



*Aplicación de un modelo de Markov con covariable para predecir el  
desbordamiento del río Carrizal*

*Application of a Markov model with covariates to predict the flooding of the  
Carrizal River*

*Aplicação de um modelo de Markov com co-variáveis para prever a cheia do Rio  
Carrizal*

Leonardo David Pinargote-Bravo <sup>I</sup>  
[leonardo.pinargote@unach.edu.ec](mailto:leonardo.pinargote@unach.edu.ec)  
<https://orcid.org/0009-0008-7738-2912>

Fernando Molina-Granja <sup>II</sup>  
[fmolina@unach.edu.ec](mailto:fmolina@unach.edu.ec)  
<https://orcid.org/0000-0003-2486-894X>

**Correspondencia:** [leonardo.pinargote@unach.edu.ec](mailto:leonardo.pinargote@unach.edu.ec)

Ciencias Técnicas y Aplicadas  
Artículo de Investigación

\* **Recibido:** 03 de junio de 2025 \* **Aceptado:** 20 de julio de 2025 \* **Publicado:** 19 de agosto de 2025

- I. Ingeniero Ambiental, Universidad Nacional de Chimborazo, Chimborazo, Ecuador.
- II. Doctor en Ingeniería de Sistemas e Informática, Universidad Nacional de Chimborazo, Chimborazo, Ecuador.

## Resumen

Ecuador es un país altamente afectado por fenómenos climatológicos como El Niño, siendo las inundaciones uno de los eventos más comunes. Este estudio propone un modelo probabilístico de cadenas de Markov de primer orden con covariable para predecir el desbordamiento del río Carrizal en la ciudad de Calceta. Se utilizaron datos mensuales de cota del río (variable de respuesta) y precipitaciones acumuladas (covariable predictora), obtenidos del INAMHI, una estación local y del proyecto Climate Hazards Group InfraRed Precipitation with Station. Las cotas fueron clasificadas en 4 categorías mediante el algoritmo K-means, y se implementó una regresión logística multinomial para estimar las probabilidades de transición entre estados. El modelo fue validado con 13 datos recientes, alcanzando una precisión del 63,6%. Se concluye que el modelo es funcional en escala mensual, pero su precisión podría mejorarse con datos de mayor resolución temporal y la inclusión de nuevas covariables. Estos resultados aportan al diseño de sistemas de alerta temprana ante desbordamientos urbanos.

**Palabras clave:** cadenas de Markov; inundaciones; modelo probabilístico; predicción hidrológica; río Carrizal.

## Abstract

Ecuador is a country highly affected by climatic phenomena such as El Niño, with flooding being one of the most common events. This study proposes a first-order probabilistic Markov chain model with covariates to predict the flooding of the Carrizal River in the city of Calceta. Monthly data on river level (response variable) and accumulated precipitation (predictor covariate) were used, obtained from INAMHI, a local station, and the Climate Hazards Group InfraRed Precipitation with Station project. The levels were classified into four categories using the K-means algorithm, and multinomial logistic regression was implemented to estimate the transition probabilities between states. The model was validated with 13 recent data points, achieving an accuracy of 63.6%. It is concluded that the model is functional on a monthly scale, but its accuracy could be improved with higher temporal resolution data and the inclusion of new covariates. These results contribute to the design of early warning systems for urban flooding.

**Keywords:** Markov chains; floods; probabilistic model; hydrological prediction; Carrizal River.

## Resumo

O Equador é um país altamente afetado por fenômenos climáticos como o El Niño, sendo as inundações um dos eventos mais comuns. Este estudo propõe um modelo probabilístico de cadeias de Markov de primeira ordem com co-variáveis para prever as cheias do rio Carrizal na cidade de Calceta. Foram utilizados dados mensais sobre o nível do rio (variável de resposta) e a precipitação acumulada (covariável preditora), obtidos do INAMHI, uma estação local, e do projeto Precipitação Infravermelha com Estação do Climate Hazards Group. Os níveis foram classificados em quatro categorias utilizando o algoritmo K-means, e a regressão logística multinomial foi implementada para estimar as probabilidades de transição entre estados. O modelo foi validado com 13 pontos de dados recentes, alcançando uma precisão de 63,6%. Conclui-se que o modelo é funcional à escala mensal, mas a sua precisão pode ser melhorada com dados de maior resolução temporal e a inclusão de novas co-variáveis. Estes resultados contribuem para o desenho de sistemas de alerta precoce para inundações urbanas.

**Palavras-chave:** Cadeias de Markov; inundações; modelo probabilístico; previsão hidrológica; Rio Carrizal.

## Introducción

Las inundaciones son un peligro natural que conlleva consecuencias devastadoras para la población, aproximadamente un tercio de las pérdidas materiales causadas por desastres naturales son debido a este fenómeno, generando pérdidas de más de 40 mil millones de dólares cada año, y han producido una cantidad de víctimas mortales de 58.700 personas entre 1970 y 2019 (OMM, 2021). Pero las inundaciones no solo causan afectaciones a la población, sino también a los ecosistemas y sitios. Arrighi (2021) analizó el riesgo global en sitios declarados patrimonios protegidos de la UNESCO, determinó que el 35% de las áreas naturales están expuestas a inundaciones. Un factor asociado con el aumento de las inundaciones es el cambio climático (Teodoro et al., 2021), o el fenómeno del Niño, que es conocido por causar inundaciones y severos daños, principalmente en la costa del Pacífico de Sudamérica (Guerrero et al., 2020).

Ecuador es un país altamente afectado por fenómenos climatológicos como El Niño y La Niña, los cuales causan considerables damnificaciones siendo las inundaciones una de las más importantes (Mena et al., 2021). Según el boletín climatológico sobre el fenómeno El Niño del Comité

Internacional para el Estudio Regional del Fenómeno El Niño (2020), en Ecuador se han dado varios eventos de precipitaciones intensas con inundaciones considerables en la zona del litoral, siendo los más importantes los suscitados en los años 1998, 2008, 2016-2017, 2022, y 2023. La predicción de estos eventos es crucial en la minimización del impacto y la respuesta ante la emergencia.

Según Randa et al. (2022) los modelos de inundación y pronóstico han evolucionado rápidamente en los últimos años, convirtiéndose en una herramienta importante en la gestión del riesgo. Sin embargo, Shehadeh y Tucker (2022) mencionan que la predicción de estos eventos es difícil, y a menudo suceden con poca antelación. Por lo tanto, se resalta la importancia del desarrollo de diferentes modelos para la planificación urbana, la mitigación del desastre y la toma de decisiones informadas, lo cual permitirá evaluar la calidad de la predicción de diferentes escenarios de inundación.

Existen distintos enfoques para clasificar los modelos de inundaciones o desbordamientos, desde modelos con enfoques empíricos, hidrodinámicos y conceptuales simplificados (Teng et al., 2017), hasta modelos deterministas y probabilísticos (Thompson y Frazier, 2014). Los modelos probabilísticos; también llamados estocásticos, como las cadenas de Markov, proporcionan una distribución probabilística de la extensión del riesgo (Ávila et al., 2019), además, permiten modelar la incertidumbre inherente en los procesos complejos (Luna y Martínez, 2013). En especial, el modelo de cadenas de Markov es un modelo probabilístico que cumple con la propiedad de Markov, la cual establece que el estado futuro de un sistema depende exclusivamente del estado actual y del tiempo transcurrido (Elavarasan et al., 2018).

Los modelos de Markov han sido ampliamente utilizados en la literatura en diferentes escenarios de procesos climatológicos. En su forma más sencilla se ha implementado para predecir la cota de los ríos definiendo un umbral de desbordamiento (San y Khin, 2016), también se ha utilizado en la predicción de caudales con cadenas de segundo orden; aquellas que dependen del estado anterior del sistema y no solo del actual, lo que mejora ligeramente la precisión del modelo (Yapo et al., 1993). En particular, modelos con covariables han demostrado mejorar la capacidad predictiva en estudios recientes (Islam & Chowdhury, 2017; Stoner & Economou, 2020; Zanin et al., 2022)

Sin embargo, a pesar de los avances significativos en modelos de predicción que usan cadenas de Markov, en la región hay pocas investigaciones referentes a este tema y su aplicabilidad para predecir desastres. Urdiales y Céleri (2018) utilizaron un modelo de Markov con covariable para

predecir precipitaciones en una zona montañosa al sur de Ecuador, donde la covariable fue la ocurrencia del fenómeno del Niño. En México, se usó de manera combinada el modelo de Markov y autómatas celulares para predecir el cambio en el uso de suelo (Reynoso et al., 2016).

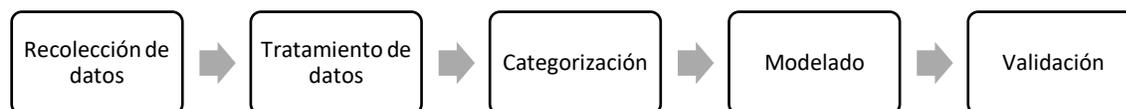
El objetivo de este estudio fue desarrollar y evaluar la eficiencia de un modelo de Markov con covariables para predecir el desbordamiento de un río en una zona urbana, en donde la variable de respuesta fue la cota del río Carrizal, ubicado en Ecuador, Calceta – Manabí, y la covariable es la precipitación. Se utilizaron datos históricos que fueron obtenidos del INAMHI, una estación local e imágenes satelitales de detección infrarrojas, clasificando las cotas por medio del algoritmo K-means, y desarrollando una regresión logística multinomial para estimar las probabilidades de transición. Mediante este trabajo se espera evaluar la aplicabilidad del modelo de cadenas de Markov con covariables en este caso de estudio, y obtener una herramienta útil para la gestión del riesgo y la respuesta temprana ante eventos de desbordamientos en zonas urbanas.

## Metodología

El cantón Bolívar está ubicado en la provincia de Manabí, en Ecuador, posee una superficie de 523,57 Km<sup>2</sup>. Su cabecera cantonal es la ciudad de Calceta y su principal afluente es el río Carrizal (GAD municipal del cantón Bolívar, 2019), el cual es objeto de esta investigación. El río Carrizal es el cauce principal de la subcuenca del Carrizal, y posee una longitud de 80,66 Km, este recorre los cantones de Bolívar y Tosagua, atravesando la cabecera cantonal de Calceta, siendo alimentado por la represa Sixto Durán Ballén que se ubica en el sitio La Esperanza, así como los ríos Tigre, Mosca y Trueno que son sus principales afluentes (Risco & Vera, 2023).

El trabajo consistió en implementar un modelo basado en cadenas de Markov para predecir el desbordamiento del río Carrizal en la zona urbana de la ciudad de Calceta. En este modelo se usan los datos históricos mensuales de las cotas en metros de la época lluviosa (entre los meses de noviembre y mayo), y los datos de precipitaciones acumuladas mensuales en milímetros (mm).

*Figura 1. Diagrama de flujo del estudio*



### Recolección de datos

Los datos se obtuvieron de la base de información del Instituto Nacional de Meteorología e Hidrología (INAMHI). Las cotas fueron registradas de la estación hidrometeorológica “Carrizal” con código H0229 y coordenadas de latitud 80G 11' 0.56" W y longitud 80G 9' 22" W, mientras que los datos de precipitaciones acumuladas mensuales se obtuvieron de la estación meteorológica “ESPAM – MFL” con código M1230, con coordenadas de latitud 0G 49' 22.7" S y longitud 80G 11' 0.56" W.

Para sacar el máximo provecho a la información de cotas, en aquellos periodos sin información de precipitaciones que sí contaban con sus valores de cotas, se utilizaron las precipitaciones del conjunto de datos de Precipitación Infrarroja de Riesgos Climáticos con Datos de Estaciones (CHIRPS por sus siglas en inglés), que obtiene datos de lluvias a través de estaciones meteorológicas y observaciones satelitales infrarrojas.

Los datos obtenidos de CHIRPS y de la estación meteorológica para los meses que sí estaban disponibles, tuvieron una correlación de 0,844 (Correlación fuerte), lo cual indica una alta cohesión para su uso en el modelo.

### Tratamiento de datos

Se realizó un tratamiento de datos con el fin de eliminar valores atípicos e incoherentes, también se eliminaron aquellos valores correspondientes a la época seca la cual no se iba a evaluar. Con esta limpieza el total de muestras fue de 123, de los cuales se dividieron 110 para el entrenamiento del modelo y 13 para la validación.

En los meses sin información de estaciones, se realizó una estimación mediante el método de medias móviles ponderado, siempre que estos estuvieran rodeados por meses que sí tuvieran información (Perry, 2011). Entre las ventajas que presenta este método está la reducción de ruidos, y la descomposición temporal de patrones estacionales al utilizar valores subyacentes (Lewinson, 2023).

El método adaptado utiliza 3 valores de cotas vecinos para estimar el valor de cota central ( $C_i$ ), cuyos subíndices se fueron valorando en función de la disponibilidad de datos.

$$C_i \approx w_1 C_{i+2} + w_2 C_{i+1} + w_3 C_{i-1}$$

Donde  $C_{i+k}$  representan las cotas en el k-ésimo mes siguiente del que se desea estimar, y los parámetros  $w_j$  representan los pesos de ponderación. Los parámetros de ponderación fueron

calculados mediante el método GRG nonlinear que viene implementado en Solver del software Microsoft Excel, aplicado al siguiente sistema.

$$\begin{aligned} \min Z &= \sum_{i=1}^N |C_i - (w_1 C_{i+2} + w_2 C_{i+1} + w_3 C_{i-1})| \\ w_j &\geq 0; \text{ para } j = 1, 2, 3 \\ w_j &\leq 1; \text{ para } j = 1, 2, 3 \\ w_1 + w_2 + w_3 &= 1 \end{aligned}$$

El modelo utiliza los  $N$  datos disponibles de otros años en ese mismo periodo para estimar los parámetros de ponderación, y así proporcionar una aproximación más real a los datos. Para cada contexto y disponibilidad de datos los subíndices fueron variando, y aplicando nuevamente el modelo se recalcularon los parámetros.

Cuando se encontraban más de 3 datos consecutivos faltantes no se estimaron debido a un considerable aumento de error y fueron descartados.

### **Categorización de datos**

Para la categorización de los datos, se utilizó un modelo de clúster K-means, el cual clasifica los grupos a partir de centroides que se van aproximando a una posición de distancia mínima entre los datos más cercanos (Pérez et al., 2019). Sin embargo, el algoritmo no clasifica de manera precisa cuando el río alcanza una cota de desbordamiento, por lo cual se validaron a partir de registros históricos aquellos años en donde se registraron desbordamientos importantes en la zona urbana de Calceta, y se tomó el valor mínimo para obtener un límite de cota crítico de desbordamiento del cauce.

Según registros históricos, los periodos en los cuales se registró un importante aumento del caudal con inundaciones en zonas urbanas de Calceta fueron: en 1988 – 1998 a causa del fenómeno de “El Niño”, en 2002 y 2008 debido a una intensa actividad climática, y en 2012 (Mendoza et al., 2023). Por lo cual, en base al registro de los meses de afectación se observó una cota mínima de desbordamiento de 4,01 metros, con esto el umbral se estableció en 4 metros. De esta forma, los clústeres se realizaron para todos los datos menores que 4, en 3 grupos (Cota baja, Cota media, Cota alta).

## Descripción del modelo

El modelo de Markov utilizado es la versión de primer orden con una sola covariable. Este modelo se caracteriza porque la probabilidad de que un sistema cambie de un estado a otro, depende únicamente de su estado actual y de la covariable que; para este caso, se trata de las precipitaciones.

$$P(X_{t+1} = j | X_t = i, Z_t = z_t) = p_{ij}(z_t)$$

Donde:

$Z_t$  es la covariable exógena en el tiempo  $t$  (precipitaciones).

$p_{ij}$  es una función que modela la transición de estados  $i \rightarrow j$ .

La función a utilizar para las transiciones es una función logística multinomial que modela los cambios desde un estado inicial  $i$  hacia los posibles estados siguientes  $j$  (Islam y Chowdhury, 2017). El modelo estima una regresión logística para cada estado actual hacia los estados siguientes, con la covariable  $Z_t$  como predictor.

$$\log\left(\frac{P(X_{t+1} = j | X_t = i, Z_t)}{P(X_{t+1} = base | X_t = i, Z_t)}\right) = \beta_{ij0} + \beta_{ij1}Z_t$$

Donde se considera  $X_t \in \{1,2,3,4\}$  como los estados del río en un tiempo  $t$ , con  $X_t = i$  siendo el estado actual  $i$ . El modelo de regresión logística puede adaptarse para asegurar que la suma de probabilidades de todos los estados posibles sume 1, esto se logra al dividir cada probabilidad de transición para la probabilidad total, proceso al cual se llama normalización.

$$P(X_{t+1} = j | X_t = i, Z_t = z_t) = \frac{\exp(\beta_{ij0} + \beta_{ij1}z_t)}{\sum_{k=1}^K \exp(\beta_{ik0} + \beta_{ik1}z_t)}$$

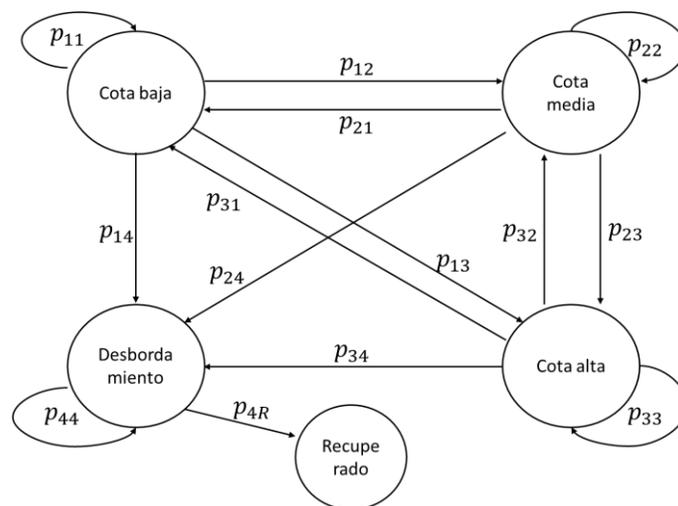
Debido a la baja cantidad de datos para el estado de desbordamiento (5 muestras), cuando  $i = 4$  (estado de desbordamiento), el modelo se ajustó a solo dos posibles estados futuros  $X_{t+1} \in \{1, 2\}$ , es decir, la variable pasa a ser binaria. A partir de que ya se ha dado un desbordamiento las posibilidades para el siguiente estado son: una recuperación del cauce a niveles normales (Estado 2), o continua en desbordamiento (Estado 1), con lo cual, el modelo logit se ajusta a su forma estándar.

$$\log\left(\frac{p(\text{desbordado})}{1 - P(\text{desbordado})}\right) = \beta_0 + \beta_1 z_t$$

Para la obtención de los coeficientes del modelo en cada transición, se utilizó el método de estimación de máxima verosimilitud (Zubaidah y Karnaningroem, 2017). Para cada transición de estados se determinó el supuesto de linealidad de los predictores y los log-odds mediante la prueba

de Box-Tidwell al 5 %, además, se realizó un análisis de sensibilidad mediante la evaluación de valores de probabilidad de los coeficientes. La estimación de los parámetros, las pruebas de validación de supuestos y el análisis de sensibilidad se realizaron mediante códigos de programación de Python 3.12.

**Figura 2.** Cadena de Markov aplicada al modelo de desbordamiento del río



La elección del modelo de primer orden se basó en la cantidad de datos disponibles, ya que un modelo de segundo orden o superior elevaba las dimensiones de la matriz de transición, lo cual demanda mayor cantidad de muestras (Shamshad et al., 2005).

### Validación del modelo

Para la validación del modelo se utilizaron los datos de los últimos 13 meses de temporada húmeda, para un total de 11 transiciones. Como métrica de valoración, se utilizó la precisión global de transición, la cual toma en consideración el estado siguiente con mayor probabilidad de transición y evalúa su similitud con el siguiente estado real de los datos (Arista et al., 2017).

$$Exactitud = \frac{\sum_{t=1}^{n_{val}} 1\{\hat{X}_{t+1} = X_{t+1}\}}{n_{val}}$$

Donde:

$n_{val}$  es el número de transiciones en el conjunto de validación.

$X_{t+1}$  es el estado real observado en el mes  $t + 1$ .

$\hat{X}_{t+1}$  es la predicción del modelo (estado siguiente con probabilidad más alta).

Además, se evaluó la matriz de confusión y con ello la métrica de recall. La validación se realizó mediante el software informático Microsoft Excel.

### Resultados y discusión

Mediante el modelo de media móvil ponderado se logró recuperar 8 datos perdidos, con lo cual se pudo aprovechar la continuidad de varios años que no poseían información completa de cotas.

En el análisis de datos se observó que la cota máxima registrada mensual entre los años 1990 y 2020 fue de 4,91 metros, correspondiente al periodo de mayo de 2012. Así mismo, la precipitación total mensual máxima fue de 474 milímetros correspondiente al periodo de febrero de 2012. Estos resultados muestran que de los años registrados, el de mayor actividad hidrometeorológica en Calceta fue el 2012, cuya evidencia también se observa en artículos de periódicos y reportajes de la época como uno de los eventos más importantes con considerables pérdidas (Palma, 2012).

Con la aplicación del algoritmo de clúster k-means al conjunto de datos menores que 4, se obtuvieron los límites de cada categoría, los cuales quedaron conformados de la siguiente manera.

*Tabla 1. Límites de cota entre un estado a otro*

Categorías	Límites	Promedio
Cota baja	< 1,68	1,26
Cota media	1,68 – 2,63	2,09
Cota alta	2,63 – 4,00	3,17
Desbordamiento	≥ 4,00	4,38

La estimación de los coeficientes permitió realizar un análisis del modelo en función de las variables. Sea  $D_i = 1 + \sum_{j \in \{1,2,3\}} \exp(\beta_{ij0} + \beta_{ij1} Z_t)$  y  $D_4 = 1 + \sum_{j \in \{1,2\}} \exp(\beta_{ij0} + \beta_{ij1} Z_t)$ , entonces la matriz de transición de estados queda conformada de la siguiente manera:

*Tabla 2. Divisores de las ecuaciones de transición para cada estado de partida*

<b>Cota baja</b>	$D_1 = 1 + \exp(-1,588556 - 0,000287 Z_t) + \exp(-4,434188 + 0,014137 Z_t)$
<b>Cota media</b>	$D_2 = 1 + \exp(1,891582 - 0,008803 Z_t) + \exp(-1,587255 + 0,010702 Z_t) + \exp(-7,432544 + 0,025512 Z_t)$
<b>Cota alta</b>	$D_3 = 1 + \exp(0,873482 - 0,010713 Z_t) + \exp(-2,041114 + 0,016970 Z_t) + \exp(-8,996617 + 0,038833 Z_t)$
<b>Desbordamiento</b>	$D_4 = 1 + \exp(3304,798234 - 8,204999 Z_t)$

Tabla 3. Matriz de transición entre estados

Estado $i$ \ Estado $j$	Cota baja	Cota media	Cota alta	Desbordamiento
Cota baja	$p_{11} = \frac{1}{D_1}$	$p_{12} = \frac{\exp(-1,588556 - 0,000287 Z_t)}{D_1}$	$p_{13} = \frac{\exp(-4,438188 + 0,014137 Z_t)}{D_1}$	$p_{14} = 0$
Cota media	$p_{21} = \frac{\exp(1,891582 - 0,008803 Z_t)}{D_2}$	$p_{22} = \frac{1}{D_2}$	$p_{23} = \frac{\exp(-1,587255 + 0,010702 Z_t)}{D_2}$	$p_{24} = \frac{\exp(-7,432544 + 0,025512 Z_t)}{D_2}$
Cota alta	$p_{31} = \frac{\exp(0,873482 - 0,010713 Z_t)}{D_3}$	$p_{32} = \frac{\exp(-2,041114 + 0,016970 Z_t)}{D_3}$	$p_{33} = \frac{1}{D_3}$	$p_{34} = \frac{\exp(-8,996617 + 0,038833 Z_t)}{D_3}$
Desbordamiento	$p_{41} = 0$	$p_{42R} = 1 - p_{44}$	$p_{43R} = 1 - p_{44}$	$p_{44} = \frac{1}{D_4}$

El modelo cumplió con el supuesto de linealidad para todos los casos (prueba de Box-Tidell con p-valor > 5 %), excepto para el estado de Desbordamiento, en el cual existe una posible no linealidad (p-valor = 0,43). Los coeficientes asociados a la variable predictora ( $Z_t$ ), se interpretan según el signo, aquellos valores positivos indican que un aumento de las precipitaciones aumenta la probabilidad relativa de transición de estados, mientras que un coeficiente negativo indica lo contrario.

En el análisis de los coeficientes de la Tabla 3 muestran las posibles transiciones de un estado a otro, la tendencia de transición de Cota baja a Cota media es relativamente baja (coeficientes bajos y negativos), de Cota baja a Cota alta es aún menos probable y hacia Desbordamiento es improbable. De Cota media la tendencia apunta a una mayor probabilidad de mantenerse en ese mismo estado, la transición a estado de Desbordamiento requiere de escenarios de lluvias extremas (coeficiente positivo, pero término constante muy negativo: -7,43). En un estado de Cota alta, se reporta un aumento significativo de la probabilidad relativa, de transición a Desbordamiento si  $Z_t$  es alta.

El análisis de sensibilidad exhaustivo presentó nociones claras sobre los coeficientes, para un estado de Cota baja en transición a Cota alta las precipitaciones no fueron significativas (p-valor=0,955), por su parte, para una transición a de Cota baja a Cota alta las precipitaciones presentaron una alta significancia (p-valor=0,002). Para un estado de Cota media, el modelo solo

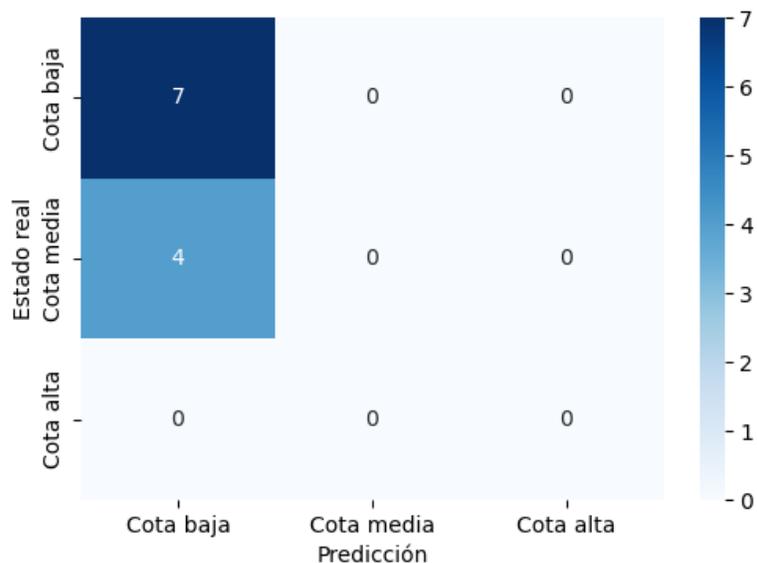
presentó significancia en la constante de transición al mismo estado de Cota media (p-valor=0,011), lo que refuerza la noción de que el estado tiene una tendencia a permanecer invariante, las precipitaciones no mostraron significancia. Para un estado de Cota alta, solo se presentó significancia con ambos coeficientes de transición a estado de Desbordamiento (p-valor=0,044 y p-valor=0,033 respectivamente), lo que indica un efecto importante de la lluvia para pronosticar desbordamientos a partir de un nivel comprendido entre 2,63 a 4 metros.

El sistema muestra una inercia fuerte, el estado de Desbordamiento es casi absorbentes, con una clasificación completa de datos, mientras que los estados intermedios muestran tendencia a mantenerse.

*Tabla 4. Comparación de predicciones con datos de validación*

Fecha	Estado actual $X_t$	Estado siguiente		Precipitaciones (mm) $Z_t$	Probabilidad de transición de estado predicho
		Real $X_{t+1}$	Predicción $\hat{X}_{t+1}$		
01/11/2014	Cota baja	Cota baja	Cota baja	1,00	0,82
01/12/2014	Cota baja	Cota baja	Cota baja	23,40	0,82
01/01/2015	Cota baja	Cota baja	Cota baja	98,00	0,80
01/02/2015	Cota baja	Cota baja	Cota baja	190,40	0,73
01/03/2015	Cota baja	Cota baja	Cota baja	146,00	0,77
01/11/2017	Cota baja	Cota baja	Cota baja	0,00	0,82
01/12/2017	Cota baja	Cota media	Cota baja	23,10	0,82
01/01/2018	Cota media	Cota media	Cota baja	48,72	0,76
01/02/2018	Cota media	Cota media	Cota baja	160,59	0,42
01/03/2018	Cota media	Cota media	Cota baja	133,13	0,52
01/04/2018	Cota media	Cota baja	Cota baja	110,72	0,59

Luego de la validación de datos que se muestra en la Tabla 4, el modelo mostró una exactitud del 63,6%. Cuyo resultado contrasta con la precisión de un modelo de Markov simple de primer orden usado por Islam et al. (2022) para predecir cotas, en el que se obtuvo un porcentaje de 71,3% para una estación, y 95,2% para otra estación en el río Fraser. La variación de resultados puede explicarse debido a la resolución temporal usada, ya que los autores de ese estudio tomaron 18 años de datos diarios. Los intervalos discretos mensuales pudieron afectar la precisión del modelo actual ante la falta de información diaria.

*Figura 3. Matriz de confusión con los datos de validación*

La predicción de un estado de Cota baja mostró ser más relevante comparado con la predicción de un estado de Cota media, además, como se muestra en la Figura 3 el modelo muestra una tendencia a mantenerse en el mismo estado para el mes siguiente. El recall o sensibilidad para los estados de Cota baja y Cota media es de 100% y 0% respectivamente, es decir, solo uno de los estados fue predicho con 100% de exactitud, en Cota media el modelo no fue capaz de predecir ningún estado siguiente.

Los resultados de este estudio se asemejaron a los obtenidos por Mulyawati et al. (2025), quienes obtuvieron una precisión del 63% en la predicción de caudales, similares características se llevaron a cabo en esa investigación en el río Ciberang, en donde se propuso una comparación del modelo de Markov de primer orden con modelos SARIMA y regresión lineal múltiple, la variación temporal discreta también fue mensual.

Sajadifar y Pakseresht (2024) propusieron un modelo de Markov con covariables para estimar la demanda de agua urbana a corto plazo en Teherán, considerando como covariables la temperatura máxima, el consumo de agua y la tasa de precipitación, obteniendo una precisión del 48% y 65%, para este estudio se utilizaron datos diarios entre 2018 y 2021.

En base a los resultados y a la revisión, se observa que la precisión del modelo está afectada por el intervalo temporal utilizado. En ese aspecto, el uso de datos diarios en vez de mensuales podría mejorar la exactitud del modelo, lo cual permitiría unas predicciones más acertadas y confiables.

Una aplicación relevante que conlleva el desarrollo de este modelo, es la implementación de un Sistema de Alerta Temprana ante Emergencias (SATE). Uno de los estudios pioneros en la aplicación de modelos de Markov para evaluar sistemas de alerta tempranas, fue el realizado por Krzysztofowicz (1983), quien propuso un enfoque bayesiano en cadenas de Markov para modelar la incertidumbre y la evolución real del sistema, el tiempo promedio de pronóstico fue de 5 a 13 horas de anticipación, el modelo demostró ser consistente y sentó las bases para los SATE's usando cadenas de Markov. El estudio de SATE's con cadenas de Markov en diversos contextos ha sido recientemente estudiado demostrando robustes (San & Khin, 2016; Shevnina & Silaev, 2019).

### **Conclusiones y recomendaciones**

El modelo de Markov con covariable presentó una precisión global de 63,6% en la predicción de transiciones de niveles del río Carrizal, lo cual es una cifra razonable para datos mensuales. El análisis del modelo permite concluir que las transiciones entre estados más frecuentes (como Cota baja y Cota media), fueron mejor representados en comparación con el estado de Desbordamiento debido a la poca cantidad de datos para esta categoría. Las precipitaciones como variable predictora mejoraron las predicciones de desbordamiento, mostrando significancia en la transición desde un estado de Cota alta.

Una de las principales limitaciones fue la resolución temporal mensual de los datos, lo cual reduce la precisión del modelo y, por ende, la capacidad para realizar predicciones más precisas y confiables. Asimismo, la baja frecuencia de eventos extremos (Desbordamiento) en resolución mensual afectó la calibración del modelo para este estado, lo que genera una representación menos robusta. Otro posible factor que podría mejorar la capacidad predictiva del modelo, es la implementación de nuevas variables exógenas que afecten el nivel del río Carrizal, algunos de estos factores pueden ser: temperatura, uso de suelo, nivel del embalse que alimenta al río, o condiciones hidrológicas previas.

Se recomienda aumentar la resolución temporal del modelo con el fin de mejorar su precisión, además, la implementación de otras covariables como temperatura, humedad, o uso de suelo podría enriquecer el modelo, en caso de no contar con esa información, la aplicación de un modelo de Markov oculto permite captar variables que no pueden ser observadas pero que afectan a las predicciones.

También se recomienda la aplicación del modelo de Markov con covariable combinado con redes neuronales recurrentes para capturar dependencias temporales más complejas.

Como proyección aplicada, se recomienda la implementación de un sistema de alerta temprana en la ciudad Calceta, lo que permitiría a las autoridades locales actualizar las probabilidades de desbordamiento del río Carrizal, y optimizar los mecanismos de prevención y de respuesta en escenario críticos.

## Referencias

1. Arista, A., Calderón, G., Fierro, A., & Nakano, M. (2017). Clasificación de Imágenes Urbanas Aéreas: Comparación entre Descriptores de Bajo Nivel y Aprendizaje Profundo. *Información tecnológica*, 28(3), 209-224. <https://doi.org/10.4067/S0718-07642017000300021>
2. Arrighi, C. (2021). A Global Scale Analysis of River Flood Risk of UNESCO World Heritage Sites. *Frontiers in Water*, 3. <https://doi.org/10.3389/frwa.2021.764459>
3. Ávila, R., Munguía, J., & Ávila, T. (2019). Modelos estocásticos en epidemiología. *Pädi Boletín Científico de Ciencias Básicas e Ingenierías del ICBI*, 6(12). <https://portal.amelica.org/ameli/journal/595/5952961022/>
4. CHIRPS. (2025). Rainfall Estimates from Rain Gauge and Satellite Observations. UC Santa Barbara. <https://www.chc.ucsb.edu/data/chirps>
5. Comité Internacional para el Estudio Regional del Fenómeno El Niño [CIIFEN]. (2020). Boletín climatológico: El Niño/La Niña en América Latina. CIIFEN website. <https://ciifen.org/el-nino-la-nina-ciifen/>
6. Elavarasan, D., Vincent, D. R., Sharma, V., Zomaya, A., & Srinivasan, K. (2018). Forecasting yield by integrating agrarian factors and machine learning models: A survey. *Computers and Electronics in Agriculture*, 155, 257-282. <https://doi.org/10.1016/j.compag.2018.10.024>
7. GAD municipal del cantón Bolívar. (2019). Plan de Desarrollo y Ordenamiento Territorial del Cantón Bolívar. <https://gadbolivar.gob.ec/pdot/>
8. Guerrero, F. A., Luyo, A. V., & Rojas, J. E. (2020). Gestión de la ayuda humanitaria ante inundaciones causadas por el Fenómeno del Niño, mediante la óptima localización de

- sensores en las quebradas y ríos—Caso del desborde del río Piura. *Engineering, Integration, and Alliances for a Sustainable Development. Hemispheric Cooperation for Competitiveness and Prosperity on a Knowledge-Based Economy*, 27-31.  
<https://doi.org/10.18687/LACCEI2020.1.1.125>
9. Instituto Nacional de Meteorología e Hidrología. (2023). Visor de estaciones meteorológicas e hidrológicas. <https://inamhi.gob.ec/info/visor>
  10. Islam, A., & Chowdhury, R. (2017). Covariate-Dependent Markov Models. En A. Islam & R. Chowdhury (Eds.), *Analysis of Repeated Measures Data* (pp. 51-66). Springer.  
[https://doi.org/10.1007/978-981-10-3794-8\\_5](https://doi.org/10.1007/978-981-10-3794-8_5)
  11. Islam, A., Ghaith, M., Hassini, S., & El-Dakhakhni, W. (2022). A Short-Term Flood Forecasting Model Using Markov Chain. En S. Walbridge, M. Nik-Bakht, K. T. W. Ng, M. Shome, M. S. Alam, A. El Damatty, & G. Lovegrove (Eds.), *Proceedings of the Canadian Society of Civil Engineering Annual Conference 2021* (pp. 555-563). Springer Nature. [https://doi.org/10.1007/978-981-19-1065-4\\_46](https://doi.org/10.1007/978-981-19-1065-4_46)
  12. Krzysztofowicz, R. (1983). A Bayesian Markov Model of the flood forecast process. *Water Resources Research*, 19(6), 1455-1465. <https://doi.org/10.1029/WR019i006p01455>
  13. Lewinson, E. (2023, octubre 30). The Comprehensive Guide to Moving Averages in Time Series Analysis. *Towards Data Science*. <https://towardsdatascience.com/the-comprehensive-guide-to-moving-averages-in-time-series-analysis-3fb2baa749a/>
  14. Luna, M., & Martínez, E. (2013). Model-based run-to-run optimization under uncertainty of biodiesel production. En A. Kraslawski & I. Turunen (Eds.), *Computer Aided Chemical Engineering* (Vol. 32, pp. 103-108). Elsevier. <https://doi.org/10.1016/B978-0-444-63234-0.50018-X>
  15. Mena, M., Scheffczyk, K., Urrutia, M., Huerte, B., & Walz, Y. (2021). Evaluación del riesgo de inundación en Ecuador. Institute for Environment and Human Security. [https://collections.unu.edu/eserv/UNU:8434/VALE\\_Flood\\_Risk\\_Assesment\\_Report\\_Ecuador\\_SPA\\_FINAL\\_META.pdf](https://collections.unu.edu/eserv/UNU:8434/VALE_Flood_Risk_Assesment_Report_Ecuador_SPA_FINAL_META.pdf)
  16. Mendoza, M., León, E. E., Díaz, A. M., & Álava, J. A. (2023). Análisis de riesgo y vulnerabilidad frente a inundaciones en sectores de la ciudad de Calceta. *Revista Científica Arbitrada Multidisciplinaria PENTACIENCIAS*, 5(7), 410-423.  
<https://doi.org/10.59169/pentaciencias.v5i7.966>

17. Mulyawati, I., Salami, I. R. S., Marselina, M., & Sabar, A. (2025). A comparison of Markov Chain, Sarima and multiple linear regression for forecasting discharge. *International Journal of Innovative Research and Scientific Studies*, 8(3), Article 3. <https://doi.org/10.53894/ijirss.v8i3.7465>
18. OECD. (2016, julio 28). *Financial Management of Flood Risk*. OECD Publishing. [https://www.oecd.org/en/publications/financial-management-of-flood-risk\\_9789264257689-en.html](https://www.oecd.org/en/publications/financial-management-of-flood-risk_9789264257689-en.html)
19. Organización Meteorológica Mundial [OMM]. (2021). Atlas de la OMM sobre mortalidad y pérdidas económicas debidas a fenómenos meteorológicos, climáticos e hidrológicos extremos. Publicaciones de la OMM. [https://www.unclearn.org/wp-content/uploads/library/1267\\_Atlas\\_of\\_Mortality\\_es.pdf](https://www.unclearn.org/wp-content/uploads/library/1267_Atlas_of_Mortality_es.pdf)
20. Osman, S., Aziz, N. A., Husaif, N., Sidek, L. M., Shakirah, A., Hanum, F., & Basri, H. (2018). Application of Stochastic Flood Forecasting Model Using Regression Method for Kelantan Catchment. *MATEC Web of Conferences*, 203, 07001. <https://doi.org/10.1051/mateconf/201820307001>
21. Palma, N. (2012, marzo 22). 1.702 personas fueron evacuadas en Calceta por inundación. *Periódico El Universo*. <https://www.eluniverso.com/2012/03/22/1/1447/1702-personas-fueron-evacuadas-calceta-tras-inundacion.html/>
22. Pérez, J., Almanza, N. N., Vega, A., Pazos, R., Zavala, C., & Martínez, A. (2019). The K-Means Algorithm Evolution. En *Introduction to Data Science and Machine Learning*. IntechOpen. <https://doi.org/10.5772/intechopen.85447>
23. Perry, M. (2011). The Weighted Moving Average Technique. En *Wiley Encyclopedia of Operations Research and Management Science*. John Wiley & Sons, Ltd. <https://doi.org/10.1002/9780470400531.eorms0964>
24. Randa, T., Krhoda, G., Atela, J., & Akala, H. (2022). Review of flood modelling and models in developing cities and informal settlements: A case of Nairobi city. *Journal of Hydrology: Regional Studies*, 43, 101188. <https://doi.org/10.1016/j.ejrh.2022.101188>
25. Reynoso, R., Valdez, J., Escalona, M., De los Santos, H., & Pérez, M. (2016). Cadenas de Markov y autómatas celulares para la modelación de cambio de uso de suelo. *Ingeniería Hidraulica y Ambiental*, 27(1), 72-81. [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S1680-03382016000100006](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1680-03382016000100006)

26. Risco, K. P., & Vera, H. P. (2023). Relación de la red de drenaje y la evolución hipsométrica con las características morfométricas de la subcuenca del río Carrizal [Tesis de licenciatura, Escuela Superior Politécnica Agropecuaria de Manabí].  
<http://repositorio.espam.edu.ec/handle/42000/2271>
27. Sajadifar, S., & Pakseresht, M. (2024). A Probabilistic Markov Chain Model for Short-term Water Demand Forecasting. *Journal of Water Management Modeling*.  
<https://doi.org/10.14796/JWMM.C519>
28. San, T. H., & Khin, M. M. (2016). River Flood Prediction Using Markov Model. En *Genetic and Evolutionary Computing* (pp. 435-443). Springer, Cham.  
[https://doi.org/10.1007/978-3-319-23204-1\\_44](https://doi.org/10.1007/978-3-319-23204-1_44)
29. San, T., & Khin, M. M. (2016). River Flood Prediction Using Markov Model. En T. T. Zin, J. C.-W. Lin, J.-S. Pan, P. Tin, & M. Yokota (Eds.), *Genetic and Evolutionary Computing* (pp. 435-443). Springer International Publishing. [https://doi.org/10.1007/978-3-319-23204-1\\_44](https://doi.org/10.1007/978-3-319-23204-1_44)
30. Shamshad, A., Bawadi, M. A., Wan Hussin, W. M. A., Majid, T. A., & Sanusi, S. A. M. (2005). First and second order Markov chain models for synthetic generation of wind speed time series. *Energy*, 30(5), 693-708. <https://doi.org/10.1016/j.energy.2004.05.026>
31. Shehadeh, K., & Tucker, E. (2022). Stochastic optimization models for location and inventory prepositioning of disaster relief supplies. *Transportation Research Part C: Emerging Technologies*, 144, 103871. <https://doi.org/10.1016/j.trc.2022.103871>
32. Shevnina, E., & Silaev, A. (2019). The probabilistic hydrological MARCS HYDRO (the Markov Chain System) model: Its structure and core version 0.2. *Geoscientific Model Development*, 12(7), 2767-2780. <https://doi.org/10.5194/gmd-12-2767-2019>
33. Teng, J., Jakeman, A., Vaze, J., Croke, B., Dutta, D., & Kim, S. (2017). Flood inundation modelling: A review of methods, recent advances and uncertainty analysis. *Environmental Modelling & Software*, 90, 201-216.  
<https://doi.org/10.1016/j.envsoft.2017.01.006>
34. Teodoro, T. A., Reboita, M. S., Llopart, M., da Rocha, R. P., & Ashfaq, M. (2021). Climate Change Impacts on the South American Monsoon System and Its Surface–Atmosphere Processes Through RegCM4 CORDEX-CORE Projections. *Earth Systems and Environment*, 5(4), 825-847. <https://doi.org/10.1007/s41748-021-00265-y>

35. Thompson, C., & Frazier, T. (2014). Deterministic and probabilistic flood modeling for contemporary and future coastal and inland precipitation inundation. *Applied Geography*, 50, 1-14. <https://doi.org/10.1016/j.apgeog.2014.01.013>
36. Urdiales, D., & Célleri, R. (2018). Pronóstico de la ocurrencia de precipitación diaria en un gradiente altitudinal en el Sur del Ecuador, mediante un generador climático. *Enfoque UTE*, 9(3), 29-41.
37. Yapo, P., Sorooshian, S., & Gupta, V. (1993). A Markov Chain Flow Model for flood forecasting. *Water Resources Research*, 29(7), 2427-2436. <https://doi.org/10.1029/93WR00494>
38. Zubaidah, T., & Karnaningroem, N. (2017). K-means Method for Clustering Water Quality Status on The Rivers of Banjarmasin, Indonesia. Department of Environmental Engineering, Institut Teknologi Sepuluh Nopember. <https://osf.io/s9n2u/download/?format=pdf>

© 2025 por los autores. Este artículo es de acceso abierto y distribuido según los términos y condiciones de la licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional (CC BY-NC-SA 4.0) (<https://creativecommons.org/licenses/by-nc-sa/4.0/>).