



RISCOS

territorium 31 (I), 2024, 19-30

journal homepage: <https://territorium.riscos.pt/numeros-publicados/>

DOI: https://doi.org/10.14195/1647-7723_31-1_2

Artigo científico / Scientific article



DESARROLLO DE UN MODELO DE REGRESIÓN LOGÍSTICA PARA EVALUAR EL RIESGO POR INUNDACIÓN
EN LA LOCALIDAD DE FONTIBÓN, BOGOTÁ D.C.-COLOMBIA*

19

DEVELOPMENT OF A LOGISTIC REGRESSION MODEL TO EVALUATE FLOOD RISK
IN THE TOWN OF FONTIBÓN, BOGOTÁ D.C.-COLOMBIA

Carolina Acero

Universidad Distrital Francisco José de Caldas (Colombia)
Facultad de Medio Ambiente y Recursos, Ingeniería Sanitaria
ORCID 0000-0000-0000-0000 caceroc@udistrital.edu.co

Felipe Ortiz

Universidad Santo Tomás (Colombia)
Decano Facultad de Estadística
ORCID 0000-0001-5272-4447 andresortiz@usta.edu.co

RESUMEN

Una parte natural del ciclo hidrológico son las inundaciones derivadas de los sistemas pluviales. Siendo, el peligro natural más frecuente y costoso, el que impacta habitualmente a la mayoría de los países. En las últimas décadas, la frecuencia de eventos de inundación ha aumentado a escala global y con ella las pérdidas económicas. La zona de estudio es la localidad de Fontibón ubicada en el occidente de la ciudad de Bogotá D.C., una de las más vulnerables a inundación debido a su geografía plana, con poca pendiente y baja capacidad para evacuar caudales. Este artículo presenta el desarrollo de un modelo de regresión logística para evaluar el riesgo de inundación en la localidad de Fontibón apoyado en información meteorológica y registros de eventos de inundación en un periodo de 21 años (2000 al 2021). El modelo es altamente sensible con una capacidad de detectar presencias de inundación del 96,15%, los meses con mayor riesgo de inundación son diciembre y enero y para octubre y diciembre del año 2023 la probabilidad de inundación es mayor del 50%.

Palabras clave: Escorrentía, inundación, planicies y riesgo.

ABSTRACT

A natural part of the hydrological cycle is flooding from storm systems, these being the most frequent and expensive natural hazard that usually impacts most countries. Flood events have become more common worldwide in recent decades, and economic losses have increased with them. The study area is Fontibón, a locality in the west of the city of Bogotá D.C., one of the most vulnerable to flooding due to its flat geography, with low slope and low flow evacuation capacity. This article presents the development of a logistic regression model to assess the risk of flooding in the Fontibón locality supported by meteorological information and records of flood events over a period of 21 years (2000 to 2021). The model is highly sensitive with a flood detection capacity of 96.15%; the months with the highest risk of flooding are December and January, and from October to December of the year 2023 the probability of flooding is greater than 50%.

Keywords: Runoff, flood, plains and risk.

* O texto deste artigo corresponde a uma comunicação apresentada no IV Simpósio Ibero-Afro-Americano de Riscos, tendo sido submetido em 27-12-2022, sujeito a revisão por pares a 18-02-2023 e aceite para publicação em 27-03-2024.

Este artigo é parte integrante da Revista *Territorium*, n.º 31 (I), 2024, © Riscos, ISSN: 0872-8941.

Introducción

Las inundaciones son calificadas como el desastre natural más desfavorable y significativo del mundo, aproximadamente corresponden al 50% de las catástrofes, por encima de hambrunas, sequías y epidemias (Douben, 2006). La presencia de eventos de inundación se ha elevado de manera acelerada, especialmente en áreas urbanas afectando el funcionamiento regular de los sectores: social, económico, financiero, político y entre otros, donde los más vulnerables son los habitantes de bajos recursos (Benjamín, 2006).

El riesgo de inundación es uno de los más frecuentes e impactantes sobre las poblaciones, se puede presentar por diferentes causas como el manejo inadecuado de los ríos, la baja capacidad de los sistemas de alcantarillado, la deforestación, entre otros (IDIGER, 2022). Uno de los principales factores que aumenta la amenaza por inundación corresponde al fenómeno de la lluvia, es usual ver que en temporada de lluvias aumentan los eventos de emergencias asociados a inundaciones por encharcamiento (Idiger, 2022). Sumado a eso se evidencian cambios significativos en la intensidad y distribución de las lluvias, asociados con la variabilidad climática y cambio climático, que extienden los periodos de inundación, con tendencia a empeorar (Pinos, Timbe, & Orellana, 2017).

Un proceso clave para la gestión del riesgo por inundación es el desarrollo de modelos de estimación y proyección del comportamiento hidrológico para obtener una reducción eficiente de este (Pinos, Timbe, & Orellana, 2017). Se ha demostrado que los modelos de gestión del riesgo en el mundo contando con criterios de responsabilidad social y sostenibilidad permiten: tener un mayor control de riesgo, contribuyendo a la planificación territorial, mitigación de impactos negativos y análisis de costo beneficio (Carla Rodríguez Cuesta, 2015). En ese sentido la realización del presente artículo de desarrollo de un modelo logístico, para evaluar el riesgo por inundación cobra importancia debido a que favorece la identificación y probabilidad de ocurrencia del evento de inundación para anticipar a los residentes de la zona.

Elementos conceptuales relevantes

El riesgo entendido como la combinación de la probabilidad de que se produzca un evento y sus consecuencias negativas, provocando daños sociales, ambientales y económicos, depende de la confluencia de factores de amenaza y vulnerabilidad (Meza, Corso, & Soza, 2010). Al encontrarse en función de la amenaza y la vulnerabilidad, el riesgo puede disminuirse, aumentarse o concretarse en una emergencia o desastre, con consecuencias para la sociedad; lo que involucra que el riesgo responde a dinámicas naturales y también

a dinámicas sociales que lo configuran y convierten continuamente (Quintero Angel, Carvajal Escobar, & Aldunce, 2012).

Para que se pueda entender el concepto de gestión del riesgo, se debe tener claro la noción de gestión integral del riesgo la cual es el conjunto de elementos, medidas y herramientas encaminadas a intervenir la amenaza o la vulnerabilidad, con el propósito de disminuir o mitigar los riesgos existentes (Oficina Federal de Protección Civil, 2014). La gestión del riesgo tiene como objetivo principal la prevención y mitigación de un desastre. La gestión del riesgo actúa mediante el desarrollo de políticas preventivas que generalmente suelen dar resultados a largo plazo disminuyendo de manera significativa en desastres que están por ocurrir o que inevitablemente ocurren (Omar Darío Cardona Arboleda, 2008).

La mitigación de riesgos es el proceso de desarrollo de opciones y acciones que, al ser implementadas, mejoran las oportunidades y reducen el impacto negativo o la probabilidad de ocurrencia de un evento en particular. La mitigación de riesgos es el eje central de la gestión de riesgos (Escuela Eurpoea por excelencia, 2021). La experiencia señala que son pocos los riesgos que realmente se pueden eliminar. Gracias a las estrategias de mitigación de riesgos es posible aceptar y tolerar riesgos, bajo un ambiente de seguridad (Jauregui, Aversa, & Salas, 2017).

En los últimos años se han presentado distintos modelos y metodologías para evaluar el riesgo de inundación, los cuales son muy heterogéneos entre si y la gran mayoría no incorporan una validación explícita de los resultados, presentando incertidumbres altas y contrastes que a menudo llevan a cuestionar su calidad (Pinos, Timbe, & Orellana, 2017). Sin embargo, a pesar de las limitaciones en el conocimiento del medio y restricciones tecnológicas que impiden las simplificaciones de los sistemas en un modelo, amplias investigaciones buscan la mayor aproximación posible, lo que ha contribuido a mejorar la gestión del riesgo por inundación (Robert, 2000).

Un modelo de regresión logística es un modelo matemático que puede ser utilizado para describir la relación entre un conjunto de variables independientes y una variable dicotómica Y (TIBC, 2022). La regresión logística resulta útil para los casos en los que se desea predecir la presencia o ausencia de una característica o resultado según los valores de un conjunto de predictores (IBM, 2021). Es similar a un modelo de regresión lineal pero está adaptado para modelos en los que la variable dependiente es dicotómica (Dos valores posibles ausencia o presencia del evento de interés) (Moral Peláez, 2014). Los coeficientes de regresión logística pueden utilizarse para estimar la razón de probabilidad de cada variable independiente del modelo (Moral Peláez, 2014).

La regresión logística (RL) es uno de los instrumentos estadísticos más expresivos y versátiles de los cuales se dispone para el análisis de datos. Su origen se remonta a la década de los sesenta (Confield, Gordon y Smith 1961); su uso se universaliza y expande desde principios de los ochenta debido, especialmente, a las facilidades informáticas con que se cuenta desde entonces. En los últimos años se ha verificado una presencia muy marcada de esta técnica, tanto en la literatura orientada a tratar temas metodológicos como en los artículos científicos (Fiuza Pérez & Rodríguez Pérez, 2000).

Enfoques Existentes

Se observa una relación directamente proporcional entre los daños/costos socioeconómicos y las amenazas por inundación debido a que si incrementa una la otra también lo hace, esto se presenta porque más personas viven en condiciones de vulnerabilidad ante estas (Venton & Hansford, 2006 como se citó (Quintero-Angel *et al.*, 2012).

Las medidas de gestión denominadas por Di Baldassarre *et al.* (2015) “Efecto del Dique” son las predominantes en Colombia para ejercer acción y respuesta ante las inundaciones, las cuales no han sido muy efectivas, y se le suma la falta de memoria histórica de la población y la ausencia de estudios específicos hidroclimáticos, físicos y ambientales de las zonas afectadas, haciendo que los eventos de inundación se repitan una y otra vez (Barrios Díaz & Salazar Galán, 2018).

En lo que respecta a los modelos para la evaluación y predicción de riesgo por inundación estos son muy heterogéneos incluyendo tantos procesos y variables como sea posible (Robert, 2000), como: la duración del evento, la temperatura del agua, el tiempo de alerta, la presencia de olas, la época del año, el día de la semana, la hora del día, la topografía, el tipo de propiedad en riesgo y muchos otros factores (Pinos, Timbe, & Orellana, 2017). Que conllevan a diferentes interpretaciones teniendo en cuenta la sensibilidad e incertidumbre.

Estado del Arte

A nivel global se evidencia que las mayores pérdidas económicas consecuencias de las inundaciones se producen en las zonas urbanas, provocando degradación o destrucción de estructuras, deterioro del contenido de las viviendas y comercios, costos del cierre de negocios, fallos en los sistemas de abastecimiento de servicios básicos, entre otros. El incremento de pérdidas por inundaciones en los últimos años ha expuesto la necesidad urgente de mejorar la gestión de este riesgo natural (Pinos, Timbe, & Orellana, 2017).

La gestión de riesgo en Colombia parte de un régimen normativo que coordina las acciones gubernamentales y

territoriales que permite delegar funciones, sustentando así la gestión del riesgo en proyectos y estrategias a corto, mediano y largo plazo desarrollados finalmente por los departamentos, municipios y regiones del país. Entonces, entre los múltiples niveles que se encargan de la gestión del riesgo en Colombia se destaca como actor principal al municipio. Dado que a nivel municipal es donde se gestiona todo el desarrollo del territorio y se generan las regulaciones para la distribución de las poblaciones (Álvarez, 2020).

De manera particular la localidad de Fontibón tiene un “Plan Local de Gestión del Riesgo y Cambio Climático” (Consejo Local de gestión del Riesgo y Cambio Climático, 2019) que va de la mano con el plan de gestión del riesgo de Bogotá que tiene como objetivo general “[...] *incrementar la capacidad del Distrito Capital para confrontar el riesgo de desastres asociados con los fenómenos naturales, socio naturales, tecnológicos, biosanitarios y humanos no intencionales, y los efectos del cambio climático, a través de la implementación de acciones de conocimiento, reducción, manejo del riesgo, mitigación y adaptación al cambio climático, que ayuden con la seguridad, bienestar, la calidad de vida de las personas, el desarrollo bajo en carbono y con capacidad de adaptación al clima*” (IDIGER, 2018b). Actualmente existe una caracterización de los escenarios de riesgo, sin embargo no se evidencian la implementación de modelos para gestionarlos.

El objetivo general de este artículo es desarrollar un modelo de regresión logística para evaluar el riesgo por inundación en la localidad de Fontibón, que permita predecir los meses con mayor riesgo. Se escogen las variables explicativas del fenómeno a través de la información meteorológica a cargo del IDEAM, y el registro de la presencia del evento otorgado por el IDIGER en un periodo de tiempo del año 2000 al 2021, con ayuda del software R Studio se establece el modelo (Ortiz, 2019).

Materiales y Métodos

Descripción del Sitio de Investigación

La zona de estudio es la localidad de Fontibón ubicada en la ciudad de Bogotá distrito capital de Colombia (fig. 1), cuenta con una extensión territorial de 33.32 km² y una población de 424,038 habitantes (CLGR-CC, 2018), se encuentra a proximidades de la cuenca de drenaje del río Fucha y el Aeropuerto Internacional El Dorado (TABLA I). Es altamente propensa al evento de inundación por su geografía plana, alto índice de contaminación y su clima frío con abundantes lluvias (precipitación promedio anual de 1465 mm) (Weather Atlas, 2022) (Fernandez Roa & Afanador, 2013). Además algunos de sus barrios se encuentran por debajo de la cota del río Bogotá (Lozada

Carvalho, 2021) y cuando este incrementa por causas invernales se desborda o genera fuertes presiones sobre los diques, lo cual puede romperlos (Consejo Local de gestión del Riesgo y Cambio Climático, 2019).

Recolección de Información

La información analizada en el artículo se solicita a entidades oficiales y pertenece a las estaciones meteorológicas instaladas en la localidad de Fontibón a cargo del IDEAM que presentaron información completa desde el año 2000 a 2021 (esta información es dada por el IDEAM de manera diaria y para el modelo se transforma en información mensual) y del IDIGER que registra en el mismo periodo de tiempo los eventos de inundación ocurridos en el sector. A continuación, se organiza la información en una base de datos en Excel (*Anexo 1 Base de datos inicial*) para ser procesada en el software R Studio, contando con seis variables explicativas: año, mes, precipitación, temperatura, humedad relativa y pendiente. Sin embargo, variables cualitativas explicativas como: cercanía a ríos o canales, estructuras hidráulicas, topografía del terreno y entre otras no son consideradas en el modelo debido a que estas variables son categorizadas como ficticias o subjetivas, que aumentan el AIC (Criterio de Información Akaike el cual mide la calidad relativa de un modelo estadístico) del modelo haciéndolo menos confiable y disminuyendo su capacidad de predicción.

Es importante resaltar que las estaciones del IDEAM cuentan con un valor significativo de incertidumbre: debido al manejo de equipos por los operarios y/o la suspensión de estaciones por mantenimiento o daño. En el caso de la estación “El Dorado Catam AUT” se evidencia ausencia de información, por eso se decide completar los datos con el software Nasa Power, el cual permite obtener información meteorológica satelital en una coordenada, mes y año determinado, evitando el descarte de datos por parte del modelo y aumentando su representación.

Análisis de Información

El modelo del riesgo por inundación en la localidad de Fontibón determina la presencia y ausencia del evento de inundación, a nivel descriptivo se realizan diagramas de caja (boxplot) con las variables cuantitativas para determinar aquellas que aporten a la explicación del fenómeno, luego se realizan gráficos de dispersión con diferentes variables cuantitativas, desagregando entre presencia o ausencia de inundación; para encontrar la relación existente entre ellas y el evento estudiado.

Se diseña un modelo de regresión logística con el fin de encontrar la relación entre la variable dicotómica (riesgo de inundación) y las variables independientes (precipitación, temperatura, humedad relativa, pendiente del terreno y eventos de inundación), se utilizó el software R Studio, en el *Anexo 2 Código de R Studio* se muestra la codificación usada.

TABLA I - Características de la zona de estudio localidad de Fontibón - Bogotá D.C.

TABLE I - Study area characteristics, Fontibón - Bogotá D.C.

Características	Zona de Estudio
Población	424,038 hab
Densidad Poblacional	127.26 hab/ha
Área	3332 ha
Suelo Urbano	3052 ha
Suelo Rural	0 ha
Suelo de Expansión Urbana	275.3 ha
Altitud	2550 msnm
Hidrología	Río Bogotá, río Fucha, quebrada San Antonio, Humedal Capellanía y humedal del Meandro del Say.
Topografía	Plana
Inclinación Promedio	1.0% a -1.0%
Promedio Temperatura Alta	18.92 °C
Promedio Temperatura Baja	10.33 °C
Humedad Relativa Promedio	82.75%
Precipitación Promedio Anual	1465 mm
Promedio Días de Lluvia Anual	305.4
Estructuras hidráulicas	Estaciones de Bombeo y Pondajes La Alameda, La Rivera y La Navarra, las cuales embalsan las aguas lluvias para depositarlas a través de plantas de bombeo al río Bogotá.
Eventos de inundación del 2000 al 2021 (Lámina de agua superior a 30 cm)	30

Fuente: (Alcaldía de Bogotá, 2021), (Google Earth, 2022), (Weather Atlas, 2022), (Lozada Carvalho, 2021), (IDIGER, 2019) e (IDIGER, 2022).
Source: (Alcaldía de Bogotá, 2021), (Google Earth, 2022), (Weather Atlas, 2022), (Lozada Carvalho, 2021), (IDIGER, 2019) e (IDIGER, 2022).

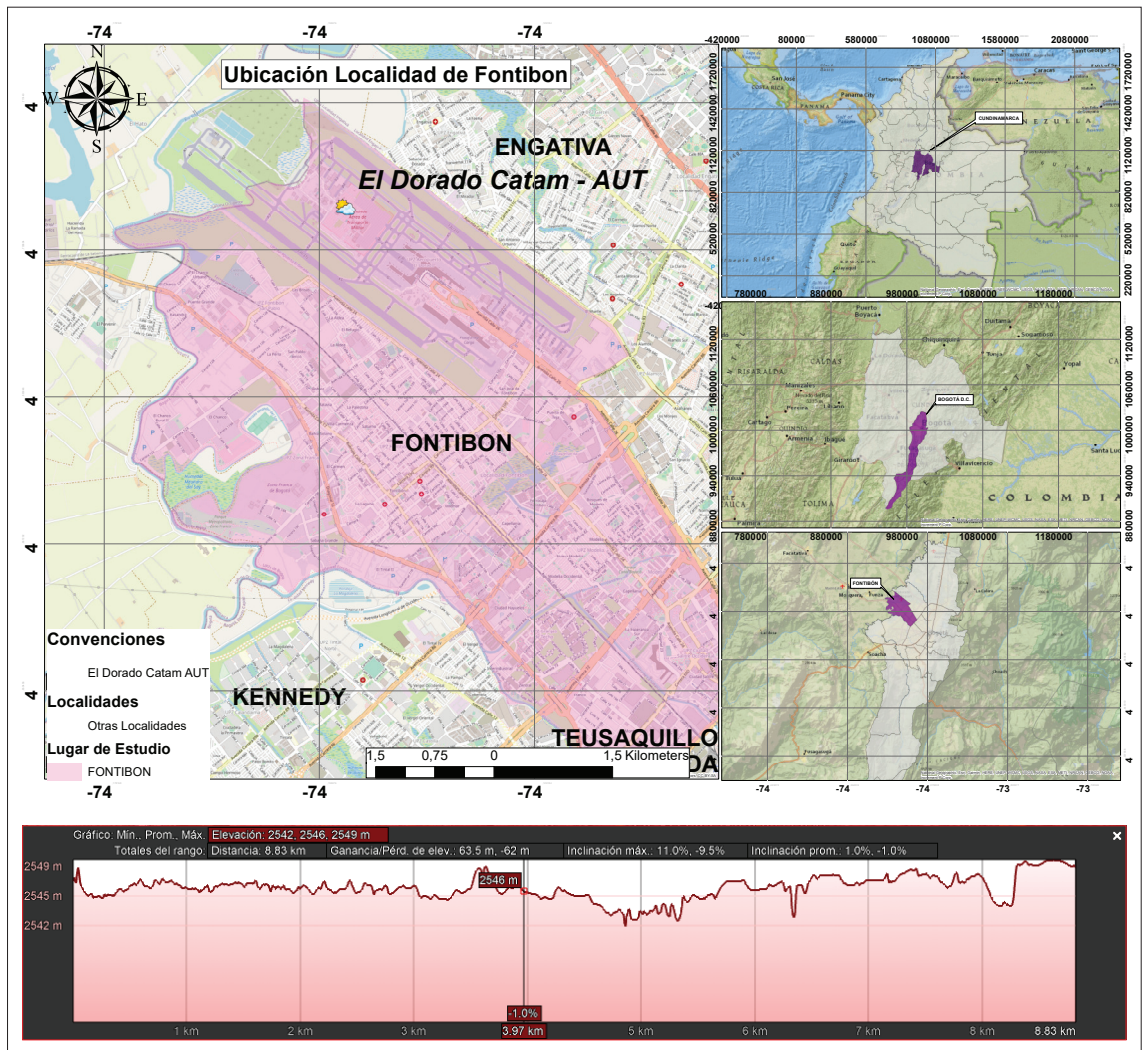


Fig. 1 - Ubicación Zona de Estudio (Fuente mapas base de ESRI:

<https://sites.google.com/site/seriescol/shapes?tmpl=%2Fsystem%2Fapp%2Ftemplates%2Fprint%2F&showPrintDialog>).

Fig. 1 - Location Study Area (Data source ESRI base maps

<https://sites.google.com/site/seriescol/shapes?tmpl=%2Fsystem%2Fapp%2Ftemplates%2Fprint%2F&showPrintDialog>).

Evaluación del Riesgo por Inundación a partir del Modelo de Regresión Logística

Para el modelamiento del riesgo por inundación en una zona de estudio específica se tienen en cuenta diferentes variables empíricas que se basan en la observación y ocurrencia del evento (Abraira, 2022).

$$\text{Logit}(\pi) = \log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki}$$

En ella, el *Logit* de la probabilidad de ocurrencia del evento de inundación, se iguala al predictor del modelo, además surge el cociente: $\left(\frac{\pi}{1-\pi}\right)$ el cual está asociado al riesgo de desarrollar el evento de interés (Ortiz, 2019) (Poma & Mercedes, 2022) (DataScientest, 2021).

Para la estimación, análisis del modelo e interpretación de los riesgos y probabilidades de inundación se tienen en cuenta los siguientes pasos:

Primer paso: Se calculan e interpretan las razones de *Odds*. El *Odds* de un suceso es el cociente de sus probabilidades de ocurrencia entre sus probabilidades de no ocurrencia, bajo unas determinadas condiciones C, lo que permite inferir cuando el riesgo aumenta o disminuye, los *Odds* ratio oscilan entre 0 e infinito (Cardenas, 2015). Razones de *Odds* mayores a uno (1) indican que aumentos en el predictor del modelo se asocian a incrementos en el riesgo y probabilidad de inundación mientras que, razones de *Odds* menores a uno que indiquen disminuciones en el predictor del modelo están asociadas con disminuciones en el riesgo y la probabilidad de inundación (Ferre Jaén, 2019).

Segundo paso: Se establece el punto de corte c con el método de Youden. Este punto de corte se establece para identificar a partir de las probabilidades estimadas, aquellos registros asociados a presencia de inundación (valores mayores al punto de corte) y aquellos asociados a ausencia de inundación (valores menores o iguales al punto de corte). El índice se expresa como la suma de sensibilidad y especificidad del test menos la unidad, de esta manera una de las interpretaciones dadas a los resultados del índice es la ganancia promedio de certidumbre neta en la clasificación de la presencia o ausencia de un evento. Bajo el razonamiento anterior se puede establecer que el índice toma valores entre 0 y 1; siendo 1 un test perfecto y cero un test inútil para detectar o descartar el evento de interés (Estrada Álvarez, 2016) (Brage Escalona, 2020).

Tercer paso: A partir del punto de corte y los registros reales de presencia o ausencia de inundación se construye la matriz de confusión, los verdaderos positivos, los falsos positivos, los falsos negativos y los verdaderos negativos, para establecer la sensibilidad y especificidad que: permiten dar cuenta de un aspecto muy importante del modelo, el cual es la capacidad que tiene para discriminar. La sensibilidad cuantifica la capacidad de detectar la condición de interés cuando la misma se encuentra presente y está dada por la probabilidad de tener un resultado positivo en presencia del evento. La especificidad, por el contrario, da cuenta de la capacidad que tiene el modelo para descartar el evento cuando este es ausente (Estrada Álvarez, 2016) (Sergas, 2014) (Lee, 2019).

Cuarto paso: Con lo anterior se cuantifica la tasa de falsos positivos (β), y la tasa de falsos negativos (α). Dado que α y β están relacionados entre sí (a través del punto de corte), suele ser costumbre reflejar esta relación, representando en el plano una curva, llamada curva ROC (siglas en inglés Receiver Operating Characteristic curve), que relaciona el porcentaje de falsos positivos del modelo (β) (en el eje horizontal) con la sensibilidad del mismo (en el eje vertical). La curva ROC permite evaluar cómo se comporta el modelo para cada punto de corte seleccionado y a partir de ella, es factible deducir cuál es el punto de corte óptimo, lo cual permite dar representación y clasificación al modelo (Estrada Álvarez, 2016) (Deride Silva, 2010).

Quinto paso: Se analiza la existencia de multicolinealidad con el factor de inflación de la varianza VIF. La Multicolinealidad es la correlación alta entre más de dos variables explicativas, lo cual no es un aspecto deseable para el modelo ya que evidencia duplicidad de información a nivel de las variables explicativas, es posible que exista un nivel de correlación entre las variables explicativas del modelo; sin embargo no debe ser alto (Uriel, 2022).

Validación del modelo

La validación del modelo de regresión logística se realiza con la evaluación de la bondad de ajuste, en el presente artículo se realiza con dos métodos:

Método McFadden Pseudo R^2 : el coeficiente de determinación, R^2 , se define como el porcentaje de variación de la variable respuesta que es explicado por el modelo de regresión ajustado (Ruiz & Londoño, 2022), los valores R^2 altos cercanos a 1 indican que una buena parte de la variación es explicada por el modelo (IBM, 2022).

Método Homer-Lemeshow consiste en comparar los valores previstos (esperados) por el modelo con los valores realmente observados. Ambas distribuciones, esperada y observada, se contrastan mediante una prueba de χ^2 . La hipótesis nula del test de Hosmer-Lemeshow es que no hay diferencias entre los valores observados y los valores pronosticados (la aceptación de esta hipótesis con un p-value mayor a 0,05 indicaría que el modelo se encuentra bien ajustado) (FBA Estadística I, 2012).

Resultados y Discusión

Varianza entre variables explicativas del riesgo por inundación según presencia o ausencia del evento

En lo ítem de precipitación se observa una diferencia importante entre el comportamiento de la precipitación bajo la presencia y ausencia de inundación, donde se percibe que si la precipitación es mayor a 100 mm es más probable que exista riesgo de inundación (fig. 2). Esta variable marca una diferencia importante para que el evento ocurra.

En las variables de humedad relativa y temperatura, las cajas son casi simétricas y se encuentran una al lado de la otra para representar ausencia y presencia, esto quiere decir que un rango en estas variables no podría determinar la presencia o ausencia del riesgo por inundación, no son variables determinantes en el modelo, debido a que en el caso de la temperatura un rango entre 16 a 19 °C que es donde se encuentra la mayor concentración de datos, no podrán definir donde tendría el mayor riesgo por inundación.

Se observan valores atípicos en las cajas de precipitación, temperatura y humedad relativa, pero revisando los datos iniciales y el rango se evidencia que los valores no son exagerados ni se encuentran fuera de contexto, es importante resaltar la incidencia de la variabilidad y cambio climáticos sobre estas variables, los cuales causan valores extremos y poco triviales en ciertas temporadas del año. Además, se puede sumar las incertidumbres de medición y manipulación de equipos.

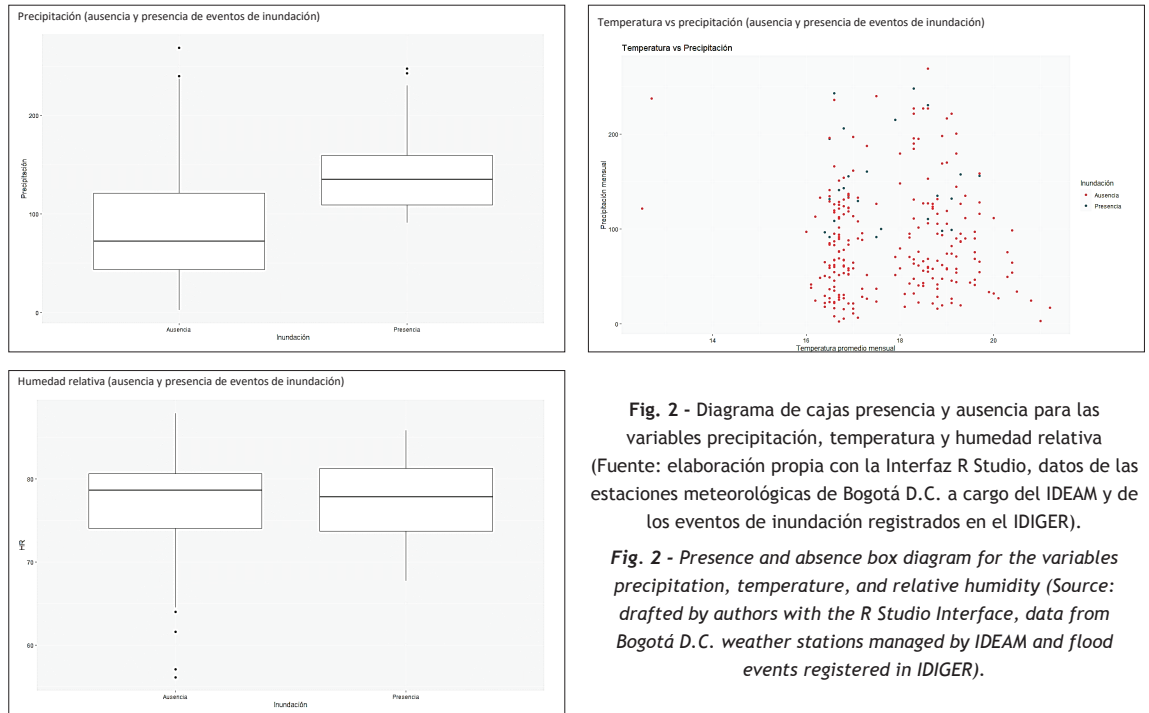


Fig. 2 - Diagrama de cajas presencia y ausencia para las variables precipitación, temperatura y humedad relativa (Fuente: elaboración propia con la Interfaz R Studio, datos de las estaciones meteorológicas de Bogotá D.C. a cargo del IDEAM y de los eventos de inundación registrados en el IDIGER).

Fig. 2 - Presence and absence box diagram for the variables precipitation, temperature, and relative humidity (Source: drafted by authors with the R Studio Interface, data from Bogotá D.C. weather stations managed by IDEAM and flood events registered in IDIGER).

Variables explicativas combinadas según presencia y ausencia del evento de inundación

En lo ítem de Año VS Precipitación no se evidencia una diferencia marcada, debido a que en la mayoría de los años fue probable evidenciar una inundación, sin embargo en la variable de precipitación se puede observar que las presencias señaladas con puntos de

color azul se ven en mayor cantidad si esta es superior a los 100 mm, por consiguiente se detalla ausencia de inundación, con precipitaciones bajas (fig. 3).

En los gráficos que combinan la precipitación vs temperatura y la precipitación vs humedad relativa, se evidencia que la temperatura y humedad relativa no marcan una diferencia sobre la ocurrencia del evento debido a que en

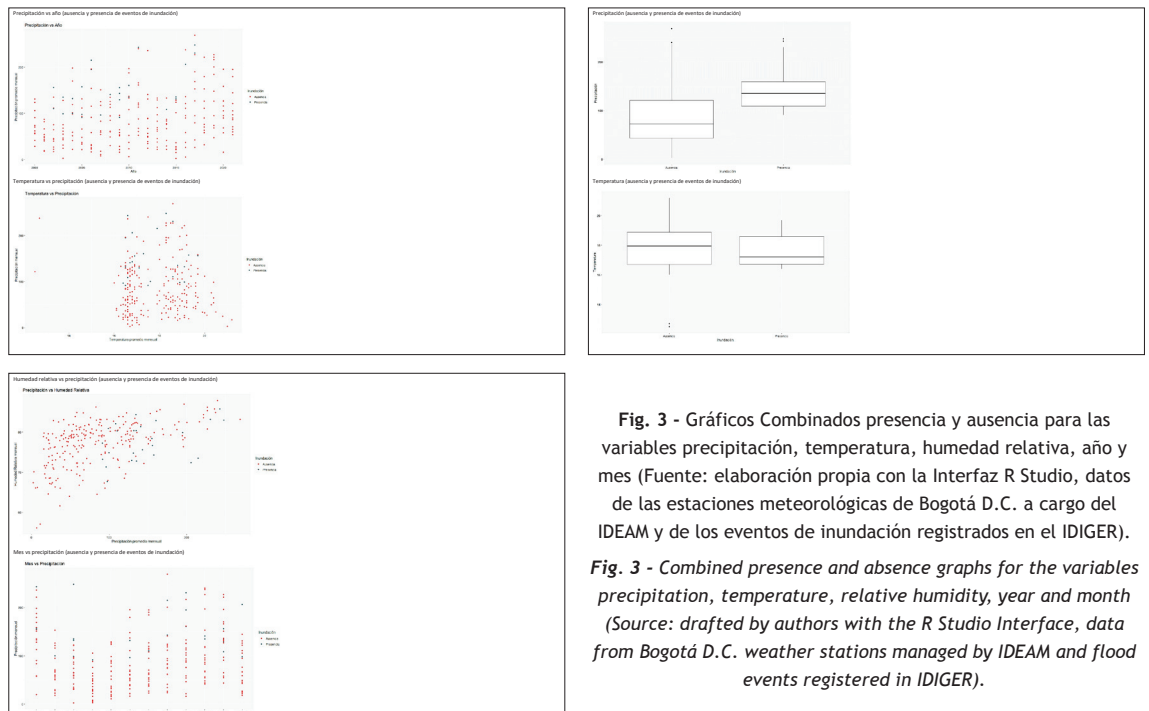


Fig. 3 - Gráficos Combinados presencia y ausencia para las variables precipitación, temperatura, humedad relativa, año y mes (Fuente: elaboración propia con la Interfaz R Studio, datos de las estaciones meteorológicas de Bogotá D.C. a cargo del IDEAM y de los eventos de inundación registrados en el IDIGER).

Fig. 3 - Combined presence and absence graphs for the variables precipitation, temperature, relative humidity, year and month (Source: drafted by authors with the R Studio Interface, data from Bogotá D.C. weather stations managed by IDEAM and flood events registered in IDIGER).

bajas o altas temperaturas y humedades relativas puede ocurrir una inundación. Se especula que posiblemente a mayor humedad relativa aumentan las probabilidades de lluvia y como se evidencia las altas precipitaciones inciden directamente sobre el riesgo de inundación.

Es importante resaltar que los datos meteorológicos cuentan con una incertidumbre significativa que podría estar afectando directamente los resultados del modelo.

Para la variable explicativa “pendiente del terreno” no es representativo ilustrar el grafico debido a que no varía en el periodo de 21 años.

En la fig. 3 en las variables de precipitación vs mes se evidencia mayor ausencia del evento de inundación en los meses de febrero, junio y agosto.

El riesgo por inundación puede aumentar por otros factores que no se tienen en cuenta en la presente investigación, debido a la dificultad de medirlos cuantitativamente, por ejemplo la capacidad de los sistemas de alcantarillado y/o los desbordamientos de cauces.

Modelo logístico

$$Logit(\pi) = Log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_{4j} X_{4ij}$$

$$Logit(\pi) = Log\left(\frac{\pi}{1-\pi}\right) = 429,6 + (-0,2106 * Año) + (0,04094 * P) + (-0,11882 * HR) + (\beta_{4j} * M)$$

Donde:

Variable Respuesta

Logit (π) = Logit de la probabilidad de que el evento de Inundación Ocurra.

$\left(\frac{\pi}{1-\pi}\right)$ = Riesgo de que ocurra el evento de Inundación

Betas

$\beta_0 = 429,6$

$\beta_1 = -0,2106$

$\beta_2 = 0,04094$

$\beta_3 = -0,11882$

β_{4j} = Dependerá del mes, enero=2,94, febrero=-14,77, marzo=-0,067, mayo= 1,37, junio=0,80, julio=1,95, agosto=0,91, septiembre=0,91, octubre=0,82, noviembre=0,74 y diciembre=3,06.

Variables Explicativas

X_{1i} = Año en el año

X_{2i} = Precipitación en el año

X_{3i} = Humedad Relativa j en el año

X_{4i} = Mes en el año , mes

AIC= 141,01 (17 iteraciones)

El modelo logístico descarta las variables de temperatura y pendiente debido a que no las encuentra representativas y aumenta el valor del criterio de Akaike AIC (se considera que cuando el AIC disminuye, mejor es el modelo).

Razón de Odds para medir el riesgo de inundación por variable

A medida que aumentan los años el riesgo de inundación disminuye en un 19 %. Esto puede deberse al esfuerzo de la ciudad por mejorar y controlar esta situación que trae problemas económicos y sociales, al pasar de los años se evidencian acciones que contribuyen a disminuir el riesgo por inundación como: las obras civiles que mejoran la capacidad de los canales, la recuperación hidráulica, adecuación de los sistemas de alcantarillado y entre otras.

Cuando se revisan las relaciones de Odds por meses se evidencia que en febrero, marzo y abril existe un menor riesgo de inundación que en los meses restantes, de mayo a enero el riesgo de inundación es mayor al 100% (TABLA II).

TABLA II - Razón de Odds.

TABLE II - Odds Ratio.

Variab	Razón de Odds	% Riesgo
Año	0,8101	19
Enero	19,059	1805,9
Febrero	0,000	100
Marzo	0,9345	6,55
Abril	0,000	100
Mayo	3,941	294,1
Junio	2,2342	123,42
Julio	7,027	602,7
Agosto	2,488	148,8
Septiembre	2,503	150,3
Octubre	2,275	127,5
Noviembre	2,103	110,3
Diciembre	21,36	2036
Precipitación	1,041	4,1
Humedad Relativa	0,828	17,2

Se evidencia que el mayor riesgo de inundación se presenta en el mes de diciembre con un valor de 2036%, esto puede deberse a la precipitación y a que es el mes con más eventos de inundación registrados por el IDIGER, es un mes vulnerable y al cual debe prestarse atención para generar estrategias de adaptación y mitigación.

Cuando aumenta la precipitación el riesgo de inundación aumenta en 4,1% y cuando aumenta la humedad relativa el riesgo de inundación disminuye en un 17,2%. El cambio de estas variables climáticas interfiere en el riesgo de inundación, y se encuentra relación con la variabilidad climática, debido a que el régimen de precipitación se ve directamente afectado por el aumento de la

humedad relativa y la temperatura, porque más calor equivale a más energía en la atmosfera, lo cual supone el aumento del vapor de agua y la actividad del sistema, contribuyendo a los fenómenos extremos con el incremento de precipitación.

Sensibilidad y Especificidad

Se observan los resultados de la Curva ROC la cual representa la sensibilidad y especificidad del modelo (fig. 4).

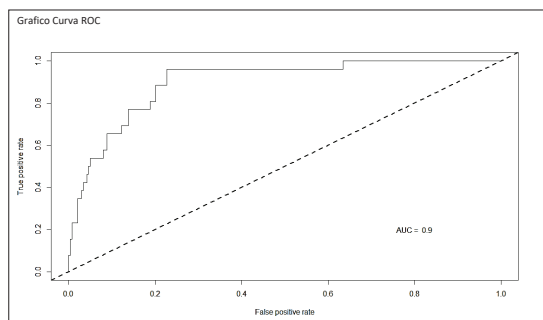


Fig. 4 - Curva ROC (Curva de Operación Relativa) (Fuente: elaboración propia con la Interfaz R Studio, datos de las estaciones meteorológicas de Bogotá D.C. a cargo del IDEAM y de los eventos de inundación registrados en el IDIGER).

Fig. 4 - ROC Curve (Relative Operation Curve) (Source: drafted by authors with the R Studio Interface, data from Bogotá D.C. weather stations managed by IDEAM and flood events registered in IDIGER).

El punto de corte del modelo es de 0,094, lo cual indica que las probabilidades estimadas por encima de ese punto son presencia y por debajo ausencia (TABLA III). A partir de esto se genera la TABLA III donde en las filas se tiene la presencia y ausencia; estimada por el modelo bajo la selección del punto de corte y en las columnas a partir de los datos recolectados. Con esto se observa lo siguiente: los verdaderos positivos son igual a 25, los verdaderos negativos son igual a 184, quiere decir que el modelo clasifico estas ausencias y presencias adecuadamente según los datos registrados. Por el otro lado se tienen 54 falsos positivos donde el modelo los clasifico como presencia y eran ausencia, y 1 falso negativo el cual clasifico como ausencia pero era presencia.

TABLA III - Punto de Corte, Conteo de ausencia y presencia, Sensibilidad y Especificidad del Modelo.

TABLE III - Cutoff Point, Absence and Presence Count, Sensitivity and Specificity of the Model.

Punto de corte	0,09498187	
Estimada/Realidad	Ausencia	Presencia
Ausencia/Falsos	184	1
Presencia/Verdaderos	54	25
Especificidad	0,7731	
Sensibilidad	0,9615	

Con lo anterior se calcula la sensibilidad y especificidad del modelo la cual indica que tan bueno es este para detectar presencias y ausencias respectivamente. Entonces se tiene que el modelo es altamente sensible con una capacidad de detectar presencias de un 96,15% y ausencias de un 77,31%. Por ende el modelo predice de mejor manera la presencia del evento, entonces cuando el modelo predice la presencia de inundación existe una probabilidad del 96,15% de que si haya una inundación. En cambio cuando predice la ausencia no es tan confiable porque existe una probabilidad del 22,69% de que si se presente el evento de inundación y este no fue detectado por el modelo.

En la fig. 4 curva ROC donde se resumen los valores de especificidad (eje y) y sensibilidad (eje x) con el punto de corte c, arroja un valor del área bajo la curva $AUC=0,90$ lo cual lo califica como un modelo “muy bueno” para clasificar el evento hidrometeorológico extremo de inundación (TABLA IV). En este caso existe una buena capacidad del modelo para clasificar la presencia y ausencia de las condiciones necesarias para que se presente una inundación.

TABLA IV - Clasificación del Modelo.

TABLE IV - Model Classification.

AUC	Interpretación
0,5-0,6	Malo
0,6-0,75	Regular
0,75-0,9	Bueno
0,9-0,97	Muy Bueno
0,97-1	Excelente

Entonces se puede decir que con precipitaciones mayores a 100 mm existe una alta probabilidad de que se inunden barrios de la localidad de Fontibón y que el modelo prediga la presencia del evento, además de que en los meses de enero, julio y diciembre existe un alto riesgo de que se presente el evento de inundación.

Ajuste del Modelo

A continuación, se muestra el ajuste del modelo por los métodos de Pseudo R^2 y Hosmer and Lemeshow (TABLA V):

TABLA V - Ajuste del Modelo.

TABLE V - Model Fit.

Ajuste del Modelo	
Pseudo R^2	0,346 - df=15
Hosmer and Lemeshow	$\chi^2=7,3041$ - df=8 - p-value=0,5042

Con el método de Pseudo R^2 , el modelo presenta un ajuste de 0,34 no es un valor bueno, debido a que entre más cercano a 1 el ajuste del modelo es mejor, pero

se resalta que el valor tampoco se encuentra cercano a cero, aunque según el ajuste del modelo por este método refleja que no es bueno para predecir. Sin embargo, con el método de Hosmer and Lemeshow el $\chi^2=7,30$ y $p\text{-value}=0,5042$, indican lo siguiente: χ^2 cuenta con un valor pequeño y $p\text{-value}$ es un valor considerado alto debido a que es mayor de 0,05, que relaciona buenos resultados en la bondad de ajuste del modelo, permitiendo razonar que es un buen modelo predictivo, a su vez este ajuste de modelo es recomendado para modelos como el de la investigación que combina variables continuas y categóricas.

Se presenta el grafico con los valores ajustados donde se evidencia levemente la tendencia a una función logarítmica (fig. 5).

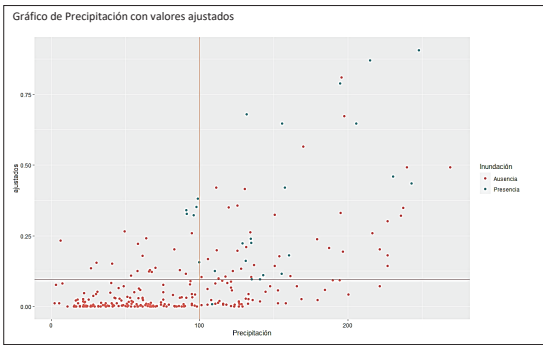


Fig. 5 - Grafico de Precipitación con valores ajustados
(Fuente: elaboración propia con la Interfaz R Studio, datos de las estaciones meteorológicas de Bogotá D.C. a cargo del IDEAM y de los eventos de inundación registrados en el IDIGER).

Fig. 5 - Precipitation graph with adjusted values
(Source: drafted by authors with the R Studio Interface, data from Bogotá D.C. weather stations managed by IDEAM and flood events registered in IDIGER).

Multicolinealidad

Se calcula la multicolinealidad que existe entre las variables explicativas del modelo (TABLA VI):

TABLA VI - Cálculo de Multicolinealidad, Factor VIF.
TABLE VI - Multicollinearity Calculation, VIF Factor.

Variab	Factor VIF
Año	1,414
Mes	1,053
Precipitación	2,079
Humedad Relativa	1,583

En la TABLA VI se calcula la multicolinealidad del modelo entre las variables explicativas con el coeficiente VIF, el cual arroja valores entre 1,053 a 2,079, los cuales no son mayores a 10 indicando que no existe problema de

multicolinealidad entre las variables, esto quiere decir que la correlación entre variables no es exacta ni fuerte, por ende no son redundantes en el modelo y explican fenómenos diferentes.

Utilidad en la gestión del riesgo

El modelo logístico permite predecir a través de los años y por mes la probabilidad de inundación. En el presente artículo se realiza el cálculo para el año 2023, sin embargo, este puede proyectar otros años futuros.

Se debe tener en cuenta lo siguiente: para la precipitación y humedad relativa, de la cual no se tienen valores conocidos se proyectan a través del promedio multianual de los 21 años analizados, para establecer un comportamiento.

Para el cálculo de la probabilidad se utiliza la siguiente formula, la cual está relacionada con el modelo establecido, ejemplo mes de enero:

$$\pi = \frac{1}{1 + e^{-(429,6 + (-0,2106 \cdot \text{Año}) + (0,04094 \cdot P) + (-0,11882 \cdot HR) + (B4 \cdot M))}}$$
$$\pi = \frac{1}{1 + e^{-(429,6 + (-0,2106 \cdot 2023) + (0,04094 \cdot 43,99) + (-0,11882 \cdot 75) + (2,94))}}$$
$$\pi = \frac{1}{1 + e^{-(-0,61)}}$$
$$\pi = \frac{1}{1 + 1,84}$$
$$\pi = 0,35$$

Con lo anterior se establece la probabilidad de inundación para el año 2023 (TABLA VII), donde se tiene lo siguiente:

TABLA VII - Valores de precipitación, humedad relativa, B4 para los meses y cálculo de la probabilidad de inundación para el año 2023.

TABLE VII - Precipitation values, relative humidity, B4 for the months and calculation of the probability of flooding for the year 2023.

	P promedio	HR promedio	B4	Probabilidad
Enero	43,99	75,00	2,94	0,35
Febrero	59,71	73,11	14,77	0,15
Marzo	103,54	76,49	-0,07	0,20
Abril	143,37	80,30		0,47
Mayo	109,05	80,33	1,37	0,46
Junio	87,91	78,33	0,80	0,21
Julio	77,12	76,13	1,95	0,41
Agosto	68,39	74,03	0,91	0,18
Septiembre	77,49	72,99	0,91	0,26
Octubre	129,95	78,49	0,82	0,59
Noviembre	119,41	82,31	0,74	0,35
Diciembre	83,86	79,28	3,06	0,65

Para octubre y diciembre la probabilidad es mayor del 50%, son meses de alto riesgo y vulnerables a la inundación en la localidad de Fontibón, se tiene que en estos meses existe el doble de probabilidad de inundación que en febrero, marzo, junio, agosto y septiembre, esto contribuye a anticiparse al riesgo y tener un cálculo aproximado del comportamiento futuro, para establecer planes de choque, a su vez deben enfocarse en meses con probabilidades que se acercan al 50% porque esto indica que meses como abril, mayo, julio, octubre y diciembre, tienen la mitad de posibilidad de inundarse, la cual es considerable.

Con la estimación de la probabilidad del riesgo, se debe proceder a la prevención y reducción de este. Preparando a las comunidades sobre el potencial riesgo en ciertos meses del año y anticiparse a lo que se pueda proyectar en años futuros.

Conclusiones

La variable explicativa más significativa del modelo es la precipitación, debido a que con lluvias mayores a 100 mm el riesgo de inundación puede aumentar en un 4,1%, además se evidencia en los diagramas de cajas y gráficos combinados que se marca una varianza y diferencia entre la presencia y ausencia del evento de inundación.

Se evidencia que el riesgo de inundación en la localidad de Fontibón aumenta considerablemente en el mes de diciembre, de igual manera en todos los meses se presenta riesgo por inundación, excepto en febrero, marzo y abril los cuales también se identifican por bajas precipitaciones.

El modelo de riesgo de inundación se clasifica como “muy bueno” según su sensibilidad y especificidad para detectar las condiciones necesarias en que se presenta un evento hidrometeorológico extremo de inundación. Además, su ajuste de bondad con el método Hosmer and Lemeshow es igualmente bueno.

Los residuales que no se pueden explicar en el modelo, se deben a otras variables que no se tuvieron en cuenta debido a su dificultad de medición y expresión de manera cuantitativa, entre las cuales cabe mencionar el desbordamiento de cauces, el abastecimiento y capacidad del sistema de alcantarillado, los coeficientes de escorrentía y entre otros, que posiblemente podrían mejorar la predicción del modelo.

El modelo no presenta problemas de multicolinealidad, haciendo referencia que la correlación entre variables no es alta, por ende, las variables no explican el mismo fenómeno y no son redundantes.

El cálculo de probabilidad proyectado por año, es de gran utilidad para anticiparse a situaciones futuras y fortalecer la reducción y el mecanismo de respuesta ante los eventos de inundación.

Referencias

- Abraira, V. (2022). *Modelos de Regresión Logística*. Madrid: Hospital Universitario Ramón y Cajal.
- ALCALDIA DE BOGOTÁ (23 de Septiembre de 2021). *Localidad de Fontibón*. Obtenido de <https://bogota.gov.co/mi-ciudad/localidades/fontibon>
- Barrios Díaz, L., & Salazar Galán, S. (2018). *Una perspectiva de la implementación de un enfoque sociohidrológico para la gestión del riesgo por inundaciones en Colombia*. Universidad Nacional de Colombia, 1-2.
- Brage Escalona, M. (2020). *Análisis de datos categóricos: regresión*. La Laguna, Tenerife: Departamento de Matemáticas, Estadística e Investigación Operativa Universidad de La Laguna.
- Cardenas, J. (1 de Diciembre de 2015). *Odd Ratio*. Obtenido de <https://networkianos.com/odd-ratio-que-es-como-se-interpreta/#toc-3>
- CLGR-CC (Junio de 2018). *Caracterización General de Escenarios de Riesgos*. Obtenido de <https://www.idiger.gov.co/documents/220605/308252/Identificaci%C3%B3n+y+priorizaci%C3%B3n.pdf/6637c155-697a-43df-8fef-c6a8b7f3488e#:~:text=DESCRIPCION%20GENERAL%20DEL%20LOCALIDAD&text=Seguimiento%20del%20estudio%20de%20la,2018%20del%2020424.038%20>
- CONSEJO LOCAL DE GESTIÓN DEL RIESGO Y CAMBIO CLIMÁTICO (2019). *Caracterización General de Escenarios de Riesgo*. Bogota, Localidad de Fontibon: Alcaldía local.
- DataScientest. (16 de Diciembre de 2021). *Regresión Logística*. Obtenido de <https://datascientest.com/es/que-es-la-regresion-logistica>
- Deride Silva, J. (2010). *Universidad de Chile*. Chile: Facultad de Ciencias Químicas y Farmacéuticas.
- ESCUELA EUROPEA POR EXCELENCIA (2021). *Mitigación de riesgos: proceso de 3 pasos para reducir frente al riesgo*.
- Estrada Álvarez, J. (2016). *El índice de Youden y su aplicación a la determinación del punto de corte en un test cuantitativo*. España: Universidad de Granada Departamento de Estadística e Investigación Operativa.
- FBA Estadística I (2012). *Fundamentos de Biología Aplicada I Estadística*. eio.usc.es.
- Fernandez Roa, Y., & Afanador, I. (2013). *Análisis de riesgo por inundación en la zona de Fontibon-Barrio Casandra*. Bogotá D.C.: Universidad Católica.
- Ferre Jaén, M. (04 de Abril de 2019). *Regresión Logística*. Obtenido de <https://gauss.inf.um.es/feir/45/>

- Fiuzza Pérez, M., & Rodríguez Pérez, J. (2000). *La regresión logística: una herramienta versátil*. Islas Canarias-España: Sociedad Española de Nefrología.
- GOOGLE EARTH (30 de Octubre de 2022). Bogotá D.C., Localidad de Fontibón, Colombia.
- IBM (31 de 08 de 2021). *Regresión lineal múltiple*. Obtenido de <https://www.ibm.com/docs/es/cognos-analytics/11.1.0?topic=tests-multiple-linear-regression>
- IBM (13 de Septiembre de 2022). *Pseudo R cuadrado*. Obtenido de <https://www.ibm.com/docs/es/spss-statistics/saas?topic=model-pseudo-r-square>
- IDEAM, & PNUD (2014). Vulnerabilidad de la región capital a los efectos del cambio climático. Documento de Apoyo Para La Toma de Decisiones, 24. URL: <http://repositorio.gestiondelriesgo.gov.co/handle/20.500.11762/19772>
- IDIGER (2019). *Caracterización General de Escenarios de Riesgo Localidad de Fontibón*. Bogotá D.C.: Alcaldía Mayor de Bogotá D.C.
- IDIGER (09 de 06 de 2022). *Caracterización General del Escenario de Riesgo por Inundación*. Obtenido de <https://www.idiger.gov.co/rinundacion>
- IDIGER (2022). *Caracterización General del Escenario de Riesgo por Inundación*. Bogotá D.C.: Instituto distrital de gestión de riesgo y cambio climático.
- Felipe, A., & Rico, O. (n.d.). Seminario de Estadística.
- Jauregui, E., Aversa, M., & Salas, R. (2017). *Estrategias para la mitigación del riesgo por inundación*. Buenos Aires: ISSN 0717 - 3997 / 0718-3607.
- Lee, D.-J. (Marzo de 2019). *Introducción al software estadístico R*. Obtenido de <https://idaejin.github.io/courses/R/2019/euskaltel/>
- Lozada Carvalho, C. (2021). *Plan Ambiental Local de Fontibón 2021-2024*. Localidad de Fontibón - Bogotá: Alcaldía Bogotá D.C.
- Meza, L., Corso, S., & Soza, S. (2010). *Gestión del riesgo de sequía y otros eventos climáticos extremos en Chile*. Chile: Oficina de Estudios y Políticas Agrarias - ODEPA.
- Meinke, H., Nelson, R., Kokic, P., Stone, R., Selvaraju, R., & Baethgen, W. (2006). Actionable climate knowledge: From analysis to synthesis. *Climate Research*, 33(1), 101-110.
DOI: <https://doi.org/10.3354/cr033101>
- Moral Peláez, I. (2014). *Modelos de regresión: lineal simple y regresión logística*. Revista Seden.
- OFICINA FEDERAL DE PROTECCIÓN CIVIL (2014). *Gestión Integral del Riesgo*. Agencia Suiza para el Desarrollo y la Cooperación.
- Pinos, J., Timbe, L., & Orellana, D. (2017). *Métodos para la evaluación del riesgo de inundación fluvial: revisión de literatura y propuesta metodológica para Ecuador*. Departamento de Recursos Hídricos y Ciencias Ambientales, Universidad de Cuenca, 147-162.
- Poma, S., & Mercedes, C. (2022). *Modelo de Regresión Logística*. UNMSM.
- Ortiz, F. (2019). *Regresión Logística*.
- Quintero, Ángel, M., Carvajal Escobar, Y., & Aldunce, P. (2012). *Adaptación A La Variabilidad Y El Cambio Climático: Intersecciones Con La Gestión Del Riesgo*. Manizales-Caldas: Universidad de Caldas.
- R CORE TEAM (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>
- Robert, L. (2000). *One World: The Health and Survival of the Human Species in the 21st Century*. Santa Fe, New Mexico: Published by health Press.
- Ruiz, J., & Londoño, C. (25 de Mayo de 2022). *Pseudo-R2 para GLM*. Obtenido de https://rpubs.com/Cata_lonc/907061
- SECRETARÍA GENERAL DE LA COMUNIDAD ANDINA (2009). Prevención y reducción de riesgos a través de los instrumentos de planificación territorial en Bogotá.
- SERGAS (2014). *Regresión logística*. Soporte Epidat Sergas.
- TIBC (09 de Octubre de 2022). *Regresión Logística*. Obtenido de <https://www.tibco.com/es/reference-center/what-is-logistic-regression>
- URIEL (31 de Octubre de 2022). *Multicolinealidad*. Obtenido de <https://www.uv.es/uriel/material/multicolinealidad3.pdf>
- WEATHER ATLAS (16 de Octubre de 2022). *Clima y previsión meteorológica mensual Fontibón, Colombia*. Obtenido de <https://www.weather-atlas.com/es/colombia/fontibon-clima>