Springer Series in Statistics, Springer New York.
- Montgomery, D. C., Peck, E. A. & Vining, G. G. (2012), *Introduction to Linear Regression Analysis*, Vol. 5.
- Moraga, P. (2019), *Geospatial Health Data: Modeling and Visualization with R-INLA and Shiny*, Chapman Hall/CRC Biostatistics Series.
- Moreno-Altamirano, A., López-Moreno, S. & Corcho-Berdugo, A. (2007), ‘Principales medidas en epidemiología’, *Revista Cubana de Higiene y Epidemiología*.
URL: <https://www.redalyc.org/articulo.oa?id=223219928011>
- Muñoz, E., Poveda, G., Arbeláez, M. P. & Vélez, I. D. (2020), ‘Spatiotemporal dynamics of dengue in Colombia in relation to the combined effects of local climate and ENSO’.
URL: <https://doi.org/10.1101/2020.08.24.20181032>
- Nigussie, T. Z., Zewotir, T. & Muluneh, E. K. (2022), ‘Effects of climate variability and environmental factors on the spatiotemporal distribution of malaria incidence in the Amhara national regional state, Ethiopia’, *Spatial and Spatio-temporal Epidemiology* **40**, 100475.
- OMS (2022), ‘Dengue y dengue grave’.
URL: <https://www.who.int/es/news-room/fact-sheets/detail/dengue-and-severe-dengue>
- Ordoñez-Sierra, G., Sarmiento-Senior, D., Gomez, J. F. J., Giraldo, P., Ramírez, A. P. & Olano, V. A. (2021), ‘Multilevel analysis of social, climatic and entomological factors that influenced dengue occurrence in three municipalities in Colombia’, *One Health* **12**.

- Ortiz, C. R., Rúa-Uribe, G., Suárez, C. A., Mafla, M., Almanza, R. & Santos, S. L. D. (2013), 'Distribución espacial de casos e incidencia de dengue: Análisis de la situación para Medellín, Colombia', *Rev. Fac. Nac. Salud Pública.* .
- Pérez Porto, J. & Gardey, A. (2014), 'Temperatura atmosférica. qué es, definición y concepto'.
URL: <https://definicion.de/temperatura-atmosferica/>
- Phanitchat, T., Zhao, B., Haque, U., Pientong, C., Ekalaksananan, T., Aromseree, S., Thaewongiew, K., Fustec, B., Bangs, M. J., Alexander, N. & Overgaard, H. J. (2019), 'Spatial and temporal patterns of dengue incidence in northeastern Thailand 2006-2016', *BMC infectious diseases* **19**, 1–12.
- Poveda, G., Rojas, W., Quiñones, M. L., Vélez, I. D., Mantilla, R. I., Ruiz, D., Zuluaga, J. S. & Rua, G. L. (2001), 'Coupling between annual and ENSO timescales in the Malaria-Climate Association in Colombia', *Environmental Health Perspectives* .
URL: <http://ehpnet1.niehs.nih.gov/docs/2001/109p489-493poveda/abstract.html>
- Riebler, A., Sørbye, S. H., Simpson, D. & Rue, H. (2016), 'An intuitive Bayesian spatial model for disease mapping that accounts for scaling', *Statistical methods in medical research* **25**(4), 1145–1165.
- Rue, H., Martino, S. & Chopin, N. (2009), 'Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations', *Journal of the Royal Statistical Society: Series b (statistical methodology)* **71**(2), 319–392.
- Segurado, P., Araujo, M. B. & Kunin, W. (2006), 'Consequences of spatial autocorrelation for niche-based models', *Journal of Applied Ecology* **43**(3), 433–444.
- Simpson, D., Rue, H., Riebler, A., Martins, T. G. & Sørbye, S. H. (2015), 'Penalising model component complexity: A principled, practical approach to constructing priors'.
- Van den Broeck, J., Brestoff, J. R. & Kaulfuss, C. (2013), *Epidemiology: principles and practical guidelines*, Springer.
- Zuur, A. F., Ieno, E. N., Smith, G. M. et al. (2007), *Analysing ecological data*, Vol. 680, Springer.