

APLICACIÓN DEL MÉTODO DE BAYES EN UN TRAMO DE ALTA
ACCIDENTALIDAD EN EL ÁREA METROPOLITANA DE BUCARAMANGA

Andrés Felipe Corredor Rodríguez

Trabajo de Grado para Optar al Título de Ingeniero Civil

Director

Miller Humberto Salas Rondón

Doctor en Gestión del Territorio e Infraestructuras del Transporte

Universidad Industrial de Santander

Facultad de Ingenierías Fisicomecánicas

Escuela de Ingeniería Civil

Ingeniería Civil

Bucaramanga

2025

Agradecimientos

Deseo expresar mi agradecimiento al profesor, Miller Salas, por su guía y valiosas aportaciones, que resultaron fundamentales para la realización de este estudio. También agradezco a mis padres, cuyo apoyo emocional y constante ánimo me permitieron culminar con éxito este proyecto.

Tabla de Contenido

	Pág.
Introducción	10
1. Objetivos	12
1.1 Objetivo general	12
1.2 Objetivos específicos	12
2. Antecedentes de metodologías en la identificación de tramos de alta accidentalidad.....	13
3. Datos	23
3.1 Obtención de datos.....	23
3.2 Identificación tramo de estudio.....	26
4. Implementación del modelo.....	28
4.1 Distribución de Poisson	29
4.2 Teorema de Bayes.....	30
4.3 Cadenas de Markov Monte Carlo (MCCM).....	32
4.4 Construcción del modelo	33
5. Resultados del análisis	37
5.1 Validación del modelo	37
5.2 Puntos con alta probabilidad de accidentes	39
6. Conclusiones	46
7. Recomendaciones	47
Referencias Bibliográficas	49
Apéndices.....	55

Lista de Tablas

	Pág.
Tabla 1 <i>Puntos de alta accidentalidad en el tramo estudiado</i>	39

Lista de Figuras

	Pág.
Figura 1 <i>Cifras anuales víctimas fatales de accidentes automovilísticos en el periodo 2013 - 2023</i>	10
Figura 2 <i>Nube de palabras más usadas en los artículos de la búsqueda bibliográfica</i>	14
Figura 3 <i>Temas en tendencia a lo largo de los años en los artículos de la búsqueda bibliográfica</i>	15
Figura 4 <i>Acercamiento a figura 3 gráfica temas en tendencia a lo largo de los años</i>	16
Figura 5 <i>Países más citados en los artículos de la búsqueda bibliográfica</i>	16
Figura 6 <i>Acercamiento a los primeros 5 países más citados en la gráfica países más citados en los artículos de la búsqueda bibliográfica</i>	16
Figura 7 <i>Países de los correspondientes autores en los artículos de la búsqueda bibliográfica.</i>	17
Figura 8 <i>Acercamiento a los primeros 5 países de los correspondientes autores en los artículos de la búsqueda bibliográfica</i>	17
Figura 9 <i>Ubicación tramo de estudio con respecto al área metropolitana de Bucaramanga</i>	27
Figura 10 <i>Acercamiento a ubicación tramo de estudio con respecto al área metropolitana de Bucaramanga.</i>	28
Figura 11 <i>Vista polígono imaginario y ubicación tramo de estudio</i>	34
Figura 12 <i>Imagen satelital con la delimitación del tramo de estudio y los puntos de alta accidentalidad debidamente georreferenciados</i>	40

Figura 13 <i>Acercamiento a imagen satelital con la delimitación del tramo de estudio y los puntos de alta accidentalidad debidamente georreferenciados</i>	40
Figura 14 <i>Imagen satelital del punto de alta accidentalidad más crítico en el tramo de estudio debidamente georreferenciados</i>	41
Figura 15 <i>Acercamiento a imagen satelital del punto de alta accidentalidad más crítico en el tramo de estudio debidamente georreferenciados</i>	42
Figura 16 <i>Imagen satelital del segundo punto de alta accidentalidad más crítico en el tramo de estudio debidamente georreferenciados</i>	43
Figura 17 <i>Acercamiento a imagen satelital del segundo punto de alta accidentalidad más crítico en el tramo de estudio debidamente georreferenciados</i>	43
Figura 18 <i>Imagen satelital del tercer punto de alta accidentalidad más crítico en el tramo de estudio debidamente georreferenciados</i>	44
Figura 19 <i>Acercamiento a imagen satelital del tercer punto de alta accidentalidad más crítico en el tramo de estudio debidamente georreferenciados</i>	45

Lista de Apéndices

	pág.
Apéndice A. Código completo del tratamiento de datos realizado en Google Collab	55
Apéndice B. Código completo del modelo desarrollado mediante Rstudio	56

Resumen

Título: Aplicación del método de bayes en un tramo de alta accidentalidad en el área metropolitana de Bucaramanga*

Autor: Andrés Felipe Corredor Rodríguez**

Palabras Clave: Alta accidentalidad, método de Bayes, seguridad vial, modelo estadístico, gestión del riesgo, tráfico urbano.

Descripción: En este estudio se identifican los puntos con alta accidentalidad dentro de un tramo específico de la malla vial del área metropolitana de Bucaramanga, empleando el método de Bayes como herramienta de análisis. Para ello, se llevó a cabo una modelación estadística mediante el empleo del software de uso libre RStudio, lo que permitió analizar los puntos con mayores índices de siniestralidad y comprender su distribución dentro de la zona de estudio. A partir de los resultados obtenidos, se evidenció que ciertos sectores de la malla vial, como las intersecciones y los desvíos, presentan una mayor concentración de accidentes. Asimismo, se identificó que los factores viales más vulnerables en dichos eventos son los peatones y los motociclistas, quienes resultan ser los actores más afectados dentro de los escenarios de siniestralidad analizados. Esta información es de gran utilidad, ya que permite una mejor comprensión de los factores de riesgo asociados a la accidentalidad en el área de estudio. Además, los hallazgos derivados de este análisis proporcionan una herramienta clave para la toma de decisiones en materia de seguridad vial, facilitando el futuro diseño de estrategias orientadas a reducir la ocurrencia de accidentes y mejorar la seguridad de los usuarios viales en el tramo analizado.

* Trabajo de Grado

** Facultad de Ingenierías Fisicomecánicas. Escuela de Ingeniería Civil. Ingeniería Civil. Director: Miller Humberto Salas Rondón. Doctor en Gestión del Territorio e Infraestructuras del Transporte

Abstract

Title: Application of the Bayesian Method to a High-Accident Section in the Metropolitan Area of Bucaramanga*

Author: Andrés Felipe Corredor Rodríguez**

Key Words: High Accident Rates, Bayesian Method, Road Safety, Statistical Modeling, Risk Management, Urban Traffic.

Description: This study identifies high-accident locations within a specific section of the road network in the metropolitan area of Bucaramanga using the Bayesian method as an analytical tool. To achieve this, a statistical modeling process was conducted through the use of the open-source software RStudio. This approach allowed for the identification and analysis of locations with the highest accident rates, providing a better understanding of their spatial distribution within the study area. The results obtained indicate that certain sections of the road network, particularly intersections and detours, exhibit a higher concentration of accidents. Additionally, the analysis revealed that the most vulnerable road users in these incidents are pedestrians and motorcyclists, who constitute the most affected groups in the accident scenarios examined. These findings are highly valuable, as they contribute to a deeper understanding of the risk factors associated with road accidents in the study area. Moreover, the insights gained from this study provide an essential tool for informed decision-making in road safety management. By leveraging these results, policymakers and transportation authorities can develop and implement targeted strategies aimed at mitigating accident risks. This, in turn, enhances the overall safety of road users and contributes to the reduction of traffic incidents within the analyzed section.

* Undergraduate Thesis

** Faculty of Physicomechanical Engineering. School of Civil Engineering. Civil Engineering Program. Advisor: Miller Humberto Salas Rondón, Ph.D. in Land Management and Transportation Infrastructure

Introducción

La elevada accidentalidad vial causa no solo pérdidas humanas irremplazables, sino también impone un significativo costo económico a las sociedades. Según la Organización Mundial de la Salud (OMS), los accidentes de tránsito provocan alrededor de 1.19 millones de muertes al año (World Health Organization, 2023), un problema que también se refleja en Colombia, donde en la red vial de aproximadamente 205.000 kilómetros (Ministerio de Transporte, 2021), se registran un número creciente de víctimas fatales año tras año de acuerdo a datos de la Agencia Nacional de Seguridad Vial (ANSV) (Agencia Nacional de Seguridad Vial, 2024a), tal como se muestra en la figura 1.

Figura 1

Cifras anuales víctimas fatales de accidentes automovilísticos en el periodo 2013 - 2023



Nota. El gráfico representa las cifras anuales de víctimas fatales de accidentes automovilísticos en Colombia en el periodo 2013 – 2023. Tomado de Agencia Nacional de Seguridad Vial (2024a).

Para esta problemática existen diversas medidas correctivas, que van desde ajustes técnicos para el mejoramiento de las vías, hasta la implementación de campañas de seguridad (Balaji et al., 2024). Sin embargo, un paso crucial en la reducción de accidentes es identificar con precisión los puntos más peligrosos en la infraestructura vial para poder corregirlos.

En este estudio, se concentrará en la identificación y clasificación de puntos de alta accidentalidad presentes en un tramo de carretera del área metropolitana de Bucaramanga. Al establecer una jerarquía de los puntos más peligrosos, se busca proporcionar una herramienta que permita priorizar las intervenciones necesarias para reducir la siniestralidad y mejorar la seguridad en estos tramos de alto riesgo.

1. Objetivos

1.1 Objetivo general

- Determinar los puntos críticos viales en un tramo de alta accidentalidad del Área Metropolitana de Bucaramanga a partir del método de Bayes.

1.2 Objetivos específicos

- Definir la pertinencia de la metodología de Bayes para la evaluación de tramos con alta frecuencia de accidentes a partir de una revisión bibliográfica.
- Identificar el tramo de malla vial del Área Metropolitana de Bucaramanga a partir de un análisis de datos disponibles.
- Analizar los datos de accidentalidad vial en el tramo seleccionado según los parámetros del método de Bayes.

2. Antecedentes de metodologías en la identificación de tramos de alta accidentalidad

En el contexto de la seguridad vial, los puntos dentro de un tramo donde existe una alta concentración de accidentes adoptan el término "punto negro" (Brijs et al., 2007). La presencia de estos puntos críticos suele estar determinada por una combinación de factores, entre los cuales pueden estar un diseño deficiente de la vía, alta densidad de tráfico, falta de señalización adecuada y condiciones climáticas adversas.

La identificación de puntos negros en la red vial ha sido abordada a través de diferentes enfoques y metodologías. Algunas investigaciones han abordado lo más relevante frente al tema, destacando métodos tales como: Jerarquización analítica (Jakimavičius, 2018), agrupación de datos (Sobral et al., 2019) y estadísticos (Brijs et al., 2007).

Estadísticamente hablando, los accidentes son totalmente aleatorios, hasta el punto que es imposible predecir las circunstancias exactas de un accidente, sin embargo, acercamientos científicos consideran suposiciones de la existencia de una tasa media de accidentes para cada intersección vial (Brijs et al., 2007). Desde esta perspectiva estudios han analizado las tendencias en cuanto a accidentes y han provisto argumentos que declaran que el conteo de accidentes sigue a la ley de probabilidad de Poisson (Lord et al., 2005). Algunos de los enfoques probabilísticos estudiados incluyen el modelo de regresión de Poisson, que es útil para analizar condiciones del tráfico e incertidumbre en los datos pero que presenta problemas si hay mucha dispersión de los datos (Miaou & Lum, 1993); los modelos logarítmicos, que transforman la relación entre las

variables explicativas y la frecuencia de accidentes, facilitan la interpretación de los resultados (Andreassen & Hoque, 1986); el método de estimación de densidad por Kernel, que aunque usa una fórmula estadística, es más un método de agrupación de datos y una muy buena forma para afinar datos y visualizar zonas de interés (Baranyai & Sipos, 2022), (Abdullah & Sipos, 2024), entre otros enfoques estadísticos.

En esta investigación se realizó una búsqueda bibliográfica haciendo uso de la base de datos Scopus. En este proceso se usaron palabras claves keywords de búsqueda para encontrar la mayor cantidad de referencias considerables al tema. No obstante, se evitó utilizar términos que limitaran la búsqueda como por ejemplo “Black Spots” o “Puntos Negros” que, si bien es un término utilizado, el término “zonas con alta accidentalidad” sería el más óptimo. Los resultados de esta búsqueda fueron analizados mediante la herramienta Bibliometrix (Aria & Cuccurullo, 2017) presente en RStudio. A continuación, se muestran algunas de las gráficas obtenidas mediante este análisis que muestran las palabras, temas y países más usadas en los títulos y resúmenes de los artículos obtenidos en la búsqueda.

Figura 2

Nube de palabras más usadas en los artículos de la búsqueda bibliográfica

Figura 4

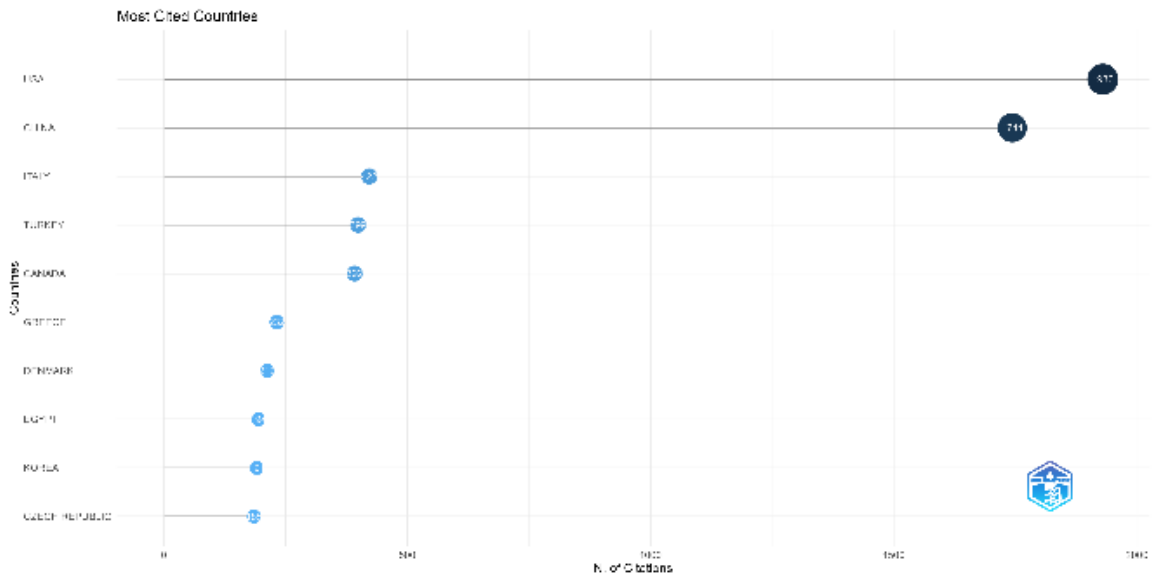
Acercamiento a figura 3 gráfica temas en tendencia a lo largo de los años



Nota. El gráfico muestra un acercamiento a la zona de interés del gráfico de tendencia de los temas de los artículos de la figura 3. Tomado de Bibliometrix (Aria & Cuccurullo, 2017).

Figura 5

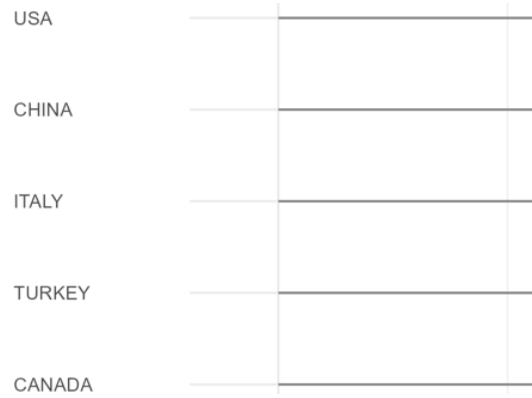
Países más citados en los artículos de la búsqueda bibliográfica



Nota. El gráfico representa los países más citados en los artículos obtenidos en la búsqueda bibliográfica. Tomado de Bibliometrix (Aria & Cuccurullo, 2017).

Figura 6

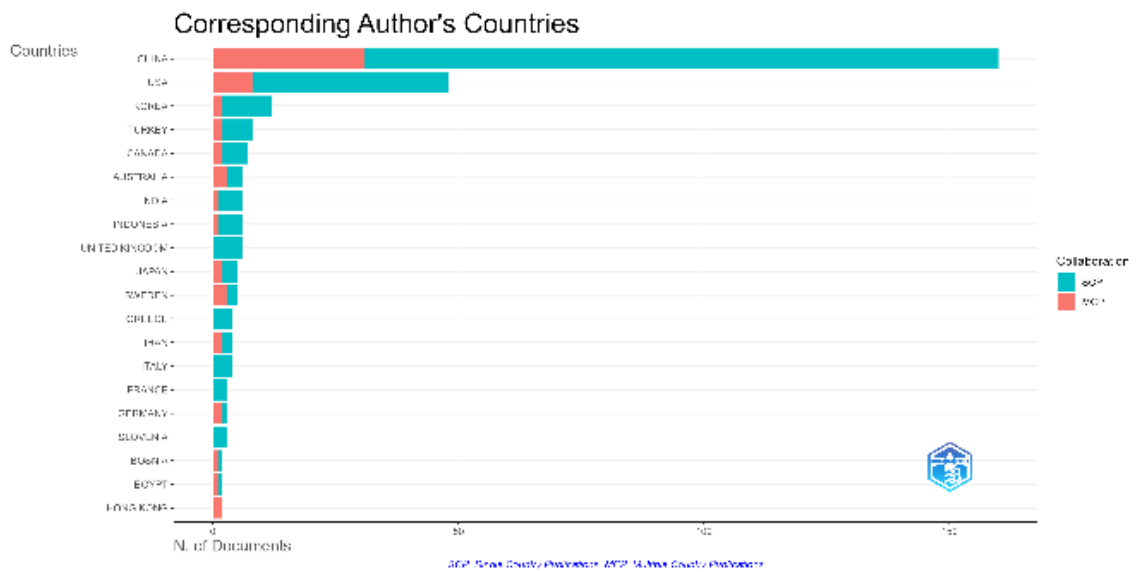
Acercamiento a los primeros 5 países más citados en la gráfica países más citados en los artículos de la búsqueda bibliográfica



Nota. El grafico muestra un acercamiento a los primeros resultados del grafico de los países más citados en los artículos de la búsqueda bibliográfica de la figura 5. Tomado de Bibliometrix (Aria & Cuccurullo, 2017).

Figura 7

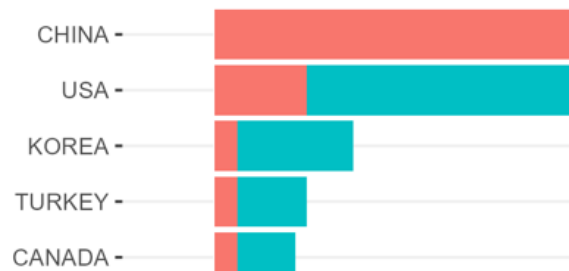
Países de los correspondientes autores en los artículos de la búsqueda bibliográfica



Nota. El gráfico representa los países de los autores en los artículos obtenidos en la búsqueda bibliográfica. Tomado de Bibliometrix (Aria & Cuccurullo, 2017).

Figura 8

Acercamiento a los primeros 5 países de los correspondientes autores en los artículos de la búsqueda bibliográfica



Nota. El gráfico muestra un acercamiento a los primeros resultados del gráfico de los países de los autores en los artículos de la búsqueda bibliográfica de la figura 7. Tomado de Bibliometrix (Aria & Cuccurullo, 2017).

La figura 2 muestra que el término Bayes theorem o traducido teorema de Bayes, se encuentra entre las palabras más usadas en los artículos encontrados en la búsqueda, de manera similar en las figuras 3 y 4, el teorema de Bayes aparece como uno de los temas en tendencia entre 2013 y 2021 mostrando así más relevancia que otros métodos. Las figuras 5, 6, 7, y 8 muestran que China y Estados Unidos ocupan los primeros puestos entre los países con más autores y citas.

Los métodos Bayesianos han sido ampliamente usados en estudios médicos para desarrollar algoritmos clínicos para estudiar diagnósticos de dengue (Caicedo et al., 2019), ideación suicida (Baños-Chaparro, 2021), casos de coronavirus (Giraldo-Ospina & Giraldo-López, 2022), mortalidades por cáncer (Torres-Avilés et al., 2015), entre otros.

Al analizar algunos artículos que involucraban aplicación del método de Bayes en el contexto vial se encontró que el interés en el uso de estos métodos en la identificación de zonas de alta accidentalidad se originó alrededor de los años ochenta (Brijs et al., 2006), de igual forma se encontró que algunas de las ventajas que conlleva el uso de modelos basados en enfoques Bayesianos jerárquicos es que permiten una clasificación de los resultados, esto posibilita examinar el riesgo de los lugares de forma comparativa y seleccionar los lugares de mayor riesgo (Brijs et al., 2007), Además, una de las razones de algunos autores para adoptar estimaciones bayesianas puede atribuirse a la baja cantidad de detalles presentes en informes de accidentes y que el modelo bayesiano permite la obtención de una clasificación de los lugares estudiados (Brijs et al., 2006).

Por otro lado, una de las posibles deficiencias del método radica en la suposición inicial de que los accidentes siguen una distribución de Poisson. Esto puede resultar limitante al describir casos extremos en los que un accidente produce múltiples víctimas, ya que esta distribución no los representa adecuadamente (Brijs et al., 2007) esto debido a que la distribución de Poisson está diseñada para manejar eventos con una dispersión similar a la media y no asigna suficiente probabilidad a valores extremos, además, que se demostró que el método Bayesiano empírico tiene deficiencias al ser aplicado en carreteras de alta velocidad (Ghadi & Török, 2017) debido a la dificultad que supone parametrizar variables como el factor humano, la cual es una de las más presentes en este tipo de vías.

El análisis de puntos de alta accidentalidad mediante el método de Bayes se basa en el uso de la regla general de Bayes la cual sirve para calcular la probabilidad condicional de un evento

dado otro (Walpole et al., 2012). Combinar este planteamiento con otros conceptos estadísticos o computacionales producen diversas variaciones y enfoques de los modelos Bayesianos, por ejemplo, al agregar un análisis estadístico previo que estime la distribución que siguen los datos genera los conocidos como modelos Bayesianos empíricos, al combinar el modelo inicial con previa definición de dependencias entre parámetros en niveles jerárquicos forma a los denominados modelos Bayesianos jerárquicos, al añadir métodos computacionales como Metropolis-Hastings o muestreo de Gibbs da origen a los modelos Bayesianos vía cadenas de Markov Monte Carlo, entre otros. Diversos estudios han abarcado estos diferentes enfoques y variaciones de los modelos bayesianos, a continuación, se muestran algunos ejemplos:

En el artículo “Application of Poisson random effect models for highway network screening” se muestra un estudio en el que se usó modelos de efecto aleatorio Poisson Log-Normal para la identificación de puntos críticos de riesgo de accidentes en vías urbanas de 4 carriles en el área central de florida entre los años 2006 y 2009 (Jiang et al., 2014).

Bayes Empírico, junto con otros métodos fueron comparados en el estudio “A comparative analysis of hotspot identification methods” en el cual se analizaron los datos de los accidentes presentados en la autopista italiana A16 durante un periodo de 5 años. En este estudio se sugiere que el método de Bayes Empírico debería usarse como el estándar en los posteriores estudios de identificación de puntos críticos (Montella, 2010).

Carl Belanger en su estudio “Estimation of safety of four-legged unsignalized intersections” utilizó el método de Bayes Empírico para la identificación de zonas de alta

accidentalidad en intersecciones de cuatro caminos no señalizadas en Quebec, Canadá (Belanger, 1994).

Maen Ghadi y Árpád Török en su artículo “Comparison Different Black Spot Identification Methods” publicado en 2017 hicieron una comparación entre métodos de detección, de agrupación y de predicción en vías con diferentes velocidades tomando al método de bayes empírico como representación de los métodos de predicción (Ghadi & Török, 2017).

En el estudio “A Bayesian model for ranking hazardous road sites” los autores estudian la implementación del modelo bayesiano mediante cadenas de Markov Monte Carlo y funciones de costo para analizar datos de accidentes de 519 intersecciones en Lovaina Bélgica (Brijs et al., 2007); posteriormente, los mismos autores en el artículo “A Model for Identifying and Ranking Dangerous Accident Locations: A Case-Study in Flanders” Continúan con la implementación de este método para analizar los datos de 23184 accidentes en la Región Flamenca de Bélgica (Brijs et al., 2006).

Así mismo, enfoques como el de Bayes jerárquico han sido abordados en artículos como “Bayesian ranking of sites for engineering safety improvements: Decision parameter, treatability concept, statistical criterion, and spatial dependence” en donde se analizan la elección de los parámetros de decisión y su consideración en la aplicación de este tipo de modelos (Miaou & Song, 2005).

Por otra parte, en el artículo “Ranking and selecting motor vehicle accident sites by using a hierarchical Bayesian model” se usó el modelo de Bayes jerárquico para analizar datos de accidentalidad en 35 sitios en Auckland Nueva Zelanda (Schlüter et al., 1997).

De igual modo, en el estudio “Bayesian identification of high-risk intersections for older drivers via gibbs sampling” se aplicó un Modelo Bayesiano Jerárquico con modelos de exposición para analizar altos riesgos para personas de edad avanzada en 102 intersecciones (Davis & Yang, 2001).

Ying MagNab, en su estudio “A Bayesian hierarchical model for accident and injury surveillance” implementa modelos Poisson Bayesianos Jerárquicos para el análisis de datos de hospitalizaciones debido a lesiones en jóvenes de 0 a 24 años en la provincia de Columbia Británica en Canadá en el periodo comprendido desde 1987 a 1996, (MacNab, 2003).

En coherencia con lo anterior, la Unión Europea indica que los procedimientos incluidos en el Sistema de Gestión de Seguridad de Infraestructura Vial (RISMS) deben implementarse en las vías primarias de la red nacional, el primer paso en el RISMS es denominado Evaluación de la red vial (RNS) el cual define metodologías para la identificación de puntos negros entre las cuales se encuentran modelos bayesianos como Bayes empírico, exceso empírico bayesiano, modelos bayesianos Poisson log-normal, entre otros (Bonera et al., 2022).

Si bien se han realizado numerosos estudios sobre los diversos métodos de identificación de puntos críticos de accidentes, los artículos seleccionados en esta sección fueron considerados los más significativos.

3. Datos

En esta sección se describe la información utilizada para el desarrollo del estudio. Se detalla el proceso de obtención de los datos, así como la identificación del tramo de estudio seleccionado para el análisis.

3.1 Obtención de datos

Para la obtención de los datos se hizo una búsqueda en las páginas web de las organizaciones de control disponibles como las direcciones de tránsito y transporte de los municipios del área metropolitana y de la ANSV. Los datos usados fueron tomados desde la página oficial de la ANSV desde un plano interactivo donde mostraban todos los decesos ocurridos a nivel nacional debido a accidentes de tránsito desde el año 2012 hasta el 2024 (Agencia Nacional de Seguridad Vial, 2024b) mediante la función query (búsqueda) de los datos de capa con información a través del servidor de entidades (Agencia Nacional de Seguridad Vial, 2024c). Debido a los parámetros de la búsqueda se obtuvieron los datos en formato JSON y en varios bloques de datos debido a que el tamaño máximo de la consulta permitida corresponde a 2000 datos, además los datos de interés para esta investigación se encontraban anidados en una estructura features correspondiente a una sola columna en los datos originales. Dados dichos inconvenientes se procesaron los datos mediante la herramienta de uso libre Google Colab, el código usado para este proceso se muestra en su totalidad en el apéndice A mostrado al final del presente documento.

Abordando el procesamiento de los datos realizado de forma más detallada este código tomó el archivo original con el resultado de todas las búsquedas realizadas y lo procesó por bloques, se procesaron los atributos anidados, se filtraron en base al municipio de ocurrencia delimitándolos al área metropolitana de Bucaramanga, se redondearon las coordenadas de ocurrencia de estos accidentes a tres decimales para que los accidentes ocurridos en un radio de 100 metros compartieran las coordenadas para ser tratados como un punto de alta accidentalidad y finalmente se descargó el archivo ya procesado.

Los datos obtenidos contenían la información correspondiente a las siguientes variables:

OBJECTID: Esta variable muestra un número de identificación único para cada deceso.

Años: Esta variable marca el año en el cual ocurrió el deceso, los datos encontrados muestran información desde el año 2012 hasta el 2024

.

UsuarioVia: Esta variable identifica el rol del usuario fallecido en el accidente. Las posibles categorías son: Peatón, Usuario de bicicleta, Usuario de moto, Usuario de otros, Usuario de vehículo individual, Usuario de transporte de carga y Usuario de transporte de pasajeros y en los casos que no se tenía la información se marca como Sin información.

FechaHecho: Esta categoría muestra la fecha en la cual ocurrió el incidente.

HoraHecho: Esta categoría muestra la hora en la cual ocurrió el incidente; en el caso que el registro estuviera incompleto se marca como NULL o se muestra la casilla en blanco.

Sexo: Masculino o Femenino dependiendo de la persona involucrada en el accidente.

Edad: Muestra la edad de la persona involucrada en el accidente.

Municipio_Ubicacion: Esta categoría muestra el municipio en el cual ocurrió el accidente.

Departamento_Ubicacion: Esta categoría muestra el departamento en el cual ocurrió el accidente.

TipoVehiculoHecho: Describe el vehículo implicado en el accidente. Las posibles opciones incluyen: Automóvil, Bicicleta, Bus, Camión, Camioneta - Campero, Motocicleta, Mototaxi, Tracción humana, Tracto – camión, en los casos en que el accidente se trataba de atropellos a peatones se marca como no aplica y en los casos que no se tiene información se acota como tal: Sin información.

Circunstancia: Define la causa o situación por la que ocurrió el accidente. Entre las posibilidades están: Exceso de velocidad, No respetar prelación, Contravía, Vehículo mal estacionado, Falta de precaución por niebla, Falta de señalización, entre otras.

ClaseAccidente: Clasifica el tipo de siniestro. Las opciones son: Atropello, Choque con objeto fijo o en movimiento, Volcamiento, Caída del ocupante, Choque con otro vehículo, y en los casos que no se cuenta con la información se denota: Sin información.

Adm_Region: Esta categoría clasifica la región en que ocurrió el accidente dependiendo de si era una zona rural o Urbana.

Y: Esta categoría muestra la coordenada correspondiente a la latitud del sitio exacto donde ocurrió el incidente.

X: Esta categoría muestra la coordenada correspondiente a la longitud del sitio exacto donde ocurrió el incidente.

Para la implementación del modelo, se seleccionaron algunas variables de entre las disponibles basándose en su relevancia para describir las características de los accidentes fatales. Las variables elegidas entre las disponibles fueron: UsuarioVia, TipoVehiculoHecho y Circunstancia.

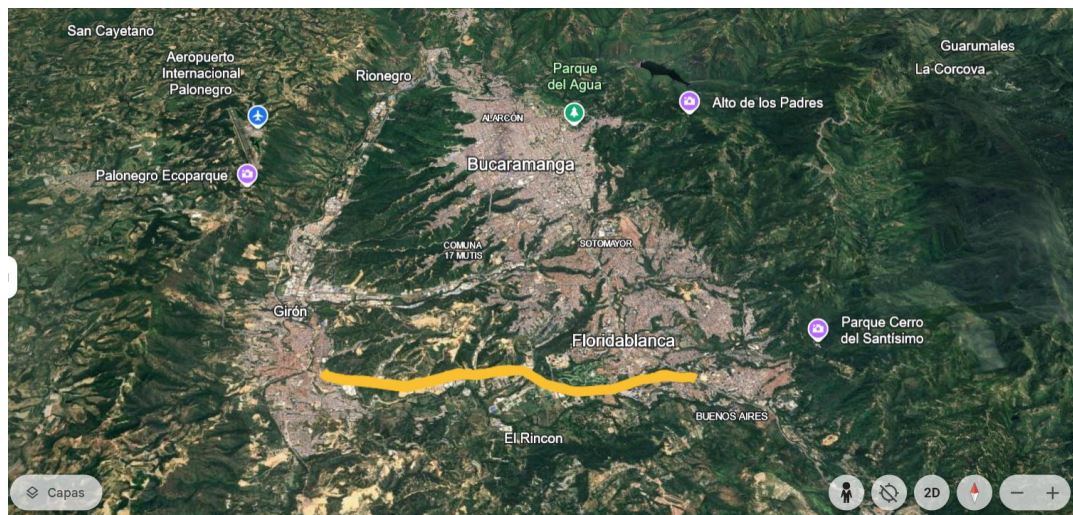
3.2 Identificación tramo de estudio

Para la elección del tramo de estudio se revisaron varios factores de las principales vías del área metropolitana de Bucaramanga como la disponibilidad de datos disponible, relevancia, uso, flujo vehicular, entre otros. Con base en estos criterios, se seleccionó un tramo de aproximadamente 7.7 kilómetros de la vía conocida como el Anillo Vial, contando como referencia inicial la estación de servicio Caneyes y como referencia final su intersección con la Avenida

Floridablanca a la altura de la sede Floridablanca de la Universidad Santo Tomás. Este tramo pertenece a la ruta 45A de la red vial nacional, la cual sirve de conexión entre la capital del país y el Caribe Colombiano pasando por la zona de los Santanderes. Este tramo cuenta con una particular combinación de factores de riesgo y numerosos registros de fallecimientos. Entre los factores que pueden aumentar el riesgo de accidentes se identifican: la coexistencia de tránsito pesado y liviano, un alto flujo vehicular asociado a las zonas industriales, un elevado tránsito peatonal en áreas residenciales y variaciones en velocidades y volúmenes de tránsito a lo largo del día. Este es un tramo clave que forma parte de la principal y más directa arteria de conexión del área metropolitana al conectar los municipios de Girón, Piedecuesta y Floridablanca. A continuación, la figura 9 y 10 se presenta el tramo de estudio seleccionado:

Figura 9

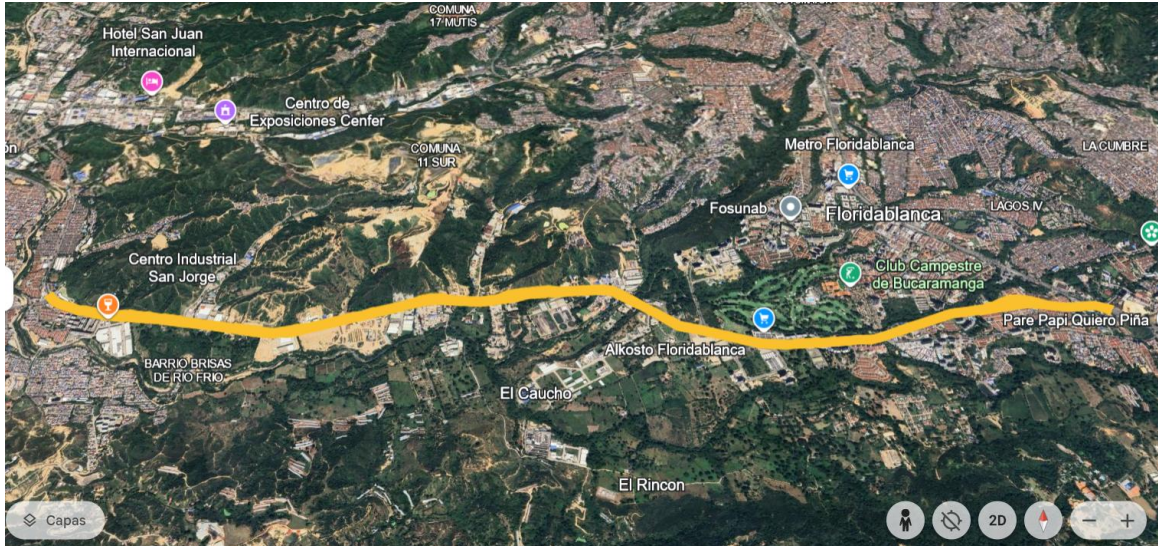
Ubicación tramo de estudio con respecto al área metropolitana de Bucaramanga



Nota. El grafico muestra la ubicación del tramo de estudio con respecto a los municipios de Bucaramanga, Girón y Floridablanca. Tomado del sitio oficial de Google Earth (Google Earth, 2024).

Figura 10

Acercamiento a ubicación tramo de estudio con respecto al área metropolitana de Bucaramanga.



Nota. El grafico muestra un acercamiento con más detalles a la figura 9 la cual muestra la ubicación del tramo de estudio con respecto a los municipios de Bucaramanga, Girón y Floridablanca. Tomado del sitio oficial de Google Earth (Google Earth, 2024).

4. Implementación del modelo

En esta sección se describe el proceso de implementación del modelo Bayesiano utilizado para el análisis de puntos con alta accidentalidad. Se presentan las bases teóricas y técnicas del modelo, comenzando con la definición de la distribución de Poisson, seguida por la aplicación del teorema de Bayes y el método de cadenas de Markov Monte Carlo (MCMC). Finalmente, se expone la construcción del modelo en su conjunto.

4.1 Distribución de Poisson

La distribución de Poisson es una distribución de probabilidad discreta que modela el número de eventos o resultados que ocurren en un intervalo de tiempo o en una región específica. Estos modelos, conocidos como experimentos de Poisson, se caracterizan por generar valores de una variable aleatoria X , que representa el número de eventos observados. Dichos intervalos pueden abarcar diferentes dimensiones, como tiempo, (por ejemplo, el número de llamadas telefónicas por hora), área (el número de ratas por acre), o volumen (el número de bacterias en un cultivo).

Un experimento de Poisson se fundamenta en el proceso de Poisson, que cumple con las siguientes propiedades:

- Independencia de los intervalos: El número de eventos que ocurren en un intervalo o región específica es independiente de los eventos que ocurren en otros intervalos o regiones disjuntas.
- Proporcionalidad en intervalos pequeños: La probabilidad de que ocurra un solo evento en un intervalo de tiempo muy corto o en una región pequeña es directamente proporcional a la longitud del intervalo o al tamaño de la región.
- Improbabilidad de eventos múltiples en intervalos pequeños: La probabilidad de que ocurran múltiples eventos en un intervalo corto o región pequeña es insignificante.

La distribución de Poisson está definida por la función de probabilidad mostrada a continuación en la ecuación 1:

$$p(x; \lambda t) = \frac{(\lambda t)^x e^{-\lambda t}}{x!}, x = 0, 1, 2, \dots \quad (1)$$

Donde:

- x es el número de eventos observados.
- λ es el número promedio de eventos por unidad de tiempo, área, distancia o volumen, calculado como $\lambda = Ar$, siendo A la tasa de ocurrencia y r el intervalo o región de interés.
- e es la base del logaritmo natural ($e \approx 2.71828$).
- $x!$ es el factorial de x .

Esta distribución es utilizada para modelar fenómenos en los que los eventos ocurren de forma independiente y aleatoria dentro del intervalo o región observada, (Walpole et al., 2012).

4.2 Teorema de Bayes

El teorema de Bayes es una herramienta fundamental en la teoría de probabilidad que permite calcular la probabilidad condicional de un evento dado otro, especialmente cuando el espacio muestral está particionado en varios subconjuntos. Se utiliza para actualizar las probabilidades iniciales (a priori) de eventos, considerando nueva información que se obtiene (a posteriori).

Si B_1, B_2, \dots, B_k constituyen una partición del espacio muestral S , y se cumple que $P(B_i) \neq 0$ para $i=1,2,\dots,k$ entonces, para cualquier evento A en S tal que $P(A) \neq 0$, la probabilidad condicional de B_i dado A está dada por la formula mostrada en la ecuación 2:

$$P(B_r | A) = \frac{P(B_r \cap A)}{\sum_{i=1}^k P(B_i \cap A)} = \frac{P(B_r)P(A | B_r)}{\sum_{i=1}^k P(B_i)P(A | B_i)} \text{ para } r = 1,2, \dots, k. \quad (2)$$

Donde:

- $P(B_i \cap A)$ representa la probabilidad conjunta de que ocurra el evento B_i y el evento A .
- $P(A|B_i)$ es la probabilidad de que ocurra A dado que ocurrió B_i
- $\sum P(B_j)P(A|B_j)$ corresponde a la suma de las probabilidades conjuntas de A con cada uno de los eventos B_j , garantizando que se evalúe sobre toda la partición del espacio muestral.

Este teorema permite abordar problemas como el cálculo de la probabilidad de que un evento específico haya sido causado por un subconjunto particular dentro de una partición, dado que se ha observado cierta información. Por ejemplo, si se selecciona un producto defectuoso de manera aleatoria, el teorema de Bayes puede emplearse para determinar la probabilidad de que dicho producto haya sido ensamblado en una máquina específica.

El cálculo incluye dos pasos fundamentales:

Determinar las probabilidades conjuntas $P(B_i \cap A) = P(B_i)P(A|B_i)$ para cada B_i de la partición.

Dividir la probabilidad conjunta del evento de interés $B_i \cap A$ entre la probabilidad total de A , representada por la suma de todas las probabilidades conjuntas.

El teorema de Bayes es ampliamente utilizado en diversas aplicaciones estadísticas y en la toma de decisiones, donde es necesario incorporar información adicional para ajustar las probabilidades iniciales, (Walpole et al., 2012).

4.3 Cadenas de Markov Monte Carlo (MCCM)

El algoritmo de Monte Carlo para Cadenas de Markov (MCCM) es una técnica de inferencia probabilística ampliamente utilizada en redes bayesianas. Este método genera muestras a partir de distribuciones posteriores complejas mediante un proceso iterativo que explora el espacio de estados. A diferencia de otros algoritmos de muestreo que generan cada muestra desde cero, MCCM basa cada nueva muestra en la anterior, modificando aleatoriamente el valor de una variable y condicionando esta elección a las variables circundantes, conocidas como el manto de Markov.

En una red bayesiana, el estado actual se define asignando valores a todas las variables. A cada paso del algoritmo, se selecciona aleatoriamente una variable no observable y se actualiza su valor mediante un muestreo condicionado a las variables en su manto de Markov. Este manto incluye sus padres, hijos y los padres de estos últimos. El algoritmo recorre así el espacio de

estados, produciendo una secuencia de muestras que tiende hacia la distribución a posteriori deseada.

El núcleo del MCCM es el muestreador de Gibbs, una variante que optimiza el proceso de muestreo al calcular la probabilidad de cada variable condicionada únicamente a su manto de Markov. Esta propiedad reduce significativamente la complejidad computacional del método, permitiendo su aplicación en problemas de gran escala.

La validez del MCCM se basa en el concepto de distribución estacionaria, que asegura que el porcentaje de tiempo que el algoritmo pasa en cada estado es proporcional a la probabilidad a posteriori de ese estado. Esto se logra mediante la propiedad de balance detallado, que establece que el flujo esperado entre cualquier par de estados es igual en ambas direcciones. Bajo supuestos estándar, existe una única distribución estacionaria que satisface estas condiciones, garantizando la convergencia del método (Russell & Norvig, 2004).

El uso del MCCM en el análisis de seguridad vial permite modelar situaciones complejas donde múltiples variables interactúan de manera probabilística. Por ejemplo, al analizar la ocurrencia de accidentes en puntos críticos de una red vial, como es el caso de este informe, el MCCM facilita la estimación de probabilidades condicionadas que serían intratables mediante métodos deterministas.

4.4 Construcción del modelo

En este trabajo se desarrolló un modelo Bayesiano basado en la distribución de Poisson haciendo uso del software RStudio, una herramienta gratuita de código abierto ampliamente usada

en análisis estadístico. El modelo fue desarrollado con un enfoque MCCM, utilizando la librería *brms*, con el objetivo de analizar la distribución de accidentes en el tramo previamente escogido. A continuación, se describe la programación y la lógica implementada.

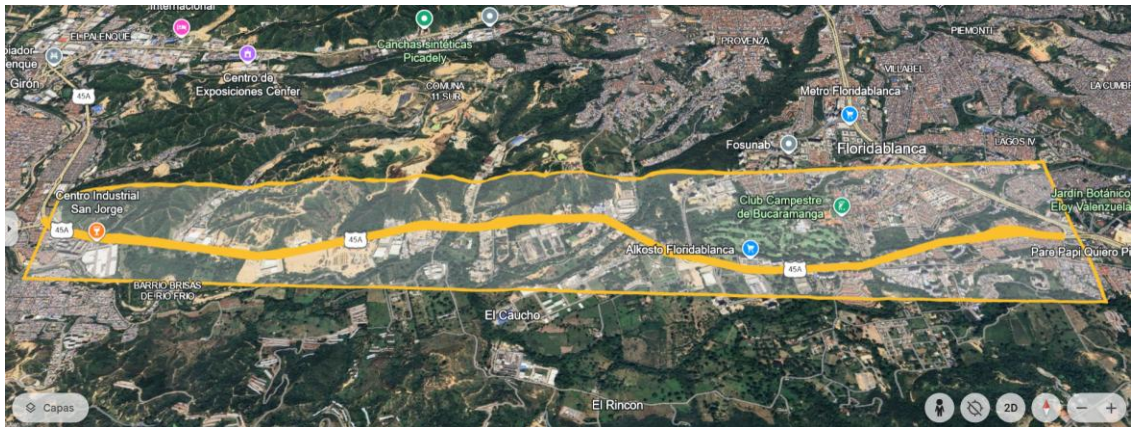
Primero, se cargaron las librerías necesarias para el análisis espacial, la manipulación de datos y la modelación bayesiana.

Luego se configuró *brms* para aprovechar múltiples núcleos de CPU, optimizando la ejecución del modelo.

Posteriormente se programó el modelo para tomar los datos de accidentes directamente desde el archivo CSV anteriormente procesado y nuevamente filtrados para seleccionar únicamente los incidentes dentro de un polígono imaginario que representa el tramo específico del Anillo Vial estudiado. Este polígono se definió mediante las siguientes coordenadas geográficas: Punto A (-73.1634449635897, 7.067735254705352), Punto B (-73.09608144715092, 7.070498654379905), Punto C (-73.09628000066337, 7.058281789801652), Punto D (-73.16295803965895, 7.060462226024185). En la siguiente imagen se muestra una visión satelital del polígono imaginario en cuestión:

Figura 11

Vista polígono imaginario y ubicación tramo de estudio



Nota. El gráfico muestra la vista del polígono imaginario usado en la programación del modelo con respecto al tramo de estudio. Tomado del sitio oficial de Google Earth (Google Earth, 2024).

Una vez establecidas las coordenadas se empleó la librería `sf` para operaciones espaciales.

Luego el conjunto de datos fue convertido a un `data.frame`, y se crearon columnas para las coordenadas redondeadas, agrupando los datos para obtener la frecuencia de accidentes por ubicación.

Luego el modelo Bayesiano se definió con las siguientes líneas presentes en el código:

```

modelo_anillo <- brm(
  num_accidentes ~ UsuarioVia + TipoVehiculoHecho + Circunstancia + (1 |
lat_round/lon_round),
  data = summary_data,
  family = poisson(),
  chains = 4,
  iter = 6000,

```

```
warmup = 3000,  
control = list(  
  adapt_delta = 0.9,  
  max_treedepth = 15)
```

Tras ajustar los parámetros del modelo, se generaron predicciones y se evaluó el desempeño del modelo mediante métricas como el error medio absoluto y el error porcentual medio.

Enseguida se creó un gráfico de dispersión para poder visualizar rápidamente las predicciones en función de las coordenadas y finalmente los resultados de predicción y validación se exportaron en archivos CSV para análisis posterior. El código del modelo es mostrado en el apéndice B mostrado en la sección de apéndices del presente documento.

Es oportuno mencionar que en este código el número de accidentes fue modelado en función de variables categóricas como el tipo de usuario, el tipo de vehículo y las circunstancias del accidente.

Asimismo, inicialmente se programó el modelo con un total de 4 cadenas independientes de Markov, cada cadena con 6000 iteraciones siguiendo el criterio utilizado en el estudio “A Bayesian model for ranking hazardous road sites” [4], es decir, en total la muestra del modelo cuenta con 24.000 iteraciones. Un aspecto relevante en el proceso de construcción del modelo es que según el criterio utilizado en el artículo anteriormente mencionado se debían usar las primeras

1.000 iteraciones de cada cadena para el periodo de adaptación o warm up, en el cual el modelo se estabiliza, pero dado que el presente modelo presenta una mayor complejidad gracias al uso de variables categóricas se estableció que de cada cadena se usaran las primeras 3.000 iteraciones. Se ajustó la tasa de aceptación y la profundidad máxima de los árboles en el muestreo del algoritmo Hamiltoniano Monte Carlo (HMC).

5. Resultados del análisis

Esta sección presenta los resultados obtenidos tras la implementación del modelo. Se detalla la validación del modelo propuesto, evaluando su precisión y consistencia en la identificación de puntos críticos. Además, se exponen los tramos con mayor probabilidad de ocurrencia de accidentes, lo que constituye un insumo clave para la toma de decisiones en seguridad vial.

5.1 Validación del modelo

Inicialmente, el modelo presentó tiempos de evaluación de gradientes bajos, lo que indica una complejidad computacional moderada. Sin embargo, ajustes adicionales fueron necesarios para mejorar la eficiencia.

En las primeras ejecuciones del modelo se observó que múltiples transiciones excedieron la profundidad máxima del árbol en el muestreo del algoritmo HMC durante el muestreo, sugiriendo una geometría compleja del espacio posterior por lo cual se incrementó este valor para mejorar la exploración del espacio paramétrico.

Además, se observó durante las primeras ejecuciones del modelo fue que el diagnóstico de convergencia mediante el factor R -hat mostró valores superiores a 2,81, lo que indica que las cadenas no convergieron adecuadamente, por lo cual se ajustó el número de iteraciones para mejorar la mezcla de cadenas.

De manera similar, en las primeras ejecuciones del modelo se notó que el tamaño efectivo de muestra (ESS) fue bajo, lo que indicaba una alta correlación entre las muestras generadas. Esto podría afectar la precisión de las estimaciones, por lo que se incrementó el número de iteraciones y se evaluó una reparametrización del modelo.

Luego de varios ajustes se estableció que el modelo tuviera 10.000 iteraciones por cada una de las 4 cadenas de Markov, que la tasa de aceptación y la profundidad máxima de los árboles del algoritmo HMC fueran de 0.95 y 20 respectivamente

Con estos ajustes finales se logró que la convergencia del modelo mediante la inspección del factor de convergencia potencial (R -hat), mostrara valores cercanos a 1 en todos los parámetros lo cual muestra que todos los parámetros convergieron correctamente. Asimismo, el tamaño de muestra efectivo (ESS) fue adecuado, garantizando la fiabilidad de las estimaciones.

Por otro lado, la validación del modelo arrojó un error medio absoluto (MAE) de 0.398, indicando una desviación promedio baja entre las predicciones y los valores reales. El error porcentual medio fue del 27.1%, con una desviación estándar del error de 0.593, lo que confirma la estabilidad y precisión del modelo en la predicción de accidentes

5.2 Puntos con alta probabilidad de accidentes

Los puntos de alta accidentalidad encontrados en el tramo estudiado se muestran en la figura 12, 13 y la tabla 1:

Tabla 1

Puntos de alta accidentalidad en el tramo estudiado

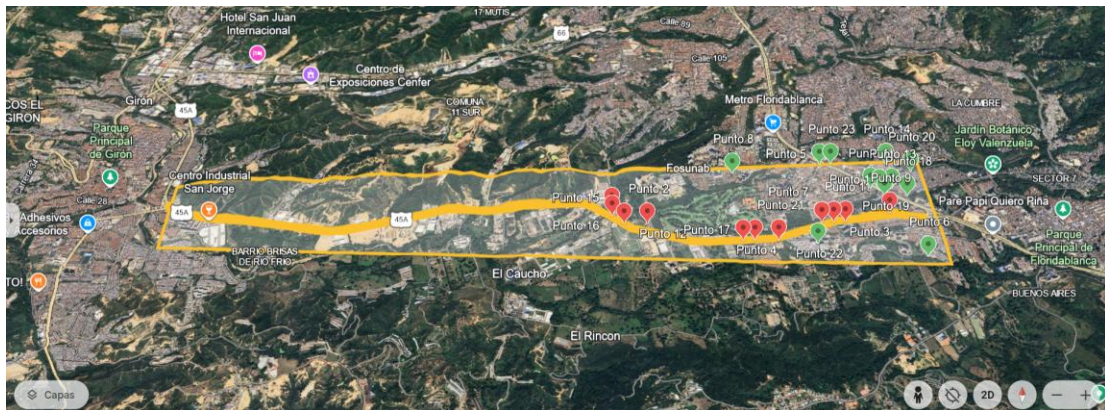
Latitud	Longitud	# accidentes	Predicción	Usuario Vía	Tipo Vehículo	Circunstancia
7.067	-73.1	5	3.198466427	Usuario de moto	Motocicleta	Sin Información
7.063	-73.121	3	2.78184712	Usuario de moto		Sin Información
7.063	-73.105	3	2.32347211	Usuario de bicicleta		Exceso De Velocidad
7.063	-73.106	2	1.651986816	Peatón		Cruzar Sin Observar
7.069	-73.113	2	1.589805487	Peatón		Cruzar Sin Observar
7.064	-73.1	2	1.523144496	Peatón		Otra
7.063	-73.123	2	1.476415116	Usuario de moto	Motocicleta	Exceso De Velocidad
7.061	-73.112	1	2.028836573	Usuario de moto	Motocicleta	Sin Información
7.07	-73.105	1	2.006868383	Usuario de moto	Motocicleta	Sin Información
7.06	-73.097	1	1.680681828	Usuario de bicicleta		Exceso De Velocidad
7.067	-73.101	1	1.496016735	Peatón		Cruzar Sin Observar
7.063	-73.104	1	1.478335958	Peatón		Otra
7.067	-73.099	1	1.303913276	Usuario de moto	Motocicleta	Exceso De Velocidad
7.07	-73.099	1	1.244894113	Peatón		Cruzar Sin Observar
7.065	-73.124	1	1.140636711	Usuario de moto		Exceso De Velocidad
7.064	-73.124	1	1.087745733	Usuario de moto		Otra
7.061	-73.113	1	1.060541053	Usuario de moto		Exceso De Velocidad
7.066	-73.098	1	1.015526392	Usuario de moto		Otra
7.066	-73.1	1	1.013192099	Usuario de moto		Otra
7.069	-73.097	1	0.999792717	Sin Información		Otra

7.061	-73.11	1	0.98930627	Usuario de moto	Impericia En El Manejo
7.061	-73.106	1	0.940223621	Usuario de moto	Otra
7.07	-73.104	1	0.937712107	Usuario de moto	Otra

Nota. Esta tabla muestra los resultados obtenidos de la ejecución del modelo en Rstudio.

Figura 12

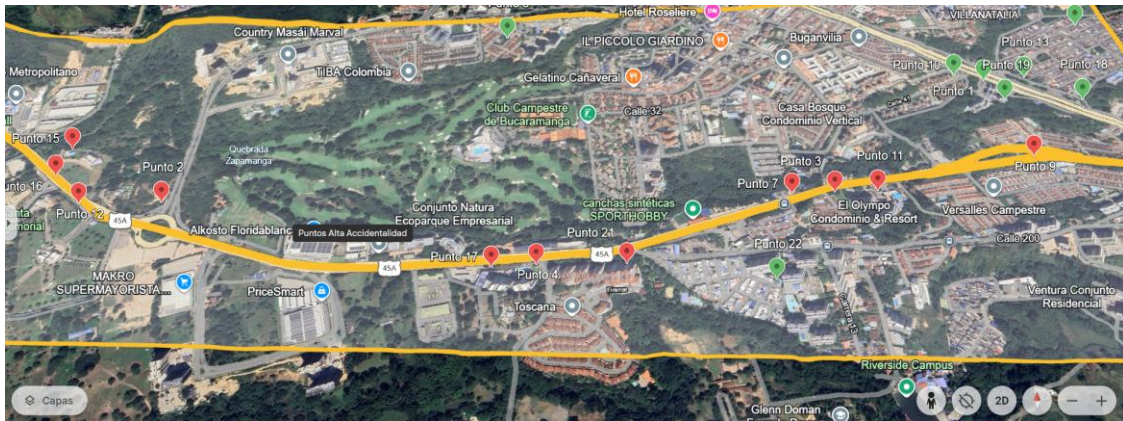
Imagen satelital con la delimitación del tramo de estudio y los puntos de alta accidentalidad debidamente georreferenciados



Nota. El grafico muestra la visualización de los puntos de alta accidentalidad obtenidos luego de la ejecución del modelo con relación al tramo de estudio. Tomado del sitio oficial de Google Earth (Google Earth, 2024).

Figura 13

Acercamiento a imagen satelital con la delimitación del tramo de estudio y los puntos de alta accidentalidad debidamente georreferenciados



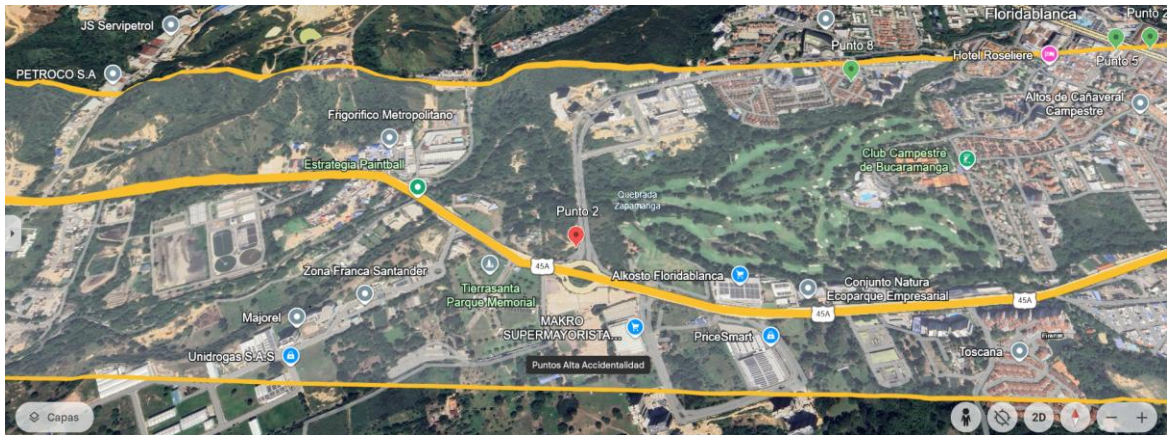
Nota. El gráfico muestra un acercamiento con más detalles a la figura 12 la cual muestra la visualización de los puntos de alta accidentalidad obtenidos luego de la ejecución del modelo con relación al tramo de estudio. Tomado del sitio oficial de Google Earth (Google Earth, 2024).

El modelo encontró un total de 23 puntos de alta accidentalidad, de los cuales 12 puntos no están en el tramo de interés, sino que pertenecen a vías cercanas. Esto se presentó debido al margen dado por la delimitación mediante las coordenadas del polígono imaginario anteriormente explicado. A continuación, se destacan los tres puntos más críticos:

El punto más crítico encontrado corresponde al punto 2 en los resultados de la modelación y se trata de la intersección de la transversal el bosque con la el tramo de la vía estudiado identificado con las coordenadas (7.067, -73.1), en este punto han ocurrido accidentes que han cobrado 3 víctimas fatales donde el factor vial predominantemente afectado es el motociclista. En las siguientes figuras se muestra un mapa con el punto en cuestión:

Figura 14

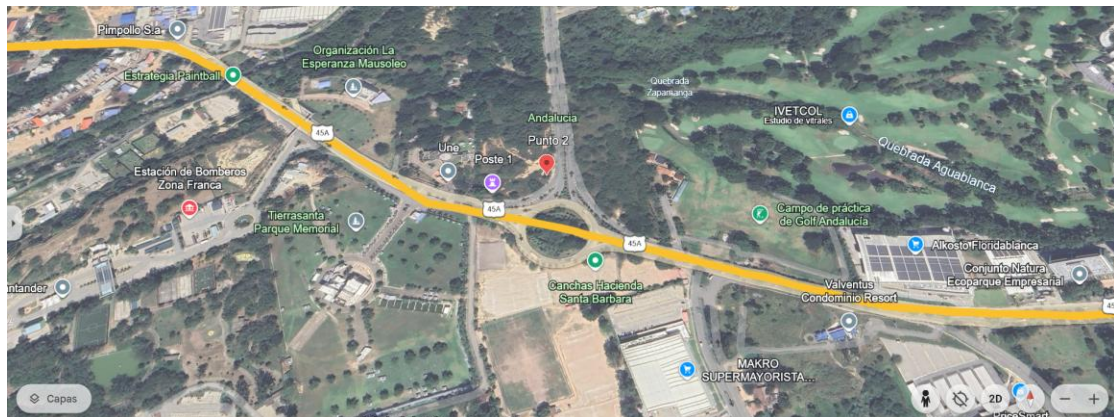
Imagen satelital del punto de alta accidentalidad más crítico en el tramo de estudio debidamente georreferenciados



Nota. El gráfico muestra la visualización del punto de alta accidentalidad más crítico obtenido en el análisis del modelo con relación al tramo de estudio. Tomado del sitio oficial de Google Earth (Google Earth, 2024).

Figura 15

Acercamiento a imagen satelital del punto de alta accidentalidad más crítico en el tramo de estudio debidamente georreferenciados

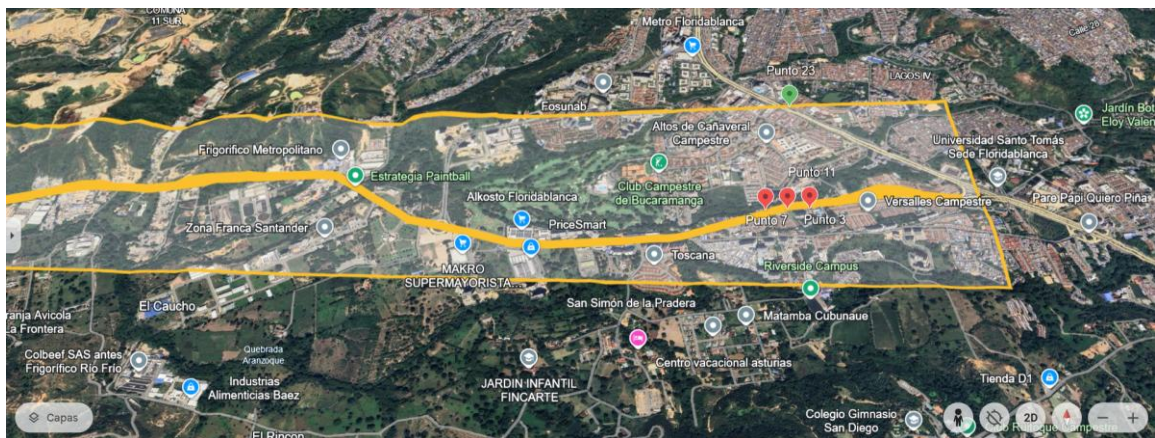


Nota. El gráfico muestra un acercamiento con más detalle a figura 14 que muestra la visualización del punto de alta accidentalidad más crítico obtenido en el análisis del modelo con relación al tramo de estudio. Tomado del sitio oficial de Google Earth (Google Earth, 2024).

En el caso del segundo punto más crítico encontrado corresponde al punto 3 en los resultados de la modelación y se encontró que está a menos de 120 metros de los puntos 7 y 11 por lo cual se asumirán estos tres como un solo punto. Estos puntos corresponden a las entradas a los conjuntos residenciales Quintas de cañaveral, Urbanización Villa Margarita y caminos de Providenza respectivamente, este punto de alta accidentalidad es identificado con las coordenadas (7.063, -73.105), en este punto han ocurrido accidentes que han cobrado 6 víctimas fatales dónde los factores viales predominantemente afectados son los biciusuarios, peatones y motociclista. Las circunstancias predominantes de estos accidentes son el exceso de velocidad y cruzar sin observar. En las siguientes figuras se muestra un mapa con el punto en cuestión:

Figura 16

Imagen satelital del segundo punto de alta accidentalidad más crítico en el tramo de estudio debidamente georreferenciados



Nota. El grafico muestra la visualización del segundo punto de alta accidentalidad más crítico obtenido en el análisis del modelo con relación al tramo de estudio. Tomado del sitio oficial de Google Earth (Google Earth, 2024).

Figura 17

Acercamiento a imagen satelital del segundo punto de alta accidentalidad más crítico en el tramo de estudio debidamente georreferenciados

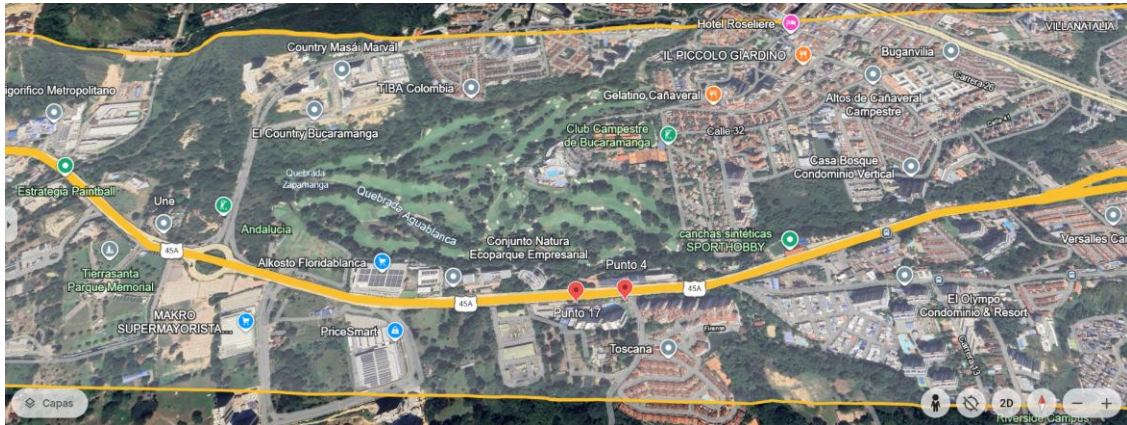


Nota. El gráfico muestra un acercamiento con más detalle a figura 16 que muestra la visualización del segundo punto de alta accidentalidad más crítico obtenido en el análisis del modelo con relación al tramo de estudio. Tomado del sitio oficial de Google Earth (Google Earth, 2024).

En el caso del tercer punto más crítico encontrado corresponde al punto 4 en los resultados de la modelación y se encontró que está a menos de 120 metros del punto 17 por lo cual se asumirán estos dos como un solo punto. Estos puntos corresponden a las entradas al conjunto residenciales Mediterrané y la tienda Altoque Anillo Vial respectivamente, este punto de alta accidentalidad es identificado con las coordenadas (7.061, -73.112), en este punto han ocurrido accidentes que han cobrado 3 víctimas fatales donde los factores viales predominantemente afectados son los peatones y motociclista. Las circunstancias predominantes de estos accidentes son el exceso de velocidad y cruzar sin observar. En las siguientes figuras se muestra un mapa con el punto en cuestión:

Figura 18

Imagen satelital del tercer punto de alta accidentalidad más crítico en el tramo de estudio debidamente georreferenciados



Nota. El grafico muestra la visualización del tercer punto de alta accidentalidad más crítico obtenido en el análisis del modelo con relación al tramo de estudio. Tomado del sitio oficial de Google Earth (Google Earth, 2024).

Figura 19

Acercamiento a imagen satelital del tercer punto de alta accidentalidad más crítico en el tramo de estudio debidamente georreferenciados



Nota. El grafico muestra un acercamiento con más detalle a figura 18 que muestra la visualización del tercer punto de alta accidentalidad más crítico obtenido en el análisis del modelo con relación al tramo de estudio. Tomado del sitio oficial de Google Earth (Google Earth, 2024).

6. Conclusiones

- La revisión bibliográfica confirmó que el método de Bayes es una herramienta adecuada para identificar puntos críticos de accidentalidad en tramos viales y evaluar escenarios con información limitada. El proceso llevado a cabo en este trabajo investigativo demostró ser viable en el contexto local, aunque con adaptaciones específicas para abordar la disponibilidad de datos en los organismos reguladores de la infraestructura vial del Área Metropolitana de Bucaramanga.
- Con base en la disponibilidad de información en las bases de datos presente en las agencias de control, las características del tráfico y de la zona se escogió un tramo del Anillo Vial el cual cuenta con una longitud aproximada de 7.7 kilómetros y se puede referenciar tomando como punto de inicio la estación de servicio Los Caneyes y como punto final la intersección de la misma con la Avenida Floridablanca a la altura de la sede Floridablanca de la Universidad Santo Tomás.
- El modelo arrojó un total de 23 puntos críticos de alta accidentalidad entre los cuales la mayoría correspondía a intersecciones y accesos a conjuntos residenciales, además el análisis estableció que los factores viales más vulnerables como peatones y motociclistas son los más afectados, además que el exceso de velocidad y las malas conductas peatonales son las principales causas de los accidentes en esta zona.

- El análisis realizado en el tramo vial del presente estudio puede servir como modelo para estudios similares en otros tramos de la ciudad e incluso en otras regiones del país, con algunos ajustes en las variables según los datos disponibles de cada área.
- La experiencia obtenida en este estudio revela que, si bien los datos disponibles permitieron realizar el análisis, una mejora en su calidad y sistematización sería determinante para incrementar la precisión y utilidad de este tipo de estudios.

7. Recomendaciones

Tomando en cuenta los resultados obtenidos de la implementación del modelo y un análisis de un registro audiovisual de autoría propia (Rodríguez, 2025), hechos en el tramo de estudio se proponen las siguientes recomendaciones para la mejora de la seguridad vial en dicha zona:

- Con base en los hallazgos del estudio, se recomienda implementar estrategias enfocadas en la protección de peatones y motociclistas, como el diseño de pasos seguros, campañas de concientización para estos grupos, y una mayor supervisión en puntos donde se presentan altos índices de accidentalidad.
- Es prioritario realizar labores de mantenimiento en el tramo analizado para reparar los huecos y grietas significativas, ya que representan un riesgo considerable para motociclistas y otros usuarios vulnerables. Asimismo, se recomienda nivelar los accesos a los conjuntos residenciales para garantizar una transición segura y fluida entre la vía principal y estas zonas.

- Los reductores de velocidad deteriorados deben ser reparados o sustituidos para garantizar su funcionalidad y contribuir al control de velocidad en las áreas críticas identificadas.
- Dado que el exceso de velocidad y las imprudencias al volante fueron factores evidenciados en el análisis, es necesario fortalecer las campañas de educación vial dirigidas tanto a conductores como a peatones, resaltando las consecuencias de estas conductas. Adicionalmente, sería pertinente la instalación de señalización más visible y cámaras de control de velocidad en los puntos críticos del tramo.
- Se sugiere establecer un sistema de monitoreo constante en el tramo analizado para evaluar el impacto de las intervenciones implementadas, identificar nuevos riesgos, y ajustar las estrategias de manera oportuna.

Referencias Bibliográficas

- Abdullah, P., & Sipos, T. (2024). Exploring the factors influencing traffic accidents: An analysis of black spots and decision tree for injury severity. *Periodica Polytechnica Transportation Engineering*, 52(1), 33-39. <https://doi.org/10.3311/PPtr.22392>
- Agencia Nacional de Seguridad Vial. (2024b). Datos de accidentalidad vial en Colombia. <https://ansv.maps.arcgis.com/sharing/rest/content/items/c62d149a31ff4899b2fea49340b5fc0a>
- Agencia Nacional de Seguridad Vial. (2024c). Servicio de datos geográficos de accidentalidad vial en Colombia. <https://services9.arcgis.com/cCK0fP0sWCjveNe8/arc>
- Agencia Nacional de Seguridad Vial. (2024a). Histórico de víctimas. Observatorio Nacional de Seguridad Vial. <https://ansv.gov.co/es/observatorio/estad%C3%ADsticas/historico-victimas>
- Andreassen, D. C., & Hoque, M. M. (1986). Intersection accident frequencies. *Traffic Engineering & Control*, 27(10), 514-517.
- Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959-975.

- Balaji, M., Velmurugan, S., & Padma, S. (2024). Economic benefit assessment of black spot improvements. *Lecture Notes in Civil Engineering*, 434, 705-722. https://doi.org/10.1007/978-981-99-6090-3_56
- Baños-Chaparro, J. (2021). Análisis bayesiano de la ideación suicida en adultos peruanos. *MEDISAN*, 25(5), 1158-1170.
- Baranyai, D., & Sipos, T. (2022). Black-spot analysis in Hungary based on kernel density estimation. *Sustainability*, 14(14), 8335. <https://doi.org/10.3390/su14148335>
- Belanger, C. (1994). Estimation of safety of four-legged unsignalized intersections. *Transportation Research Record*, 1467, 23-29.
- Bonera, M., Barabino, B., & Maternini, G. (2022). A straightforward framework for road network screening to Lombardy Region (Italy). *Sustainability*, 14(19). <https://doi.org/10.3390/su141912424>
- Brijs, T., Karlis, D., Van Den Bossche, F., & Wets, G. (2007). A Bayesian model for ranking hazardous road sites. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 170(4), 1001-1017. <https://doi.org/10.1111/j.1467-985X.2007.00486.x>

- Brijs, T., Van Den Bossche, F., Wets, G., & Karlis, D. (2006). A model for identifying and ranking dangerous accident locations: A case study in Flanders. *Statistica Neerlandica*, 60(4), 457-476. <https://doi.org/10.1111/j.1467-9574.2006.00341.x>
- Caicedo, D. M., et al. (2019). Desarrollo de algoritmos clínicos para el diagnóstico del dengue en Colombia. *Biomédica*, 39(1), 170-185. <https://doi.org/10.7705/biomedica.v39i2.3990>
- Davis, G. A., & Yang, S. (2001). Bayesian identification of high-risk intersections for older drivers via Gibbs sampling. *Transportation Research Record*, 1746, 84-89. <https://doi.org/10.3141/1746-11>
- Ghadi, M., & Török, Á. (2017). Comparison of different black spot identification methods. *Transportation Research Procedia*, 25, 1105-1112. <https://doi.org/10.1016/j.trpro.2017.12.104>
- Giraldo-Ospina, C. E., & Giraldo-López, K. (2022). Estrategia de predicción de pacientes con casos de coronavirus; reporte de experiencia Pereira, Risaralda, marzo-abril 2020. *Medicas UIS*, 35(1), 57-69. <https://doi.org/10.18273/revmed.v35n1-2022006>
- Google Earth. (2024). Ubicación del área de estudio [Captura de pantalla de imagen satelital]. <https://earth.google.com/>
- Jakimavičius, M. (2018). Analysis and assessment of Lithuanian road accidents by AHP method. *Baltic Journal of Road and Bridge Engineering*, 13(3), 238-260. <https://doi.org/10.7250/bjrbe.2018-13.414>

- Jiang, X., Abdel-Aty, M., & Alamili, S. (2014). Application of Poisson random effect models for highway network screening. *Accident Analysis & Prevention*, 63, 74-82. <https://doi.org/10.1016/j.aap.2013.10.029>
- Lord, D., Washington, S. P., & Ivan, J. N. (2005). Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: Balancing statistical fit and theory. *Accident Analysis & Prevention*, 37(1), 35-46. <https://doi.org/10.1016/j.aap.2004.02.004>
- MacNab, Y. C. (2003). A Bayesian hierarchical model for accident and injury surveillance. *Accident Analysis & Prevention*, 35(1), 91-102. [https://doi.org/10.1016/S0001-4575\(01\)00093-8](https://doi.org/10.1016/S0001-4575(01)00093-8)
- Miaou, S.-P., & Lum, H. (1993). Modeling vehicle accidents and highway geometric design relationships. *Accident Analysis & Prevention*, 25(6), 689-709. [https://doi.org/10.1016/0001-4575\(93\)90034-T](https://doi.org/10.1016/0001-4575(93)90034-T)
- Miaou, S.-P., & Song, J. J. (2005). Bayesian ranking of sites for engineering safety improvements: Decision parameter, treatability concept, statistical criterion, and spatial dependence. *Accident Analysis & Prevention*, 37(4), 699-720. <https://doi.org/10.1016/j.aap.2005.03.012>
- Ministerio de Transporte. (2021). Transporte en cifras Estadísticas 2020 (1a ed.). <https://mintransporte.gov.co/loader.php?lServicio=Tools2&lTipo=descargas&lFuncion=descargar&idFile=28303>

- Montella, A. (2010). A comparative analysis of hotspot identification methods. *Accident Analysis & Prevention*, 42(2), 571-581. <https://doi.org/10.1016/j.aap.2009.09.025>
- Rodríguez, F. (2025, enero 21). Video Tramo de Estudio Anillo Vial [Video]. YouTube. <https://youtu.be/pOeBRW-Uo1M>
- Russell, S. J., & Norvig, P. (2004). *Inteligencia artificial: un enfoque moderno*. Pearson Educación. Disponible en <https://www-ebooks7-24-com.bibliotecavirtual.uis.edu.co/?il=3299>
- Schlüter, P. J., Deely, J. J., & Nicholson, A. J. (1997). Ranking and selecting motor vehicle accident sites by using a hierarchical Bayesian model. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 46(3), 293-316. <https://doi.org/10.1111/1467-9884.00084>
- Sobral, T., Galvão, T., & Borges, J. (2019). Visualization of urban mobility data from intelligent transportation systems. *Sensors*, 19(2), 332. <https://doi.org/10.3390/s19020332>
- Torres-Avilés, F., Moraga, T., Núñez, L., & Icaza, G. (2015). Mortalidad por cáncer pulmonar en Chile. *Cadernos de Saúde Pública*, 31, 1975-1982. <https://doi.org/10.1590/0102-311X00194313>
- Walpole, R. E., Myers, S. L., Ye, K., & Myers, R. H. (2012). *Probabilidad y estadística para ingeniería y ciencias*. Pearson Educación. Disponible en <https://www-ebooks7-24-com.bibliotecavirtual.uis.edu.co/?il=3375>

World Health Organization. (2023). World Health Organization. <https://www.who.int/es/news-room/factsheets/detail/road-traffic-injuries>

Apéndices

Apéndice A. Código completo del tratamiento de datos realizado en Google Collab

```
import json

import pandas as pd

import numpy as np

archivo_json = 'DatoscompletosColab.json'

data = []

with open(archivo_json, encoding='utf-8') as file:

    for line in file:

        try:

            data.append(json.loads(line))

        except json.JSONDecodeError as e:

            print(f"Error procesando la línea: {e}")

municipios = ['Bucaramanga', 'Giron', 'Piedecuesta', 'Floridablanca']

accidentes = []

for entry in data:

    for feature in entry['features']:

        atributos = feature['attributes']

        municipio = atributos.get('Municipio_Ubicacion', "").strip()

        if municipio in municipios:
```

```
atributos['X'] = round(atributos['X'], 3) # Aprox. 100 metros de precisión
atributos['Y'] = round(atributos['Y'], 3)
accidentes.append(atributos)

df_accidentes = pd.DataFrame(accidentes)
df_accidentes.to_csv('Datosfiltrados.csv', index=False)
df_accidentes.head()

from google.colab import files
files.download('Datosfiltrados.csv')
```

Apéndice B. Código completo del modelo desarrollado mediante Rstudio

```
library(sf)
library(dplyr)
library(brms)
library(ggplot2)
library(parallel)

options(mc.cores = parallel::detectCores())

Mode <- function(x) {
```

```
ux <- unique(x)

ux[which.max(tabulate(match(x, ux)))]

}

datos <- read.csv("C:/Users/jaime/OneDrive/Desktop/Datosfiltrados.csv")

poligono_anillo_vial <- st_polygon(list(rbind(
  c(-73.1634449635897, 7.067735254705352), # Punto A
  c(-73.09608144715092, 7.070498654379905), # Punto B
  c(-73.09628000066337, 7.058281789801652), # Punto C
  c(-73.16295803965895, 7.060462226024185), # Punto D
  c(-73.1634449635897, 7.067735254705352) # Volver al punto A
)))

datos_sf <- st_as_sf(datos, coords = c("X", "Y"), crs = 4326)

poligono_sf <- st_sfc(poligono_anillo_vial, crs = 4326)

datos_anillo_vial <- datos_sf %>%
  st_filter(poligono_sf)

datos_anillo_vial_df <- as.data.frame(st_drop_geometry(datos_anillo_vial))
```

```
datos_anillo_vial_df$lat_round <- round(st_coordinates(datos_anillo_vial)[, 2], 6)
datos_anillo_vial_df$lon_round <- round(st_coordinates(datos_anillo_vial)[, 1], 6)

summary_data <- datos_anillo_vial_df %>%
  group_by(lat_round, lon_round) %>%
  summarise(
    num_accidentes = n(),
    UsuarioVia = Mode(UsuarioVia),
    TipoVehiculoHecho = Mode(TipoVehiculoHecho),
    Circunstancia = Mode(Circunstancia),
    .groups = 'drop'
  )

modelo_anillo <- brm(
  num_accidentes ~ UsuarioVia + TipoVehiculoHecho + Circunstancia + (1 |
lat_round/lon_round),
  data = summary_data,
  family = poisson(),
  chains = 4,
  iter = 10000,
  warmup = 3000,
  control = list(
    adapt_delta = 0.99,
```

```
    max_treedepth = 20
  )
)

summary_data$prediccion_anillo <- fitted(modelo_anillo)[, "Estimate"]

validacion_anillo <- summary_data %>%

  summarise(

    error_medio_absoluto = mean(abs(prediccion_anillo - num_accidentes)),

    error_porcentual_medio = mean(abs((prediccion_anillo - num_accidentes) /
num_accidentes) * 100),

    desviacion_error = sd(prediccion_anillo - num_accidentes)

  )

cat("Resultados de Validación - Anillo Vial\n")

print(validacion_anillo)

ggplot(summary_data, aes(x = lon_round, y = lat_round)) +

  geom_point(aes(color = prediccion_anillo), size = 4, alpha = 0.6) +

  scale_color_viridis_c(option = "magma") +

  labs(

    title = "Predicciones de Accidentes - Anillo Vial",

    x = "Longitud Redondeada",

    y = "Latitud Redondeada",
```

```
    color = "Predicción"

  ) +

  theme_minimal()

tabla_ordenada <- summary_data %>%

  arrange(desc(prediccion_anillo)) %>%

  select(lat_round, lon_round, num_accidentes, prediccion_anillo, UsuarioVia,
TipoVehiculoHecho, Circunstancia)

write.csv(tabla_ordenada,

"C:/Users/jaime/OneDrive/Desktop/tabla_predicciones_accidentes_anillo_vial.csv", row.names =
FALSE)

write.csv(validacion_anillo,

"C:/Users/jaime/OneDrive/Desktop/validacion_modelo_anillo_vial.csv", row.names = FALSE)

saveRDS(modelo_anillo, "C:/Users/jaime/OneDrive/Desktop/modelo_anillo_vial.rds")

cat("Tabla de puntos ordenados por predicción de accidentes (mayor a menor):\n")

print(tabla_ordenada)
```