



Informe mensual de actividades

Anexo WP1.3.2 Modelo de Alertas para
movimientos en masa detonados por lluvias -
Equipo Geotecnia

Enero de 2026

Contrato
CCT 261 de 2025

Contenido

WP1.3	Información para la gestión y herramientas operacionales que permitan la consolidación del observatorio de amenazas y riesgos	2
	Modelo de Alertas por Movimientos en Masa detonados por lluvias	2
	Avances Mensuales	18
Referencias		45

WP1

Observatorio de amenazas y riesgos

Este WP tiene como objetivo generar conocimiento que sirva como insumo para la alerta temprana y la toma de decisiones a partir del monitoreo, análisis, modelación y generación de alarmas y alertas sobre múltiples componentes físicos, a través de actividades de investigación y operaciones que permitan avanzar en la caracterización, entendimiento, seguimiento y pronóstico de múltiples amenazas ambientales y evaluación de riesgos asociados.

Se avanzará en la consolidación del observatorio de múltiples amenazas, el cual es fundamental dada la variedad de amenazas que pueden impactar la sociedad y el territorio. Para este fin, el WP1 está constituido por tres grandes grupos de actividades: WP1.1 - Caracterización y monitoreo de variables y procesos asociados con las amenazas. El grupo WP1.2 - Modelación y pronóstico de la calidad del aire, las lluvias y otros fenómenos asociados con las amenazas. Por último, el grupo WP1.3 - Información para la gestión y herramientas operacionales que permitan la consolidación del observatorio de amenazas y riesgos en su primera fase, para el apoyo a emergencias, se enfoca en la aplicación (transferencia) de esta investigación a las operaciones para generar alertas y apoyar la gestión, a través de la consolidación de un observatorio multi-amenaza.

WP1.3

Información para la gestión y herramientas operacionales que permitan la consolidación del observatorio de amenazas y riesgos

Modelo de Alertas por Movimientos en Masa detonados por lluvias

Introducción

Los movimientos en masa representan una de las amenazas naturales más significativas, que causan extensas pérdidas de vidas y daños ambientales cada año (Petley, 2012). La lluvia es uno de los factores más comunes que desencadenan estos eventos (Guzzetti et al., 2000; Kirschbaum et al., 2010; Petley, 2012). A nivel regional, una de las estrategias más efectivas para estudiar la lluvia que puede provocar un movimiento en masa es a través de los umbrales de lluvia. Un umbral se define como la cantidad mínima de precipitación a partir de la cual es probable que se produzcan movimientos en masa (Reichenbach et al., 1998; Guzzetti et al., 2007; Segoni et al., 2015). Por esta razón, al desarrollar herramientas para sistemas de alerta temprana de movimientos en masa, los umbrales de lluvia se consolidan como un instrumento esencial. Una de las ventajas de los sistemas de alerta temprana basados en umbrales críticos de lluvia radica en que el monitoreo de la precipitación es relativamente sencillo y económicamente eficiente en extensas zonas geográficas (IEWP, 2005).

Los umbrales de lluvia se establecen mediante diversas técnicas y metodologías. Fundamentalmente, hay algunos enfoques para su cálculo: mediante modelos estadísticos y modelos físicos (Guzzetti et al., 2000; Kirschbaum et al., 2010; Petley, 2012; Wu et al., 2015).

Los umbrales estadísticos se determinan a partir de la recopilación de datos históricos de precipitaciones, que pueden estar o no vinculados a registros históricos de movimientos en masa. Este análisis implica una revisión estadística detallada de los registros antiguos para identificar la correlación entre la ocurrencia de deslizamientos y las lluvias que los provocaron (Gariano et al., 2015; Wu et al., 2015). Los umbrales definidos a través de métodos estadísticos suelen utilizar modelos frecuentistas (Brunetti et al., 2010; Peruccacci et al., 2012, 2017), modelos bayesianos (Guzzetti et al., 2008) y la probabilidad condicionada (Berti et al., 2012).

Por otro lado, los umbrales de lluvia establecidos mediante modelos físicos se basan en procesos físicos y utilizan métodos numéricos. Estos modelos analizan principalmente la interacción entre las precipitaciones y las condiciones del terreno, utilizando modelos hidrológicos y de estabilidad para establecer dicha relación (Iverson, 2000; Guzzetti et al., 2008; Baum et al., 2008; Alvioli et al., 2014). Para la elaboración de estos umbrales, es esencial contar con datos detallados sobre las condiciones hidrológicas, la geología, la morfología y las características específicas de los suelos que influyen directamente en la activación de los movimientos en masa (Aleotti, 2004; Crosta and Frattini, 2001).

En Colombia, aún no se ha establecido una metodología específica y detallada para implementar un Sistema de Alerta Temprana regional dirigido a movimientos en masa inducidos por lluvias. Si bien en el Valle de Aburrá, al igual que en otras partes del país, se han llevado a cabo investigaciones sobre la temática, la relación entre las condiciones de lluvia y los deslizamientos ha sido insuficientemente estudiada (Hidalgo and Vega, 2014).

Moreno et al. (2006) fueron pioneros en modelar la influencia de la lluvia en los movimientos en masa en Antioquia. Propusieron umbrales de precipitación basados en la lluvia acumulada durante tres días anteriores y los 15 días que le siguen, relacionando estos valores con la ocurrencia de deslizamientos. Por su parte, Aristizábal et al. (2010) determinaron umbrales específicos para el Valle de Aburrá, sugiriendo combinaciones de lluvias acumuladas de 0 a 100 mm en 5 días con lluvias antecedentes de 60 mm en 30 días, 160 mm en 60 días y 200 mm en 90 días. Recientemente, Marin and Velásquez (2020) presenta una metodología que establece umbrales basados en la Intensidad-Duración de la lluvia, utilizando un modelo físico aplicándolo a 108 subcuencas del Valle de Aburrá. A pesar de los estudios mencionados, aún no se ha consolidado una herramienta que opere en tiempo real y que proporcione información acerca de las lluvias que pueden desencadenar movimientos en masa en la región. Si bien actualmente el IDEAM presenta un mapa diario de las condiciones de susceptibilidad por movimientos en masa para cada municipio, teniendo en cuenta la lluvia diaria, no existe documentación que presente una metodología replicable en otros entornos y escalas.

Este documento presenta un resumen general de la metodología desarrollada para un modelo regional de alertas por movimientos en masa, basado en la lluvia antecedente. Dicho modelo, formulado en vigencias anteriores y actualmente operado diariamente por el equipo de geotecnia, genera mapas de probabilidad de ocurrencia de movimientos en masa clasificados por barrios y veredas. Además, se describen los análisis y procedimientos implementados para reducir la incertidumbre y ofrecer productos más confiables. Finalmente, se resalta la importancia de continuar evaluando las limitaciones del modelo e incorporar nuevas metodologías que fortalezcan su mejora continua.

Metodología

En la Figura 1 se presenta un esquema general de la metodología propuesta para el modelo de alertas. Esta se compone de cuatro pasos principales: **1) Procesamiento de los datos, 2) Ajuste y definición de umbrales, 3) Espacialización, y 4) Generación del mapa de alertas.** Es posible que con el tiempo el modelo se ajuste y algunos de estos componentes se modifiquen. A continuación, se explica en detalle cada uno de los pasos.

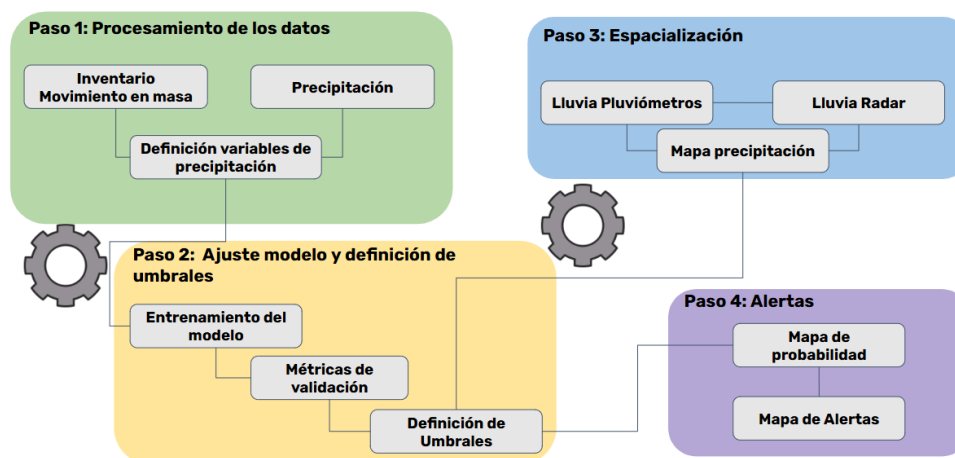


Figura 1. Esquema general de la metodología propuesta para el modelo de alertas por movimientos en masa detonados por lluvias.

1. Procesamiento de datos

Inventario de movimientos en masa: Se utilizaron dos fuentes principales para esta base de datos: el Departamento Administrativo de Gestión del Riesgo y Desastres (DAGR) y el Semillero de Investigación Geohazards de la Universidad Nacional de Colombia, sede Medellín. El DAGRD proporciona datos de emergencias reportadas a través de la línea de atención 123, cubriendo el período de septiembre de 2020 a marzo de 2023, con 4,153 eventos. El Semillero Geohazards complementa esta información con registros desde 1970 hasta 2023, para los demás municipios del Valle de Aburrá, fuera de Medellín.

Se seleccionó el período de 2013 a 2023 para alinear con los registros históricos de lluvias de SIATA. Se realizó una depuración, asegurando que los eventos seleccionados se relacionaban principalmente con lluvias y no con otras causas, como factores antrópicos. Para esto, se seleccionaron aquellos eventos que tuvieron más de 20 mm de acumulado en el día previo a la ocurrencia del evento y que tuviera más de 300 mm de acumulados durante los 90 días antecedentes. Finalmente, se obtuvieron 801 eventos relacionados principalmente con lluvias. El 97 % de estos eventos provienen del inventario del DAGRD, destacando la importancia de Medellín debido a su densidad poblacional y su avanzada sistematización en el registro de eventos de movimientos en masa. La Figura 2 muestra la distribución espacial de los eventos del inventario recopilado.

A pesar de la rigurosa depuración realizada en los datos de movimientos en masa y precipitación, existen limitaciones significativas que deben considerarse al interpretar y valorar los resultados obtenidos. Estas limitaciones incluyen incertidumbres en la ubicación espacial y temporal de los movimientos en masa, la calidad de los datos de lluvia registrados, el área de influencia de las estaciones pluviométricas y las incertidumbres inherentes a las metodologías aplicadas. Aunque se asume que los movimientos en masa analizados fueron desencadenados por lluvias, otros factores, como la actividad antrópica, podrían haber contribuido. La exactitud en la fecha y hora de cada evento está sujeta a la precisión del organismo o entidad que completó el registro, lo cual puede introducir un grado de subjetividad. Asimismo, la localización de los eventos no siempre refleja con exactitud el lugar del suceso; en algunos casos, se basa en direcciones o áreas conocidas, con un margen de error que podría alcanzar decenas de metros.

Datos de precipitación: Para el análisis de la precipitación, se tomó el registro pluviómetro de la red de sensores del proyecto SIATA. Se seleccionaron 35 estaciones con el registro histórico más extenso, que abarca desde el año 2012 hasta la actualidad. Estas estaciones fueron las primeras instaladas por el proyecto y registran datos cada minuto. La Figura 2 muestra la ubicación de los pluviómetros utilizados en los análisis, junto con la proximidad de los eventos del inventario de movimientos en masa. Para correlacionar los eventos de lluvia con los movimientos en masa del inventario, se consideraron los eventos que ocurrieron dentro de un radio de 5 km alrededor de cada pluviómetro. En casos donde un evento estuviera en el rango de dos estaciones (dentro del radio de 5 km), se asoció al pluviómetro más cercano al sitio del evento. Así, cada evento del inventario tiene asignado el pluviómetro más próximo.

Algunas limitaciones del registro pluviométrico se asocian a la variación en la altura de las estaciones pluviométricas, aumentando la incertidumbre sobre la lluvia específica que desencadenó el movimiento en masa. Además, la calidad de los datos de lluvia puede verse afectada por diversos factores, como daños en las estaciones, fallos de energía eléctrica del lugar o fallos en el registro y almacenamiento de los datos, lo que impide una representación precisa de las precipitaciones reales que provocaron los eventos.

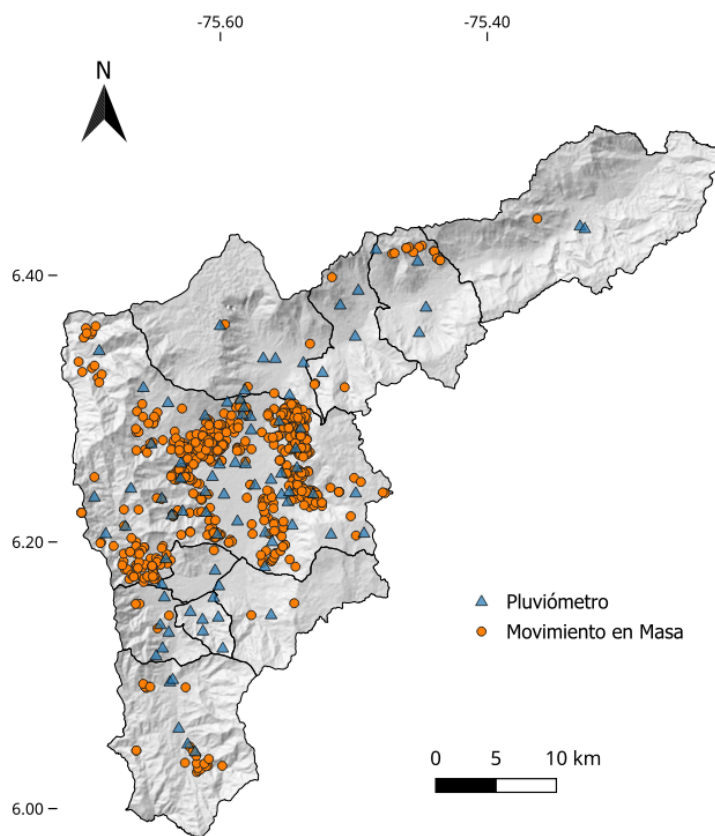


Figura 2. Mapa del Valle de Aburrá mostrando las ubicaciones de los eventos de los inventarios recopilados, con un total de 801 eventos. Los eventos del DAGRD (Medellín) abarcan el período de 2020 a 2023, y los de Geohazards de 2013 a 2023. Los 35 pluviómetros seleccionados del SIATA están marcados con triángulos azules.

Definición variables de precipitación: Los umbrales de lluvia Reciente - Antecedente, referidos en la literatura como umbrales de lluvia antecedente, están relacionados con los tipos de umbrales de precipitación que se derivan utilizando metodologías estadísticas y empíricas o modelo basados en datos (Balducci et al., 2006; Chleborad et al., 2006). Estos umbrales se basan en el acumulado total de precipitación registrado en los días previos a un movimiento en masa. Para establecer estos umbrales, el estudio se centró en el análisis de dos variables específicas de precipitación antecedente. La primera variable, nombrada precipitación antecedente a Corto Plazo (CP), refleja el acumulado de lluvia registrado en los días previos al evento. Por ejemplo, un CP de 7 días evalúa la precipitación acumulada en esos 7 días previos al evento. La segunda variable, llamada precipitación de Largo Plazo (LP), representa el acumulado de lluvia en los días anteriores al período considerado para CP. Por ejemplo, un LP de 90 días indicaría la precipitación acumulada durante los 90 días antes del período de 7 días de CP. La Figura 3 muestra un gráfico esquemático de una serie temporal de lluvia ilustrando estas dos variables.

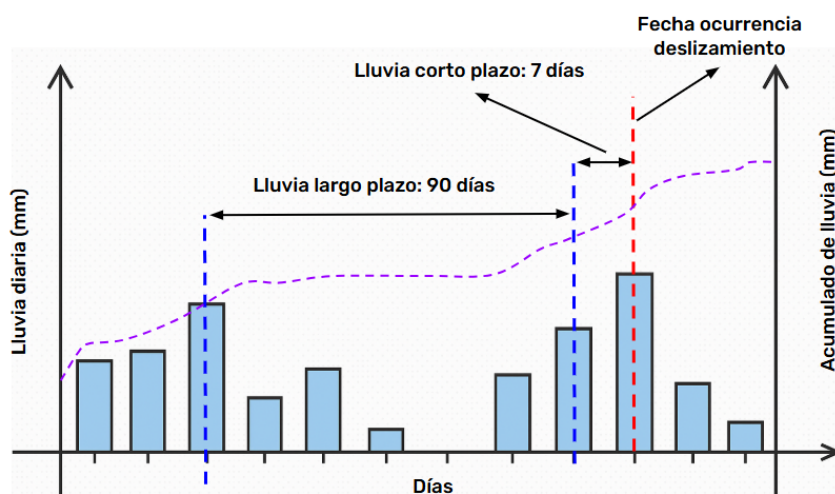


Figura 3. Esquema de una serie de tiempo de precipitación donde se indican las dos variables de precipitación antecedente que utiliza el modelo: Lluvia de Corto Plazo y Lluvia de Largo Plazo.

Se realizó el cálculo de la lluvia antecedente para diferentes intervalos temporales correspondientes a cada evento. Se realizaron variadas combinaciones de días para la CP y LP. Para CP, los períodos estudiados fueron de 1, 3, 5 y 7 días, y para LP, de 15, 30, 60 y 90 días. Combinando estos intervalos de CP y LP, se buscó determinar cuáles ofrecían mayor claridad para distinguir entre los patrones de precipitación que incrementan la probabilidad de movimientos en masa y aquellos que no.

2. Ajuste modelo y definición de umbrales

Entrenamiento del modelo y ajuste: Para la definición del modelo y ajuste se emplearon algunas técnicas de clasificación supervisada. Se exploraron tres algoritmos: Regresión Logística (RL), Máquina de Vectores de Soporte (SVM) y Análisis Discriminante Lineal (LDA), elegidos por su prevalencia en la literatura y su eficacia en clasificar linealmente entre grupos, diferenciando los eventos de lluvia que causan deslizamientos de los que no. Adicionalmente, se usaron 801 eventos de lluvia que no resultaron en movimientos en masa tomados aleatoriamente entre 2013 y 2023, que superaron los 5 mm de precipitación diaria y 100 mm en 90 días, para entrenar y evaluar los modelos, dividiendo los eventos en aquellos que resultaron y no resultaron en movimientos en masa.

Métricas de validación: Para cada combinación de días considerada, se evaluó el rendimiento de los tres algoritmos de clasificación anteriormente mencionados. Con el objetivo de identificar cuál combinación de días y qué algoritmo mostraban un desempeño superior en la clasificación de eventos de lluvia que generan movimientos en masa frente a los que no, se calcularon diversas métricas de rendimiento. Estas incluyen Accuracy, F1-Score, la curva ROC y el Área Bajo la Curva ROC (AUC). Para el entrenamiento y evaluación de cada algoritmo y combinación de días, se utilizó una división del 80 % de los datos para entrenamiento y el 20 % restante para evaluación.

Definición de umbrales y categorías de alertas: Los umbrales definidos se basaron en las métricas obtenidas de la Curva ROC. Esta curva es un gráfico cartesiano bidimensional que describe el rendimiento y la eficacia de un modelo, mostrando la tasa de verdaderos positivos (eje y: TPR) contra la tasa de falsos positivos (eje x: 1-TNR) para diferentes escenarios de probabilidad. En este contexto, el TPR, también conocido como Sensibilidad, mide la proporción de casos de deslizamiento correctamente identificados (es decir, deslizamientos que superaron el umbral) entre todos los eventos de lluvia con deslizamientos observados. Por ejemplo, un TPR del 95 % indica que el modelo identifica correctamente el 95 % de las lluvias que provocaron deslizamientos. Por otro lado, el TNR evalúa la capacidad del modelo para identificar correctamente los días de lluvia sin deslizamientos, es decir, la proporción de días lluviosos correctamente clasificados sin deslizamientos (que no superan el umbral) entre todos los eventos de lluvia observados sin deslizamientos.

Considerando lo anterior, un umbral óptimo se puede definir cuando tanto el TPR como el TNR son altos. Por esta razón, un TPR del 95 % puede interpretarse como un umbral con una probabilidad de excedencia del 5 %, un criterio común en la literatura que sugiere que el 95 % de las lluvias asociadas con deslizamientos superan el umbral definido, mientras que solo el 5 % no lo hace. En este análisis, se extrajeron tres umbrales de probabilidad. El primer umbral se asocia con un TPR del 95 %, el segundo corresponde al punto óptimo en la Curva ROC, que es el punto más cercano a la clasificación perfecta, es decir, un TPR y TNR del 100 % (esquina superior izquierda de la gráfica ROC). El tercer umbral se define con una probabilidad asociada a un TNR del 95 %, es decir, la más alta proporción de días sin deslizamientos correctamente clasificados.

En resumen, el **umbral alto** corresponde a una tasa de verdaderos positivos del 95 % (es decir, un 5 % de falsos negativos), mientras que el **umbral bajo** se asocia a una tasa de verdaderos negativos del 95 % (equivalente a un 5 % de falsos positivos). Con base en estos valores, se definieron tres categorías de probabilidad de ocurrencia de movimientos en masa: **Baja**, para valores inferiores al umbral bajo; **Media**, para valores comprendidos entre el umbral bajo y el umbral alto; y **Alta**, para valores superiores al umbral alto. En la Figura 4 se presenta un ejemplo gráfico de cómo se definieron los umbrales y cada una de las categorías de probabilidad.

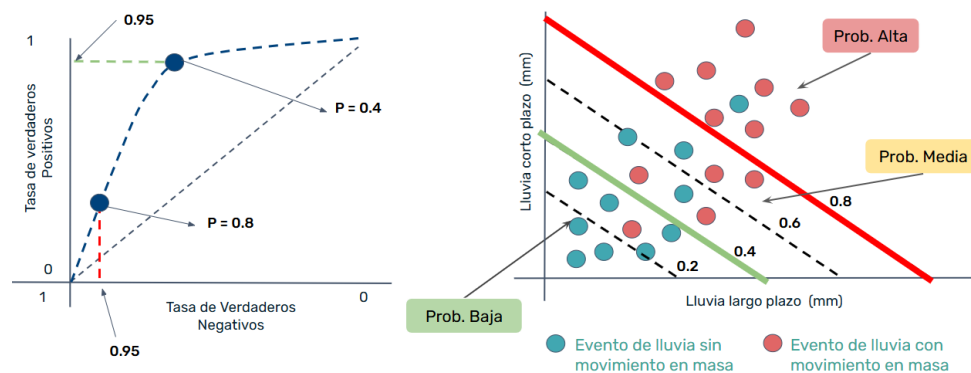


Figura 4. Ejemplo de la definición de los umbrales de lluvia a partir de la tasa de verdaderos positivos (95 %) y la tasa de falsos negativos (5 %). Se ilustran las categorías de probabilidad de ocurrencia de movimientos en masa: Baja, Media y Alta.

3. Especialización de la precipitación acumulada

Para evaluar los umbrales definidos en todo el Valle de Aburrá es necesario que las dos variables de entrada —la lluvia de corto plazo (CP) y la lluvia de largo plazo (LP)— estén representadas espacialmente en toda el área de estudio. Para ello, el modelo se apoya en dos fuentes de información de precipitación registradas por el proyecto SIATA: la red de pluviómetros y el radar meteorológico.

Cada fuente presenta ventajas y limitaciones. La red de pluviómetros ofrece mediciones puntuales de alta confiabilidad, pero no representa de manera continua la variabilidad espacial de la lluvia en toda la región. Por su parte, el radar permite caracterizar los patrones espaciales de la precipitación, aunque el cálculo del acumulado puede presentar mayor incertidumbre, dado que se trata de una medición indirecta realizada a una altura específica.

Con el fin de aprovechar lo mejor de ambas fuentes, se empleó una técnica de interpolación que combina la precisión de los acumulados medidos por la red de pluviómetros con la capacidad espacial del radar. La técnica utilizada corresponde al **Kriging Simple con Medias Locales (KSML)**.

El Kriging Simple con Medias Locales (SKLM) es una variante del Kriging Simple que incorpora una media local variable en lugar de una media constante. En este caso, la media local se estima utilizando datos auxiliares, como los datos de radar, que proporcionan una mejor representación de la variabilidad espacial en la región de estudio. Permite que la media local cambie de un punto a otro, utilizando datos de radar para capturar mejor los patrones espaciales.

La estimación de la precipitación en un punto no muestreado s_0 se realiza combinando la media local en esa ubicación con una suma ponderada de las diferencias entre los valores observados en los pluviómetros y la media local en las ubicaciones de esos pluviómetros. La fórmula es:

$$Z^*(s_0) = m(s_0) + \sum_{i=1}^N \lambda_i [Z(s_i) - m(s_i)], \quad (1)$$

donde:

- $Z^*(s_0)$ es la estimación de la precipitación en el punto no muestreado s_0 ,

- $m(s_0)$ es la media local en la posición s_0 , derivada de los datos de radar,
- $Z(s_i)$ es el valor de precipitación medido en el pluviómetro ubicado en s_i ,
- $m(s_i)$ es la media local en el punto s_i , también obtenida del radar,
- λ_i es el peso asignado a cada observación en el pluviómetro s_i , que se determina minimizando la varianza de la estimación, es decir, buscando la mayor precisión posible.

Para validar su eficacia, se seleccionaron fechas aleatorias con eventos de lluvia en cada uno de los trimestres del año. Esto permitió evaluar el comportamiento del modelo de interpolación tanto en temporadas con alta incidencia de lluvias como en aquellas con menor actividad o días con lluvia. En la Figura 5 se presenta un ejemplo de evento de lluvia específico registrado con la red pluviométrica y con el radar meteorológico y cómo sería el mapa final interpolado.

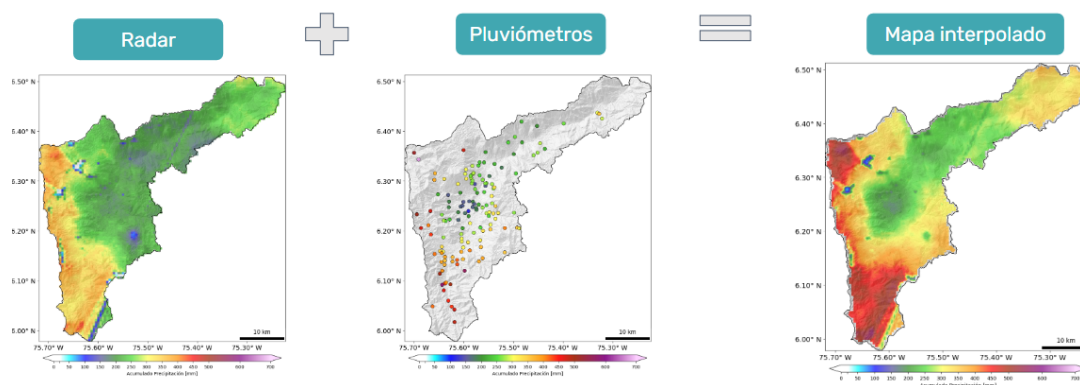


Figura 5. Ejemplo de evento de lluvia específico registrado con la red pluviométrica y con el radar meteorológico y cómo sería el mapa final interpolado.

4. Mapa de alertas

Finalmente, se presenta el mapa de alertas con las categorías previamente definidas. En términos generales, el modelo integra el mapa de lluvia acumulada de corto plazo con el de largo plazo, y evalúa en cada píxel de la región si estos valores superan los umbrales establecidos. De esta manera, se obtiene un mapa de probabilidad para todo el Valle de Aburrá, como se observa en la Figura 6.

Posteriormente, este mapa de probabilidad se cruza con una máscara que excluye las grandes áreas con pendientes menores a 5° , que corresponden principalmente a zonas planas asociadas a la llanura de inundación del río Aburrá y a algunos valles de quebradas mayores. En estas áreas, la probabilidad de que se presenten movimientos en masa detonados por lluvia es prácticamente nula, por lo que se clasifican en la categoría **No aplica**.

Luego, el mapa resultante se intersecta con la capa vectorial de barrios y veredas del Valle de Aburrá. La categoría de probabilidad asignada a cada barrio o vereda corresponde a la categoría más frecuente entre los píxeles contenidos en su polígono. Por ejemplo, si en la vereda Piedras Blancas la mayoría de los píxeles presentan una probabilidad baja, entonces la vereda se clasifica con probabilidad baja.

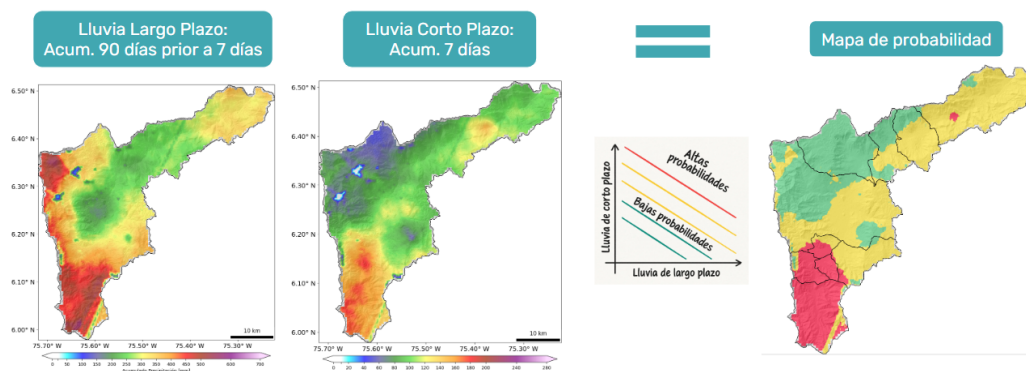


Figura 6. Mapas de las variables de lluvia de corto plazo y largo plazo, evaluadas en los umbrales establecidos para obtener las diferentes categorías de probabilidad de ocurrencia de movimientos en masa.

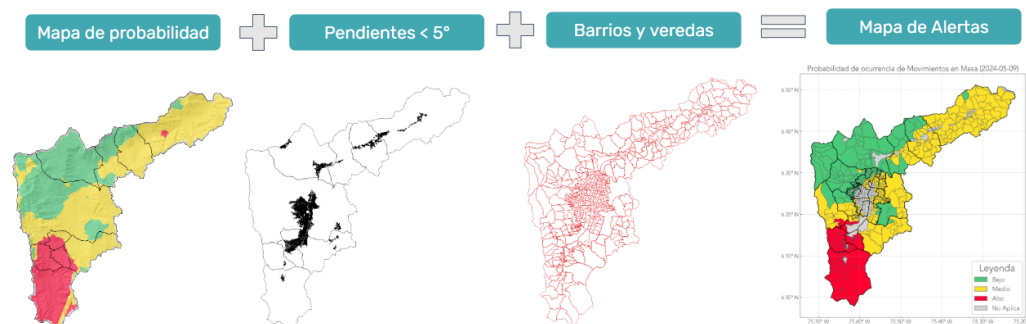


Figura 7. Mapa final de alertas por probabilidad de ocurrencia de movimientos en masa detonados por lluvias, agregado por veredas y barrios de los distintos municipios del Valle de Aburrá.

Resultados

Los resultados de las métricas de Accuracy y F1-Score, expresados en porcentajes, se presentan en la Tabla 1 para cada combinación de días y algoritmo de clasificación evaluados. Es notorio que los tres modelos de clasificación presentaron un rendimiento óptimo. Todas las combinaciones de días exhibieron un Accuracy y un F1-Score superiores al 70 %, lo cual sugiere que casi cualquier selección de días logra distinguir eficazmente entre las lluvias que incrementan la probabilidad de ocasionar movimientos en masa y aquellas que no. Las combinaciones más efectivas, de acuerdo con las métricas evaluadas, son las de 1 día versus 90 días, 3 días versus 90 días, 5 días versus 90 días y 7 días versus 90 días, destacadas en negrita en la Tabla 3.

Tabla 1. Evaluación de algoritmos de clasificación con Accuracy y F1-Score para todas las combinaciones de días con lluvia registrada por pluviómetros.

Pluvios CP vs. LP	Regresión Logística		SVM		LDA	
	Accuracy	F1-score	Accuracy	F1-score	Accuracy	F1-score
1dvs.15d	83.49	83.49	85.36	85.36	80.06	80.06
1dvs.30d	86.29	86.29	88.79	88.79	83.18	83.18
1dvs.60d	88.47	88.47	87.85	87.85	83.49	83.49
1dvs.90d	89.72	89.72	89.1	89.1	88.16	88.16
3dvs.15d	77.26	77.26	77.26	77.26	76.95	76.95
3dvs.30d	77.88	77.88	77.88	77.88	74.77	74.77
3dvs.60d	79.13	79.13	79.44	79.44	80.37	80.37
3dvs.90d	81.31	81.31	83.49	83.49	79.75	79.75
5dvs.15d	72.9	72.9	73.83	73.83	73.52	73.52
5dvs.30d	76.32	76.32	75.39	75.39	74.45	74.45
5dvs.60d	76.95	76.95	80.06	80.06	76.95	76.95
5dvs.90d	80.69	80.69	80.37	80.37	80.37	80.37
7dvs.15d	73.52	73.52	73.83	73.83	74.77	74.77
7dvs.30d	78.19	78.19	77.57	77.57	78.19	78.19
7dvs.60d	78.82	78.82	79.13	79.13	79.13	79.13
7dvs.90d	81.31	81.31	81.93	81.93	81.31	81.31

Finalmente, de las combinaciones más efectivas (1 día vs. 90, 3 días vs. 90 días, 5 días vs. 90 días y 7 días vs. 90 días) se escogieron 3 combinaciones de lluvias para la evaluación en tiempo real en la plataforma SIAMMO, estos fueron la combinación de 1 días vs. 90 días, 3 días vs. 90 días y las de 7 días vs. 90 días, esto por qué las primeras combinaciones refleja un poco las lluvias diaria o detonante, mientras que la segunda se asocia más a la lluvia antecedente que posiblemente aporta a la humedad en el suelo días anteriores a un evento de lluvia detonante.

En las Figuras 8, 9 y 10 se presentan los resultados de cada modelo evaluado. El gráfico de la izquierda representa la combinación de 90 días prior a los 7 días contra los 7 días y sus respectivos valores de probabilidades de lluvia crítica, presentados con diferentes gamas de colores. También se muestran los eventos de lluvia que detonaron movimientos en masa, así como aquellos que no lo hicieron y que formaron parte del testeo o evaluación del modelo, junto con los tres umbrales definidos por la Curva ROC, que se muestran en el segundo gráfico a la derecha, donde se indican con puntos cada uno de los valores de los umbrales. Se puede observar que los tres modelos presentaron un comportamiento similar tanto en la curva ROC como en los umbrales definidos. Por esta razón, para la combinación de días de 1 día contra 90 días anteriores a un día, solo se mostró el resultado obtenido con el modelo de Regresión Logística, como se observa en la Figura 11.

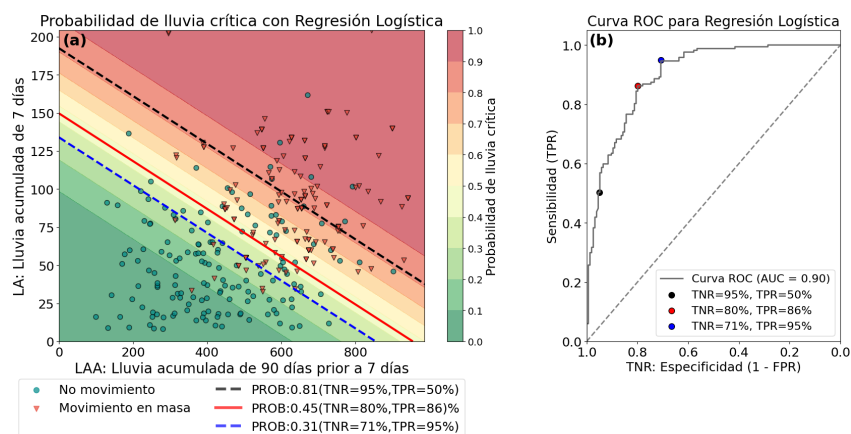


Figura 8. (a) Combinación de CP de 7 días y LP de 90 días prior a 7 días y su respectiva probabilidad hallada con Regresión Logística. Las líneas negra, roja y azul representan los umbrales determinados con la Curva ROC mostrados en forma de puntos en el gráfico (b). También se muestra los datos de validación utilizados para construir la curva ROC (puntos verdes y triángulos rojos). (b) Curva ROC para la combinación de CP de 7 días y LP de 90 días prior a 7 días con Regresión Logística. El punto azul representa el umbral hallado con un TPR de 95 %, el punto rojo es el punto Óptimo y el negro el TNR de 95 %.

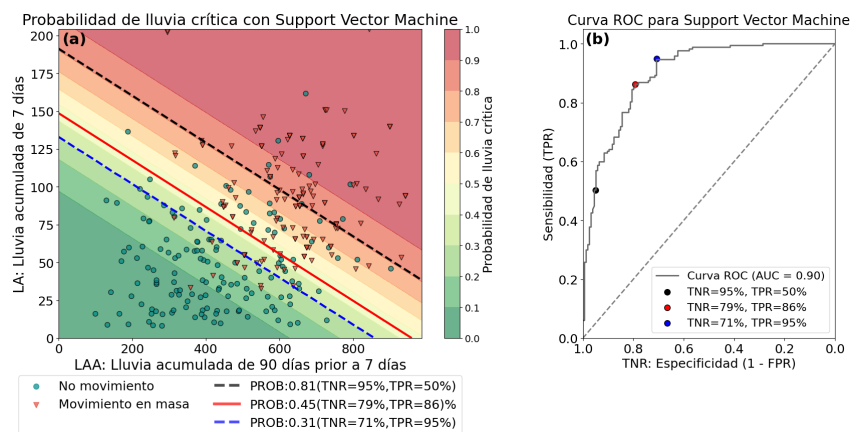


Figura 9. (a) Combinación de CP de 7 días y LP de 90 días prior a 7 días y su respectiva probabilidad hallada con Support Vector Machine. Las líneas negra, roja y azul representan los umbrales determinados con la Curva ROC mostrados en forma de puntos en el gráfico (b). También se muestra los datos de validación utilizados para construir la curva ROC (puntos verdes y triángulos rojos). (b) Curva ROC para la combinación de CP de 7 días y LP de 90 días prior a 7 días con Support Vector Machine. El punto azul representa el umbral hallado con un TPR de 95 %, el punto rojo es el punto Óptimo y el negro el TNR de 95 %.

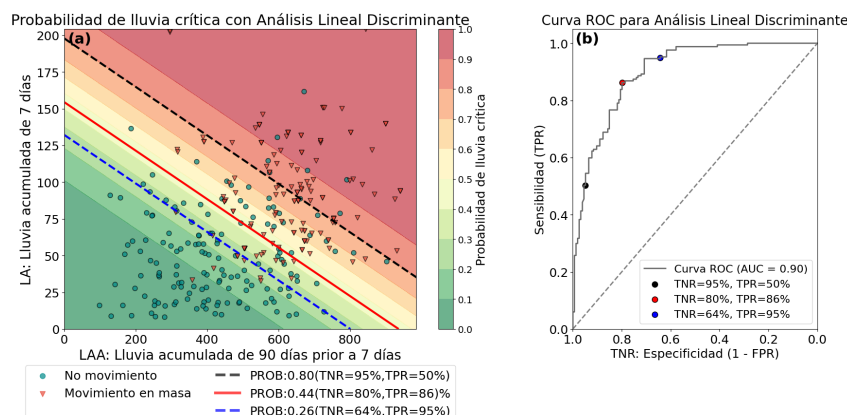


Figura 10. (a) Combinación de CP de 7 días y LP de 90 días prior a 7 días y su respectiva probabilidad hallada con Análisis Lineal Discriminante. Las líneas negra, roja y azul representan los umbrales determinados con la Curva ROC mostrados en forma de puntos en el gráfico (b). También se muestra los datos de validación utilizados para construir la curva ROC (puntos verdes y triángulos rojos). (b) Curva ROC para la combinación de CP de 7 días y LP de 90 días prior a 7 días con Análisis Lineal Discriminante. El punto azul representa el umbral hallado con un TPR de 95 %, el punto rojo es el punto Óptimo y el negro el TNR de 95 %.

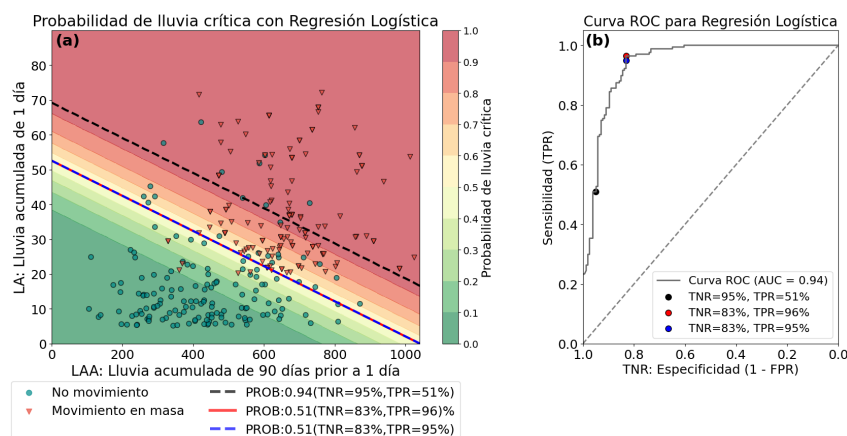


Figura 11. (a) Combinación de CP de 1 día y LP de 90 días prior a 1 día y su respectiva probabilidad hallada con Regresión Logística. Las líneas negra, roja y azul representan los umbrales determinados con la Curva ROC mostrados en forma de puntos en el gráfico (b). También se muestra los datos de validación utilizados para construir la curva ROC (puntos verdes y triángulos rojos). (b) Curva ROC para la combinación de CP de 1 día y LP de 90 días prior a 1 días con Regresión Logística. El punto azul representa el umbral hallado con un TPR de 95 %, el punto rojo es el punto Óptimo y el negro el TNR de 95 %.

Finalmente, se realiza una evaluación visual del desempeño del modelo mediante la comparación de los mapas generados para distintos eventos de lluvia que ocasionaron emergencias y afectaciones en varios sectores del Valle de Aburrá. Adicionalmente, se incluye un caso de lluvia en el que no se reportaron movimientos en masa, con el fin de contrastar los resultados.

En la Figura 12 se presentan, en la parte superior izquierda, los mapas de lluvia acumulada de corto plazo (1 día) y de largo plazo (90 días). A la derecha se observa el mapa de probabilidad correspondiente al 25 de mayo de 2023, fecha en la que se registró un número significativo de movimientos en masa de tipo flujo en la vereda San Andrés del municipio de Girardota. En este evento se contabilizaron alrededor de 350 deslizamientos que afectaron vías, viviendas y parcelaciones.

De manera similar, en la Figura 13 se muestra el resultado para el 9 de mayo de 2024, considerando una lluvia acumulada de corto plazo de 7 días y una acumulación de largo plazo de 90 días. En este caso, el modelo indicó una alta probabilidad de movimientos en masa en el sur del valle, lo que coincidió con la ocurrencia de un flujo en el sector Himalaya, en el municipio de La Estrella, que obligó a la evacuación de más de 100 personas.

En las Figuras 14 y 15 se ilustran otros dos eventos relevantes: el del 29 de abril de 2025 en el corregimiento de Altavista (Medellín) y el del 6 de mayo de 2025 en el sector Los Balsos (El Poblado). Ambos eventos estuvieron asociados a lluvias intensas en los días previos y generaron pérdidas materiales considerables.

Finalmente, en la Figura 16 se presenta el mapa correspondiente al 30 de enero de 2025, fecha en la que se registró un evento de lluvia sin reportes de movimientos en masa en el Valle de Aburrá. En este caso, el modelo asignó predominantemente la categoría de probabilidad **Baja**, lo cual concuerda con la ausencia de emergencias y representa el comportamiento esperado del sistema en este tipo de situaciones.

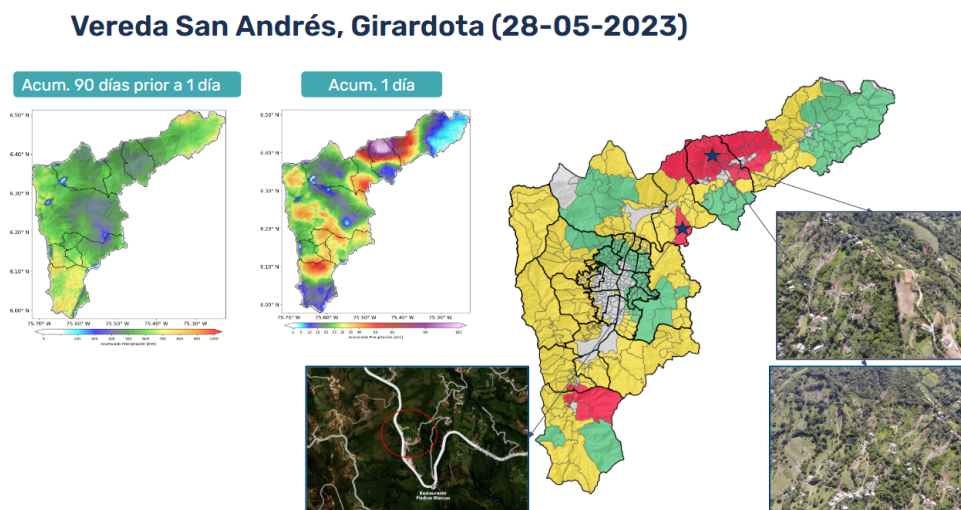


Figura 12. Mapas de lluvia acumulada de corto plazo (1 día), largo plazo (90 días) y mapa de probabilidad para el 25 de mayo de 2023, cuando se registraron alrededor de 350 movimientos en masa en la vereda San Andrés, Girardota.

El Himalaya, La Estrella (9-05-2024)

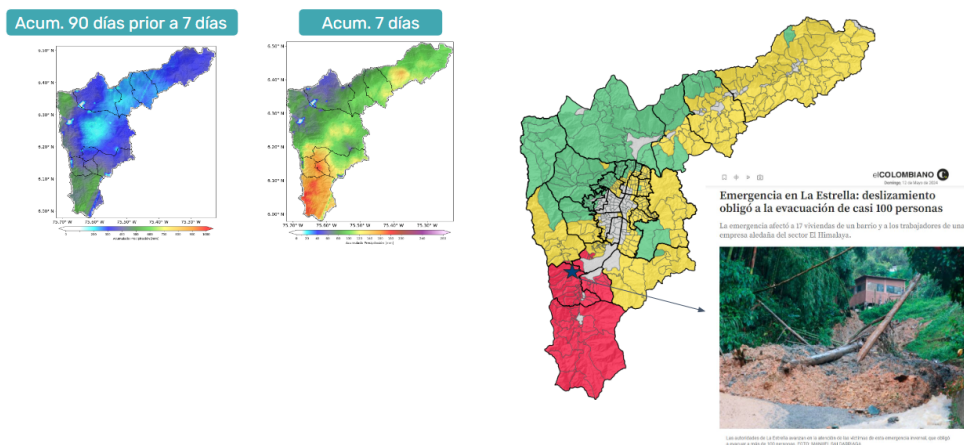


Figura 13. Mapas de lluvia acumulada y probabilidad para el 9 de mayo de 2024, día en que se presentó un flujo en el sector Himalaya (La Estrella), que obligó a la evacuación de más de 100 personas.

Altavista, Medellín (29-04-2025)

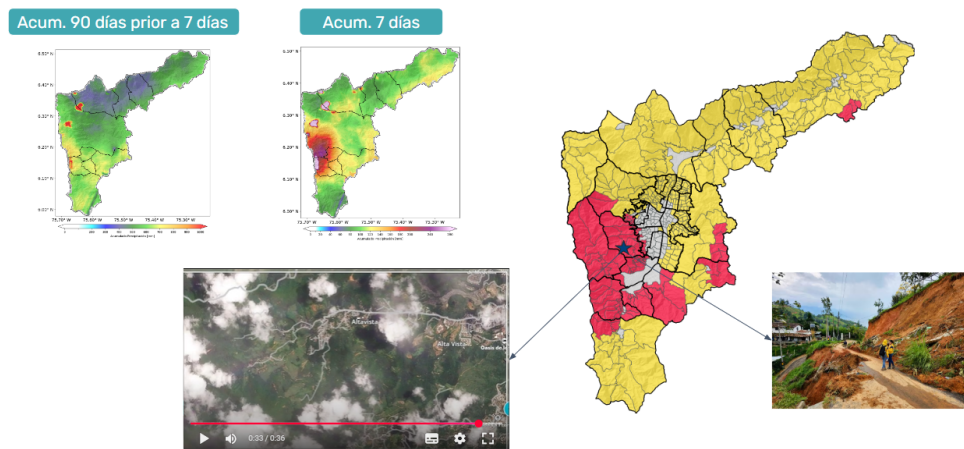


Figura 14. Mapas de probabilidad para el 29 de abril de 2025, eventos de deslizamientos registrados en el corregimiento de Altavista (Medellín).

Los Balsos, Medellín (06-05-2025)

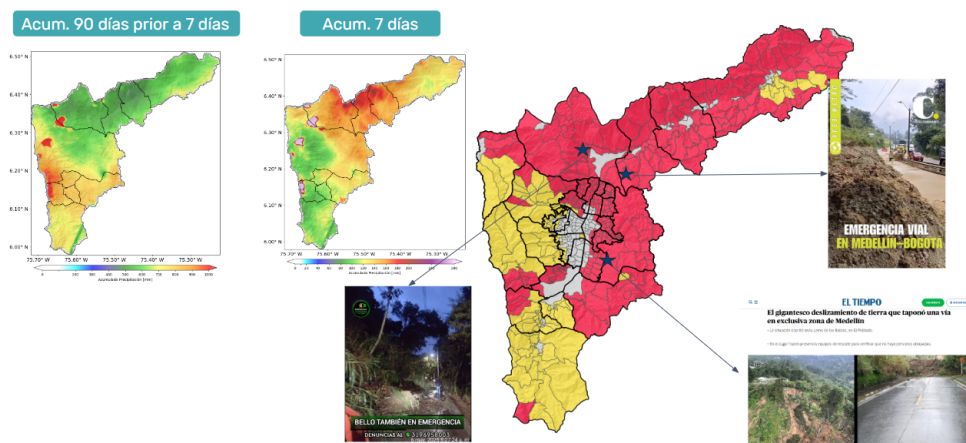


Figura 15. Mapas de probabilidad para el 6 de mayo de 2025, evento ocurrido en el sector Los Balsos (El Poblado, Medellín).

Evento de lluvia (30-01-2025)

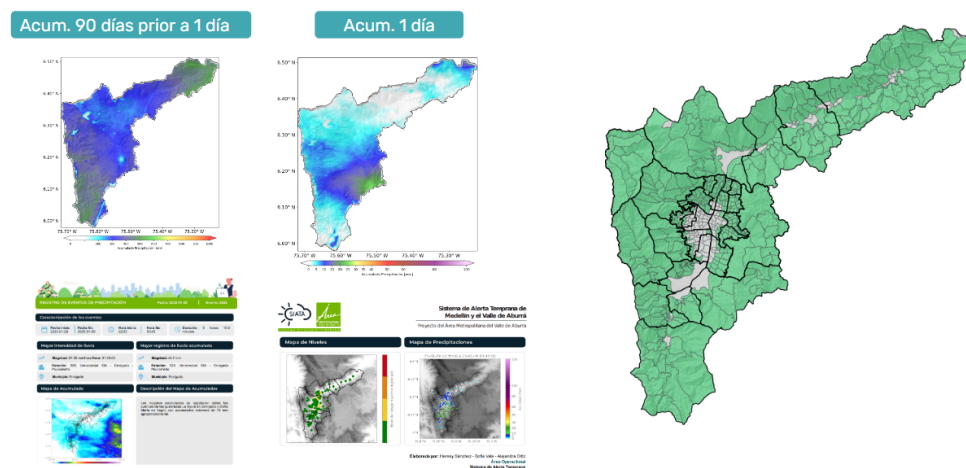


Figura 16. Mapa de probabilidad correspondiente al 30 de enero de 2025, evento de lluvia en el que no se registraron movimientos en masa en el Valle de Aburrá.

Discusión y conclusiones

Este estudio presenta una metodología para un modelo regional de alertas por movimientos en masa detonados por lluvia, basada en dos variables de precipitación antecedente: *corto plazo* (CP) y *largo plazo* (LP). Se entrenaron y evaluaron tres clasificadores supervisados (Regresión Logística, SVM y LDA) con un conjunto de eventos con deslizamientos (801) y un conjunto equivalente de eventos sin deslizamientos seleccionados aleatoriamente. Los resultados muestran desempeños consistentes (Accuracy y F1-score $> 70\%$) y destacan como combinaciones más informativas aquellas que usan LP de 90 días junto con CP

de 1, 3, 5 y 7 días. Con base en ello, se seleccionaron tres configuraciones (1–90, 3–90 y 7–90 días) para operación en tiempo casi real.

Operacionalmente, la metodología integra: (i) umbrales derivados de la Curva ROC (umbral alto con TPR = 95 %, umbral bajo con TNR = 95 %), (ii) fusión y espacialización de la precipitación mediante Kriging Simple con Medias Locales (KSML), combinando la precisión puntual de los pluviómetros con el patrón espacial del radar, y (iii) agregación de probabilidades por unidades administrativas (barrios y veredas) tras enmascarar zonas de pendiente $< 5^\circ$ (*No aplica*). Estudios de caso recientes evidencian coherencia espacial entre las áreas clasificadas con probabilidad *Alta* y la ocurrencia de afectaciones reportadas.

Pese a los resultados alentadores, deben considerarse varias fuentes de incertidumbre: (i) sesgos y vacíos del inventario (concentración espacial de reportes, precisión variable en fecha/hora y localización de eventos), (ii) incertidumbres instrumentales y operativas de la red de pluviómetros (fallos, mantenimiento, entorno local) y del radar (calibración, atenuación, conversión a acumulados) y (iii) la suposición de fronteras aproximadamente lineales en los clasificadores aplicados. Además, los umbrales son estáticos en el tiempo y no capturan explícitamente estacionalidad, variabilidad intra-anual o condicionantes climáticos de gran escala.

Recomendaciones y trabajo futuro

Se proponen las siguientes líneas de mejora para robustecer el sistema:

- **Evaluación y calibración probabilística:** complementar la ROC/AUC con curvas precisión–recuperación y diagramas de fiabilidad.
- **Validación espaciotemporal rigurosa:** esquemas de validación cruzada por estación/sector y retroevaluaciones por temporadas, para medir generalización fuera de muestra.
- **Enriquecimiento de predictores:** explorar otras ventanas CP/LP y variables auxiliares (índices topográficos, litología, cobertura/uso del suelo, proxies de humedad antecedente), y métodos que capturen no linealidades (bosques aleatorios, *gradient boosting*) con control de sobreajuste.
- **Umbrales dinámicos y regionalización:** permitir que los umbrales varíen por subregiones hidroclimáticas o clusters de lluvia, y por estado climático (p. ej., temporadas o fases ENSO).
- **Fusión radar–pluviómetros:** evaluar variantes del KSML y esquemas de corrección por sesgo/escala del radar, cuantificando la incertidumbre de la interpolación (mapas de varianza kriging).
- **Métricas operativas:** incorporar indicadores orientados a alerta (POD, FAR, CSI) y análisis de costos/beneficios de falsas alarmas vs. omisiones, alineados con las necesidades de gestión del riesgo.

La estrategia propuesta genera mapas de probabilidad interpretables y operables para apoyo a la toma de decisiones a escala regional. Si bien persisten incertidumbres asociadas a datos, modelos y agregación espacial, los resultados indican capacidad para discriminar condiciones de *Alta* vs. *Baja* probabilidad en concordancia con eventos observados. La implementación de las mejoras sugeridas —en especial calibración probabilística, validación espaciotemporal y umbrales dinámicos— fortalecerá la confiabilidad del sistema y su utilidad práctica para la gestión del riesgo por movimientos en masa detonados por lluvia en el Valle de Aburrá.

Productos

Poner los mapas, siammo y el repositorio.

Tabla 2. Productos relacionados con análisis de lluvia y movimientos en masa.

Producto / Tema	Descripción resumida
Repositorio Código Git-Hub	https://github.com/siatalandslides/alert_model
Mapas diarios	https://siata.gov.co/geotecnia/alertas_diarias/
Mapas en la plataforma SIAMMO	https://siata.gov.co/geotecnia/siammo.html

Avances Mensuales

Mes de Julio

Durante el mes de julio se enfocaron esfuerzos en mejorar los mapas de lluvia acumulada interpolados a partir de la red pluviométrica y del radar meteorológico. Se identificó que el principal problema provenía de la información del radar en zonas altas, como el Cerro Padre Amaya o el Cerro La Baldía, donde el haz del radar impacta el relieve, generando valores artificialmente altos de precipitación.

En conjunto con el equipo de meteorología, se evaluaron diferentes estrategias para corregir este sesgo. Finalmente, se implementó una máscara que excluye estos valores no representativos, evitando que distorsionen la interpolación. Las pruebas realizadas con diferentes metodologías concluyeron que la opción más efectiva fue dejar estas áreas como “sin datos”, permitiendo que el proceso de interpolación las completara usando la información de las celdas vecinas.

A partir de esta mejora, los mapas de interpolación de lluvia dejaron de verse afectados por valores extremos irreales, y los mapas de probabilidad resultantes mostraron coherencia con las condiciones reales. Se espera seguir avanzando en esta línea, no solo filtrando la información del radar con mayor precisión, sino también adoptando técnicas de interpolación que consideren el error en los datos, de forma que estos sean filtrados o eliminados en lugar de amplificados, garantizando así productos de lluvia más fiables.

Mes de agosto

Máscara datos acumulados de Radar Durante el mes de julio se aplicó una máscara que eliminaba (dejaba “sin datos”) las celdas de acumulados de lluvia muy altos del radar—típicos cuando el haz impacta en zonas montañosas muy altas—. Esto resolvía los picos irreales, pero dejaba huecos que luego distorsionaban la interpolación (Lluvia pluvios + Radar). En agosto se mejoró el procesamiento con una estrategia robusta que se aplica en dos pasos: (i) **detección** de picos altos usando estadística robusta local y (ii) **relleno** (*inpainting*) de las celdas marcadas como picos altos con el valor del píxel válido más cercano. El objetivo es obtener campos acumulados realistas sin suavizar en exceso la lluvia acumulada real.

El procesamiento se puede dividir en los siguientes pasos:

- **Acumulado temporal:** Sea $R(x, y, t)$ la lluvia horaria de radar. Para un periodo $[t_0 - D, t_0]$, el

acumulado es:

$$A(x, y) = \sum_{t \in [t_0 - D, t_0]} R(x, y, t), \quad (2)$$

donde t_0 es la fecha objetivo y D los días previos a acumular (p.ej., $D = 90$).

- **Detección robusta de picos (MAD local):** Se recorre la malla $A(x, y)$, para cada celda i , se evalúa A_i respecto a su vecindario W_i (vecindario cuadrado de tamaño $s \times s$ centrado en i , aquí $s = 9$):

$$m_i = \text{med}\{A_j : j \in W_i\}, \quad (3)$$

$$\text{MAD}_i = \text{med}\{|A_j - m_i| : j \in W_i\}, \quad (4)$$

$$\sigma_i = 1,4826 \text{ MAD}_i. \quad (5)$$

Aquí, m_i es la mediana local, MAD_i mide la dispersión robusta y σ_i convierte MAD a una escala comparable con la desviación estándar bajo normalidad. Se marca como *pico alto* la celda i si

$$(A_i - m_i) > k \sigma_i \quad \text{y} \quad \sigma_i > \varepsilon, \quad (6)$$

donde k controla cuán estricta es la detección y ε evita activar la máscara en zonas prácticamente uniformes. **Parámetros usados:** $s = 9$ (aprox. $4,5 \times 4,5$ km si la celda es de 500 m), $k = 3,5$, $\varepsilon = 0$.¹

- **Máscara y relleno por vecino más cercano:** La detección anterior produce una máscara booleana M sobre todo el Valle de Aburrá, donde $M(p) = 1$ (verdadero) indica “pico alto” y $M(q) = 0$ indica celda válida. Para cada celda marcada $p=1$ se busca el píxel válido más cercano:

$$q^* = \arg \min_{q: M(q)=0} \|p - q\|_2,$$

y se define el campo corregido

$$\hat{A}(p) = A(q^*), \quad \hat{A}(q) = A(q) \text{ si } M(q) = 0.$$

Esta estrategia preserva bordes y acumulados reales al no promediar sobre áreas extensas.

Elección de parámetros y consideraciones

- **Ventana s :** ventanas mayores son más robustas al ruido, pero pueden pasar por alto picos pequeños; $s = 9$ equilibra robustez y sensibilidad a la escala de 500 m.
- **Umbral k :** valores típicos $k \in [3, 5]$. Aquí $k = 3,5$ captura ecos orográficos intensos sin afectar núcleos convectivos reales.
- **Tolerancia ε :** útil para evitar disparos en campos muy planos (baja variabilidad).
- **NaN y bordes:** si todo el vecindario es inválido, no se rellena; opcionalmente puede extenderse la máscara a celdas NaN.

Resultados Tras aplicar la corrección, los mapas de acumulado de D días muestran una reducción clara de columnas y píxeles extremos alineados con los picos montañosos, donde la señal del radar no es

¹En operación, $\varepsilon \in [0, 1, 0, 5]$ mm ayuda a reducir falsos positivos en campos muy homogéneos.

clara, manteniendo el acumulado de lluvia real. La fracción de celdas corregidas se mantiene baja (como es esperable para artefactos localizados) y los máximos regionales descienden a rangos físicamente reales para las fechas de acumulados ensayadas.

En las Figuras 17, 18 y 19 se presentan, respectivamente: (i) el mapa de acumulado sin posprocesamiento, (ii) la máscara que identifica los píxeles no realistas y (iii) el acumulado final corregido con la metodología. En la Figura 17 se observan valores anómalamente altos tanto en el acumulado de 7 días como en el acumulado de 90 días previos; la Figura 19 muestra la máscara utilizada, donde se destacan las celdas que no corresponden a valores reales (picos muy altos); y la Figura 18 evidencia que, tras el procesamiento descrito, los mapas de acumulado son más acordes con la realidad, lo cual también se refleja en la escala de valores.

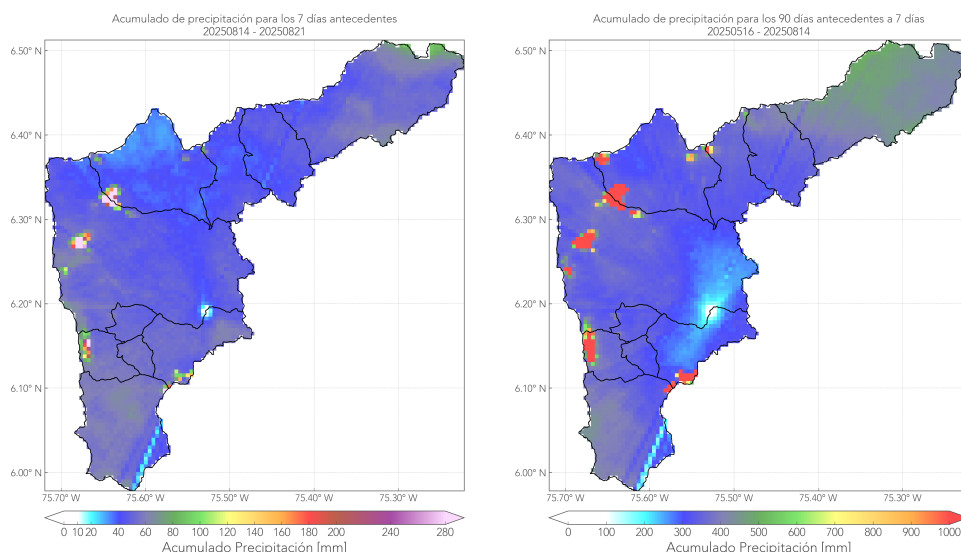


Figura 17. Mapa de acumulado de radar para los 7 días anteriores al día 21 de agosto de 2025 y los 90 días anteriores a esos 7 días. Se observan en ambos mapas valores extremos de lluvia superiores a los 1000 mm en ciertas zonas que coinciden con puntos donde el haz del radar choca con partes altas de montaña, principalmente hacia el occidente.

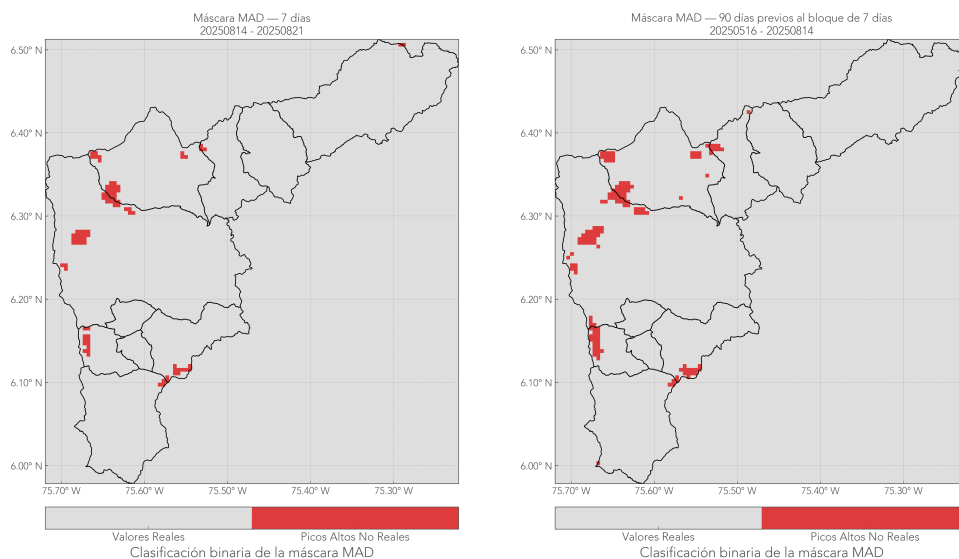


Figura 18. Máscara de las zonas identificadas como valores muy altos, fuera de lo normal, coherente con puntos donde el haz del radar choca con partes altas de montaña, principalmente hacia el occidente.

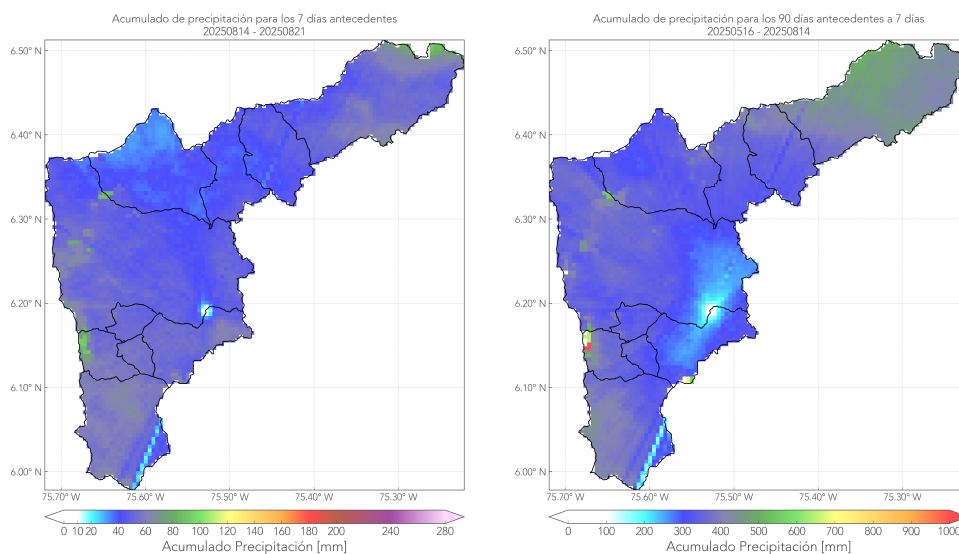


Figura 19. Mapa de acumulado de radar corregido con la metodología del vecino más cercano. Se observan que los picos muy altos ya no están.

Fragmentos de implementación (Python) Máscara de picos (MAD local):

```
def mascara_spike_mad(da, size=9, k=3.5, min_mad=0.0):
    arr = da.values
    def _med(x): return np.nanmedian(x)
    def _mad(x):
```



```
m = np.nanmedian(x)
return np.nanmedian(np.abs(x - m))
med = ndi.generic_filter(arr, _med, size=size, mode="nearest")
mad = ndi.generic_filter(arr, _mad, size=size, mode="nearest")
sigma = 1.4826 * mad
valid = np.isfinite(arr) & np.isfinite(med) & np.isfinite(sigma) & (sigma > min_mad)
mask = valid & ((arr - med) > (k * sigma))
return xr.DataArray(mask, coords=da.coords, dims=da.dims, name="mask_spike_mad")
```

Relleno por vecino más cercano:

```
def rellenar_vecino_mas_cercano(da, bad_mask):
    arr = da.values.copy()
    mask = bad_mask.values # | ~np.isfinite(arr) (activar si se desea)
    if mask.all():
        return da.copy()
    inds = ndi.distance_transform_edt(mask, return_distances=False, return_indices=True)
    out = arr.copy()
    out[mask] = arr[inds[0][mask], inds[1][mask]]
    return xr.DataArray(out, coords=da.coords, dims=da.dims, attrs=da.attrs, name=da.name)
```

Uso en la cadena de acumulación:

```
da_perodo, da_90d = acumulado.generar_mapa_acumulado()
mask_mala = mascara_spike_mad(da_perodo, size=9, k=3.5, min_mad=0.0)
da_perodo_fix = rellenar_vecino_mas_cercano(da_perodo, mask_mala)
```

Cálculo de la incertidumbre del Modelo de Alertas Uno de los objetivos principales de esta vigencia es realizar una evaluación detallada de la **incertidumbre del modelo de alertas**. Para ello es necesario plantear una metodología que permita abordar de manera integral los diferentes componentes del modelo y, en lo posible, cuantificar dicha incertidumbre.

Durante el mes se avanzó en la revisión y recopilación de estudios que proponen metodologías para evaluar la incertidumbre y validar sistemas de alerta temprana por movimientos en masa en distintas regiones del mundo. Estos trabajos constituyen una base conceptual y metodológica importante para orientar el proceso local. Entre los estudios más relevantes se destacan:

- *Assessing the performance of regional landslide early warning models: the EDuMaP method*. Revista *Natural Hazards and Earth System Sciences*, 2016 (Calvellido and Piciullo, 2016).
- *Adapting the EDuMaP method to test the performance of the Norwegian early warning system for weather-induced landslides*. Revista *Natural Hazards and Earth System Sciences*, 2017 (Piciullo et al., 2017).
- *Critical rainfall thresholds for triggering shallow landslides in the Serchio River Valley (Tuscany, Italy)*. Revista *Natural Hazards and Earth System Sciences*, 2012 (Giannecchini et al., 2012).
- *Calibration and validation of rainfall thresholds for shallow landslide forecasting in Sicily, southern Italy*. Revista *Geomorphology*, 2015 (Gariano et al., 2015).

- *Rainfall thresholds for the forecasting of landslide occurrence at regional scale.* Revista *Landslides*, 2012 (Martelloni et al., 2012).
- *Territorial early warning systems for rainfall-induced landslides.* Revista *Earth-Sciences Reviews*, 2018 (Piciullo et al., 2018).

Estos estudios plantean diferentes enfoques para la calibración, validación y análisis de incertidumbre en modelos de alerta por deslizamientos, lo que permitirá seleccionar e implementar las metodologías más adecuadas para el contexto del Valle de Aburrá.

Mes de septiembre

Durante agosto el trabajo se orientó al procesamiento de los acumulados de lluvia obtenidos por el radar meteorológico, incorporando una metodología que mitiga varias de sus limitaciones.

En septiembre el esfuerzo se centró en mejorar y actualizar la sección de *Movimientos en masa* del reporte hidrometeorológico semanal. Hasta ahora, el reporte evaluaba los umbrales únicamente con información de estaciones pluviométricas puntuales; por lo tanto, no se realizaba una espacialización a nivel de barrios y veredas de los municipios del Valle de Aburrá. En la Figura 20 se ilustra el formato anterior del reporte, en el que la probabilidad de movimientos en masa se presentaba sólo por estación.

Para contextualizar, el reporte se divide en dos partes: (i) **Resumen semanal** y (ii) **Condiciones actuales**. En el formato previo, el *Resumen semanal* mostraba un gráfico de barras con el número de estaciones en probabilidad alta (rojo), media (amarillo) y baja (verde) para cada día de la semana anterior (lunes a domingo; ver Figura 20). A la derecha, en *Condiciones actuales*, se incluía un mapa con la ubicación de las estaciones del valle coloreadas según la probabilidad del día de emisión del reporte (los lunes), acompañado de una tabla con los conteos por nivel de probabilidad y los municipios con estaciones en probabilidad alta (ver Figura 20). Finalmente, en la parte inferior se incluía una explicación breve y general del modelo de alertas.

Durante septiembre se actualizó este reporte para ofrecer alertas y un resumen semanal **especializados a nivel de barrio y vereda**, y no sólo por estaciones. Se diseñó un nuevo formato (Figura 21) que preserva la estructura original (izquierda: resumen semanal; derecha: condiciones actuales) pero incorpora:

- **Resumen semanal (izquierda):** gráfico de barras apiladas por municipio, que resume la proporción de barrios y veredas en probabilidades alta, media y baja para cada día de la última semana. Además, se listan los barrios/veredas que acumulan tres o más días en probabilidad alta (rojo), como insumo para priorizar monitoreo continuo, emisión de reportes y, de ser necesario, inspecciones de campo.
- **Condiciones actuales (derecha):** mapa de probabilidad de movimientos en masa para cada barrio o vereda de los municipios del Valle de Aburrá, correspondiente a la fecha de emisión del reporte (los lunes), junto con una tabla resumen del número de barrios/veredas en cada nivel de probabilidad y el listado de municipios asociados.

Asimismo, se actualizó la explicación metodológica del modelo en la sección inferior del reporte, aclarando de manera concisa en qué consiste, cómo se interpretan los niveles de probabilidad (alta/rojo, media/amarillo, baja/verde) y cómo se integra la información de lluvia antecedente.

Con esta nueva versión se espera mejorar la comprensibilidad del modelo de alertas por movimientos en masa detonados por lluvia y aumentar la utilidad de la información para la toma de decisiones.

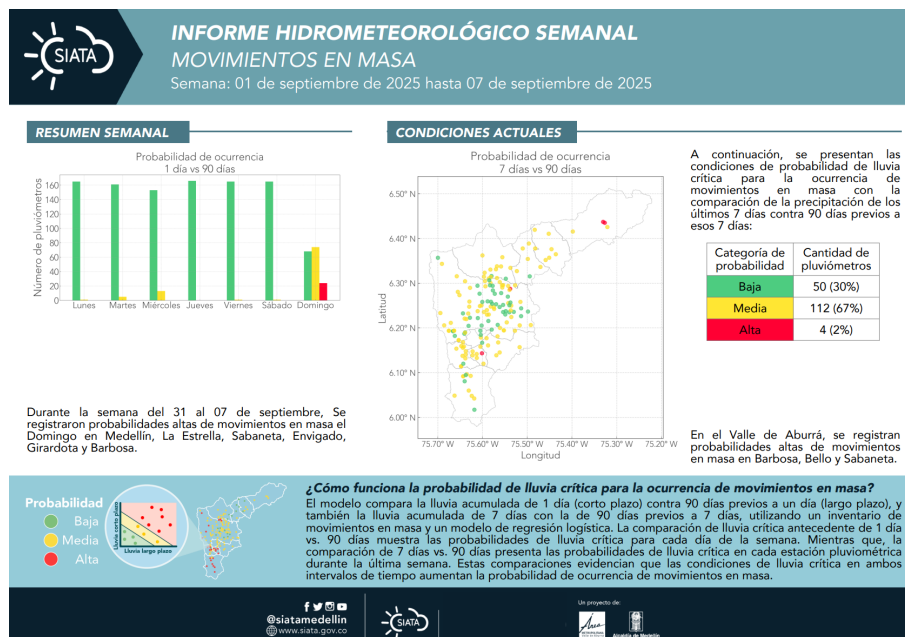


Figura 20. Sección *Movimientos en masa* del reporte hidrometeorológico semanal (versión anterior basada en estaciones pluviométricas).

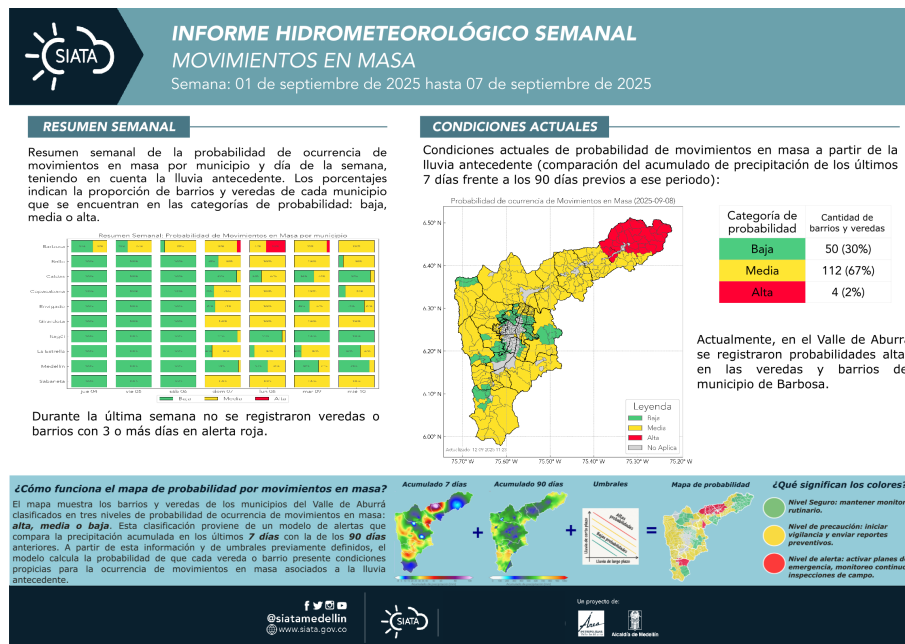


Figura 21. Nueva versión de la sección *Movimientos en masa* del reporte hidrometeorológico: resultados espacializados por barrios y veredas, con resumen semanal por municipios y condiciones actuales.

Cálculo de la incertidumbre del Modelo de Alertas Paralelamente, durante el mes de septiembre se avanzó en uno de los objetivos propuestos para esta vigencia: la cuantificación de la incertidumbre del modelo de alertas.

Durante el mes de agosto se realizó una revisión de literatura con el propósito de identificar metodologías científicas aplicadas a la evaluación del desempeño e incertidumbre de sistemas y modelos de alerta temprana por movimientos en masa.

Como resultado de esta búsqueda, se identificó que la metodología que mejor se ajusta a la estructura y funcionamiento del modelo de alertas actualmente operado es la propuesta por Calvello y Piciullo (2016) en el artículo titulado “Assessing the performance of regional landslide early warning models: the EDuMaP method” Calvello and Piciullo (2016).

Dicho trabajo plantea una metodología denominada EDuMaP (Event–Duration Matrix–Performance), orientada a evaluar y cuantificar la incertidumbre de modelos regionales de alerta temprana por deslizamientos a partir del análisis conjunto de los eventos de deslizamiento registrados y las alertas emitidas por el sistema en un periodo determinado.

Se decidió adoptar esta metodología para la cuantificación de la incertidumbre del modelo de alertas actualmente en operación. A continuación, se presenta una descripción general de su estructura y etapas principales:

El método EDuMaP evalúa el desempeño técnico de un modelo de alerta temprana mediante tres pasos sucesivos:

- I. **Análisis de eventos:** Se identifican y clasifican los **eventos de deslizamiento (LE)** y los **eventos de alerta (WE)** registrados en un mismo periodo temporal. Los deslizamientos se agrupan según su magnitud o densidad espacial, mientras que las alertas se agrupan según su nivel (baja, media, alta, muy alta).
- II. **Construcción de la matriz de duración:** Se genera una matriz temporal en la que cada celda representa el tiempo total durante el cual coincidieron un nivel de alerta (baja, media, alta, muy alta) y una clase de deslizamiento específicos (Eventos pequeños, intermedios y grandes). Esta matriz permite cuantificar cuánto tiempo el sistema estuvo emitiendo alertas correctas, falsas alarmas o sin alertas frente a eventos reales.
- III. **Evaluación del desempeño:** A partir de la matriz de duración se aplican criterios de clasificación y de grado de corrección, los cuales permiten identificar:
 - **Alertas Correctas (CA):** coincidencia entre alerta y evento real.
 - **Falsas alarmas (FA):** alerta sin evento asociado.
 - **Alertas omitidas (MA):** evento sin alerta emitida.
 - **Verdaderos negativos (TN):** ausencia de alerta y de evento.

Luego, se calculan indicadores estadísticos de desempeño tales como eficiencia, tasa de detección, razón de falsas alarmas y poder predictivo, entre otros.

Estos indicadores permiten cuantificar la precisión, sensibilidad y confiabilidad del modelo, y por ende, estimar su nivel de incertidumbre operacional.

En síntesis, la metodología EDuMaP ofrece un marco cuantitativo y reproducible para analizar la correspondencia temporal entre las alertas emitidas y los eventos de deslizamiento observados, permitiendo identificar de manera objetiva los aciertos, errores y omisiones del sistema, y con ello evaluar su incertidumbre y capacidad predictiva.

Mes de octubre

En agosto se realizó una revisión de la literatura para identificar metodologías de evaluación y cuantificación de incertidumbre en modelos de alerta por movimientos en masa aplicadas en otros contextos y potencialmente acoplables al sistema operativo actual. Con base en esa revisión, en septiembre se decidió adoptar la metodología EDuMaP (Event–Duration Matrix–Performance), propuesta por Calvello y Piciullo (2016) en el artículo “Assessing the performance of regional landslide early warning models: the EDuMaP method” Calvello and Piciullo (2016).

En octubre se llevó a cabo una evaluación preliminar del desempeño del modelo de alertas actualmente en operación. La metodología EDuMaP ya había sido presentada de forma general en el informe de septiembre; en este mes se aplicó al modelo para la primera temporada de lluvias de 2025, considerando el periodo comprendido entre el 1 de abril de 2025 y el 30 de junio de 2025, cuando la influencia de las lluvias fue mayor y se registraron eventos significativos de movimientos en masa en distintas zonas del Valle de Aburrá.

Como se explicó en el informe de septiembre, EDuMaP se organiza en tres componentes:

- I. **Análisis de eventos:** identificación y clasificación de eventos de deslizamientos (LE, por sus siglas en inglés *Landslide Events*) y eventos de alerta (WE, *Warning Events*).
- II. **Matriz de duración:** construcción de una matriz temporal en la que cada celda de la matriz representa el tiempo total durante el cual coinciden un nivel de alerta (p. ej., baja, media, alta, muy alta) y una clase de deslizamiento (p. ej., pequeño, intermedio, grande). Esta matriz cuantifica el tiempo en alertas correctas, falsas alarmas y ausencia de alerta frente a eventos reales.
- III. **Evaluación del desempeño:** a partir de la matriz de duración se aplican criterios de clasificación y de grado de corrección para identificar:
 - **Alertas correctas (CA):** coincidencia entre alerta y evento real.
 - **Falsas alarmas (FA):** alerta sin evento asociado.
 - **Alertas omitidas (MA):** evento sin alerta emitida.
 - **Verdaderos negativos (TN):** ausencia simultánea de alerta y de evento.

Posteriormente, se calculan indicadores estadísticos de desempeño —como eficiencia/precisión global, tasa de detección (POD), razón de falsas alarmas (FAR), poder predictivo (PPV), índice de éxito crítico (CSI), entre otros— que permiten cuantificar la precisión, sensibilidad y confiabilidad del modelo y, en consecuencia, estimar su nivel de incertidumbre operacional.

A continuación, se presentan cada uno de los tres pasos anteriores, explicando la metodología utilizada en cada caso y los principales resultados obtenidos:

- I. **Análisis de eventos:** Esta etapa se divide en dos partes: los **eventos de deslizamientos (LE)** y los **eventos de alerta (WE)**.
 - **Eventos de deslizamientos (LE):** Para la identificación de los eventos de deslizamiento se tomaron los registros más significativos ocurridos durante el periodo de estudio, comprendido entre el **1 de abril y el 30 de junio de 2025**. En este intervalo se recopilamos un total de **98 registros** de movimientos en masa en distintas zonas del Valle de Aburrá, ocurridos en diferentes fechas.

En la Figura 22 se presenta el mapa del Valle de Aburrá subdividido en veredas, sobre el cual se ubican los puntos correspondientes a cada uno de los eventos registrados durante la temporada de lluvias, que fueron la base para la definición de los eventos de deslizamientos.

Para la construcción de la matriz de duración, los eventos de movimientos en masa deben clasificarse en **cuatro categorías** ver Tabla 23, según la tabla presentada en el artículo de referencia de Calvello y Piciullo (2016) Calvello and Piciullo (2016).

De acuerdo con dicha tabla, existen **tres criterios** para clasificar los eventos de deslizamientos:

- **Criterio absoluto:** considera únicamente el número de deslizamientos registrados en una zona determinada.
- **Criterio relativo:** se basa en la densidad de deslizamientos por kilómetro cuadrado dentro de cada zona.
- **Criterio mixto:** combina los dos anteriores, ajustando el número de deslizamientos por el tamaño del área analizada.

En este caso, dado que el inventario disponible era relativamente pequeño, se optó por utilizar el **criterio absoluto** para la clasificación de los eventos.

Dado que la matriz de duración se expresa en unidades de tiempo (horas), fue necesario determinar la **fecha y hora de inicio y fin** de cada evento. Para ello, se definió que, si en una misma vereda o barrio ocurren varios deslizamientos separados por menos de **12 horas**, estos se consideran parte de un mismo evento. Si transcurren más de 12 horas entre uno y otro, se considera un evento nuevo. Este procedimiento se aplicó de forma continua desde abril hasta junio de 2025 para todas las zonas con registros de deslizamientos.

El resultado final fue una tabla que contiene la **fecha de inicio y fin del evento**, su **duración en horas**, el **municipio**, el **nombre de la vereda o barrio**, el **número de deslizamientos** asociados y la **clase del evento (LE_class)** asignada de acuerdo con la tabla de clasificación.

Tras la agrupación temporal, de los 98 registros iniciales quedaron definidos **32 eventos de deslizamiento**, ya que varios puntos correspondían al mismo episodio. Estos 32 eventos fueron utilizados posteriormente para la construcción de la matriz de duración y la evaluación del desempeño del modelo.

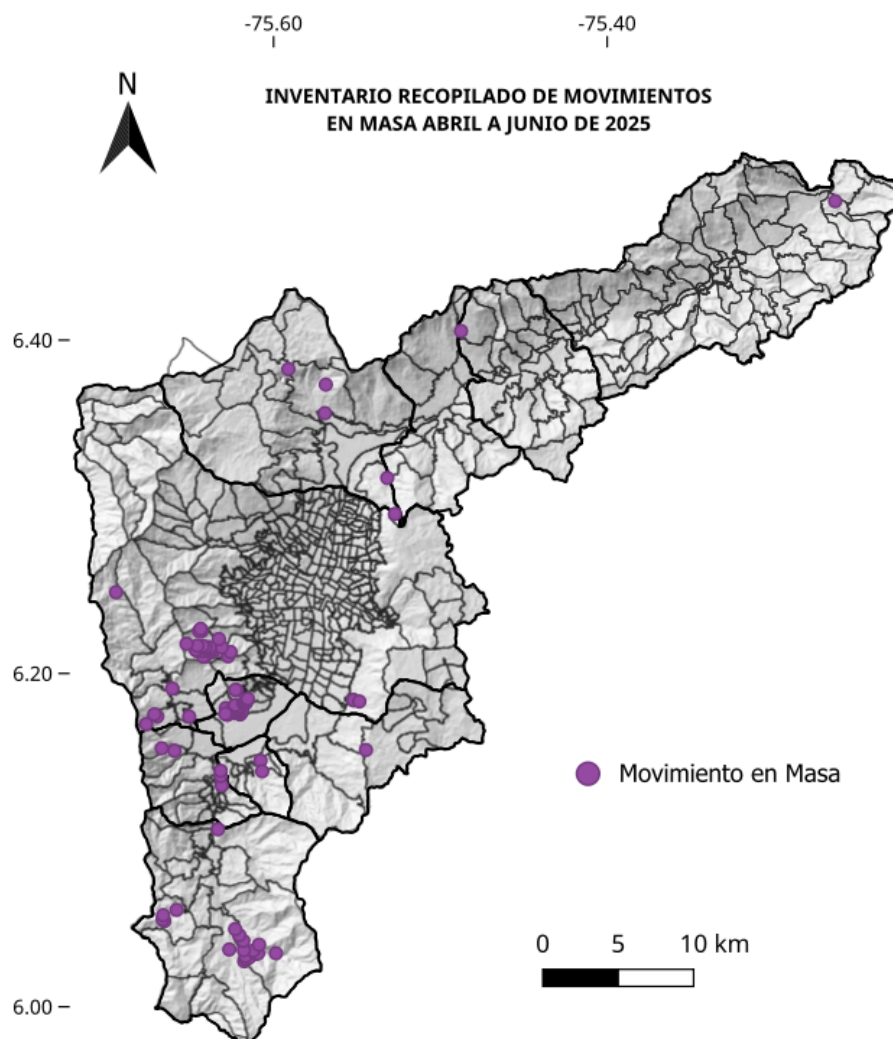


Figura 22. Sección Mapa del Valle de Aburrá subdividido en veredas y barrios con la ubicación del inventario utilizado para la evaluación del modelo.

LE class	Absolute criterion (no. of landslides)	Relative criterion (no. of landslides/area)	Mixed criterion
1	0	0	0
2	1	from 0.001 to 0.02 km ²	1
3	2 to 10	from 0.021 to 0.1 km ²	from 2 to MIN(10; 0.1 km ²)
4	> 10	> 0.1 km ²	> MIN(10; 0.1 km ²)

Figura 23. Tabla tomada de Calvello and Piciullo (2016) donde se explica las diferentes categorías de eventos de deslizamientos.

- **Eventos de alerta (WE):** Los eventos de alerta corresponden a los periodos en los que el modelo de alertas emitió una probabilidad de ocurrencia de movimientos en masa **media o alta** en una o varias zonas. Un evento de alerta se define como el intervalo de tiempo en el que el modelo pasa de una condición de **alerta baja (categoría 1)** a una de **alerta media (categoría 2)** o **alerta alta (categoría 3)**, y finaliza cuando el sistema retorna nuevamente a la condición de **alerta baja**.

Por tanto, en este caso la clasificación de las alertas se establece en **tres clases**:

- **Categoría 1:** probabilidad baja (alerta verde).
- **Categoría 2:** probabilidad media (alerta amarilla).
- **Categoría 3:** probabilidad alta (alerta roja).

Los eventos WE se construyeron a partir de los mapas de alertas generados automáticamente por el modelo cada día a las **4:00 a.m.**, almacenados en el directorio correspondiente (*/mnt/investigacion/geotecnia/historico_modelo/geojsons/*). Se analizaron las variaciones diarias de cada zona (vereda o barrio) entre el **1 de abril y el 30 de junio de 2025**, identificando los periodos en que la probabilidad de deslizamiento aumentó (inicio del evento) y cuando retornó a su nivel base (fin del evento).

Con esta información se obtuvieron los eventos de alerta (WE) para cada zona, incluyendo su **fecha de inicio y fin, duración, y nivel máximo de alerta alcanzado**.

- II. **Matriz de duración:** constituye el elemento central de la metodología EDuMaP, ya que permite evaluar cuantitativamente el desempeño del modelo de alertas al relacionar, en el tiempo, los **niveles de alerta emitidos (WE)** con las **clases de eventos de deslizamientos (LE)**.

Como se muestra en la Figura 24, la matriz se representa como una tabla de doble entrada en la que las **filas** corresponden a las clases de *Warning Events* (WE) y las **columnas** a las clases de *Landslide Events* (LE). Cada celda d_{ij} contiene el **tiempo total (en horas)** durante el cual coexistieron simultáneamente una clase i de alerta y una clase j de deslizamiento. Matemáticamente, cada elemento de la matriz se expresa como:

$$d_{ij} = \sum_{\Delta T} \text{time}_{ij}$$

donde time_{ij} representa las horas en las que una alerta de clase i estuvo activa mientras se registraban deslizamientos de clase j . De esta forma, la suma de todos los elementos $\sum_{i,j} d_{ij}$ es igual al total de horas analizadas dentro del periodo de estudio.

La Figura 24b ilustra el principio temporal de superposición entre los eventos de alerta y de deslizamientos: las líneas horizontales representan la evolución temporal de los niveles de alerta (parte inferior) y las clases de deslizamientos (parte superior). Cada intersección temporal contribuye a un valor d_{ij} dentro de la matriz. Por ejemplo, si durante 10 horas una zona presenta un nivel de alerta alto (WE=3) y simultáneamente se registran deslizamientos clasificados como intermedios (LE=3), esas 10 horas se suman al elemento d_{33} .

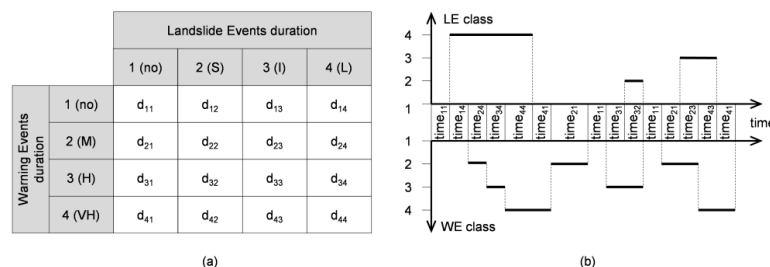


Figure 5. Structure of the duration matrix (a) and graphical exemplification of the temporal analysis needed for its computation (b).

Figura 24. Estructura conceptual de la matriz de duración (a) y representación gráfica del análisis temporal requerido para su cálculo (b). Fuente: Calvello y Picicullo (2016).

Aplicación al caso de estudio:

El análisis se realizó con una **resolución temporal horaria** ($\Delta t = 1$ hora), de modo que cada hora del periodo fue asignada simultáneamente a una clase de alerta (WE) y a una clase de evento de deslizamiento (LE).

El procedimiento se implementó mediante un algoritmo en Python que realiza los siguientes pasos:

- Carga las tablas de eventos de deslizamientos (LE) y de alertas (WE).
- Define la ventana temporal del análisis y genera un índice horario desde el 1 de abril al 30 de junio de 2025.
- Para cada vereda o barrio del Valle de Aburrá, construye dos series horarias:
 - WE(t):** representa el nivel de alerta activo en cada hora (1 = baja, 2 = media, 3 = alta).
 - LE(t):** representa la clase de evento de deslizamiento en esa hora (1 = sin evento, 2 = pequeño, 3 = intermedio, 4 = grande).
- Cada par horario (WE(t), LE(t)) se contabiliza dentro de la celda correspondiente de la matriz d_{ij} .
- Finalmente, las horas se acumulan para todas las zonas, obteniendo una matriz general para todo el Valle de Aburrá.

Este procedimiento permitió cuantificar el tiempo total (en horas) de coincidencia entre las distintas combinaciones de niveles de alerta y clases de eventos de deslizamientos. El resultado se resume en la siguiente tabla:

Tabla 3. Matriz de duración obtenida para el periodo comprendido entre el 1 de abril y el 30 de junio de 2025. Los valores representan horas acumuladas de coexistencia entre niveles de alerta (WE) y clases de deslizamiento (LE).

WE class	LE (1)	LE (2)	LE (3)	LE (4)
1 (Baja)	454,252	7	0	0
2 (Media)	0	0	0	0
3 (Alta)	461,228	13	12	3

En esta matriz se observa que la mayor parte del tiempo las zonas permanecieron en **alerta baja** sin presencia de deslizamientos (celda d_{11}), lo cual es esperable dado que los eventos de movimiento

en masa son relativamente poco frecuentes. Sin embargo, se identificaron también horas en las que coincidieron alertas altas ($WE=3$) con deslizamientos reales ($LE=2-4$), correspondientes a las **alertas efectivas o correctas**.

- III. **Evaluación del desempeño:** A partir de la *matriz de duración* del paso anterior, se evaluó el modelo siguiendo los dos criterios propuestos por Calvello and Piciullo (2016): el **criterio A** (clasificación tipo contingencia) y el **criterio B** (grado de corrección tipo semáforo). En ambos casos, y como recomienda el artículo, se excluye del análisis la celda d_{11} (horas sin alerta y sin deslizamientos), pues suele dominar y enmascara la comparación entre aciertos y errores informativos.

Alert classification criterion (A)		Landslide events			
		1 (no)	2 (S)	3 (I)	4 (L)
Warning events	1 (no)	n/a	TN	MA	
	2 (M)				
	3 (H)	FA	CA		
	4 (VH)				

		Landslide events			
		1 (no)	2 (S)	3 (I)	4 (L)
Warning events	1 (no)	n/a	TN	MA	MA
	2 (M)	TN	TN	MA	MA
	3 (H)	FA	FA	CA	CA
	4 (VH)	FA	FA	CA	CA

Figura 25. Criterios de evaluación propuestos por EDuMaP: (**izq.**) *criterio A* (CA, FA, MA, TN); (**der.**) *criterio B* (verde, amarillo, rojo, morado). Esquema adaptado de Calvello and Piciullo (2016).

Criterio A: clasificación tipo contingencia

Se agrupan las horas de la matriz de duración en cuatro categorías operativas (Fig. 25, izquierda):

- CA (Correct Alerts) :** $WE \geq 2 \wedge LE \geq 2$ (hubo alerta y hubo deslizamiento)
FA (False Alerts) : $WE \geq 2 \wedge LE = 1$ (hubo alerta sin deslizamiento)
MA (Missed Alerts) : $WE = 1 \wedge LE \geq 2$ (no hubo alerta y sí hubo deslizamiento)
TN (True Negatives) : $WE = 1 \wedge LE = 1$ (ni alerta ni deslizamiento).

Con estas cuatro cantidades se calculan los indicadores habituales (trabajando sobre la parte informativa, es decir, excluyendo d_{11} salvo que se indique lo contrario):

$$\begin{aligned} \text{Hit Rate / POD / Sensitivity (HR)} &= \frac{CA}{CA + MA} \\ \text{Precisión / PPV / Predictive Power (PPV)} &= \frac{CA}{CA + FA} \\ \text{Threat Score / CSI (TS)} &= \frac{CA}{CA + FA + MA} \\ \text{Missed Alert Rate (FNR)} &= \frac{MA}{CA + MA} = 1 - \text{HR} \\ \text{False Alert Rate (FAR)} &= \frac{FA}{CA + FA} = 1 - \text{PPV} \\ \text{Accuracy (sin TN) } (I_{\text{eff}}) &= \frac{CA}{CA + FA + MA} = \text{TS} \\ \text{Accuracy (con TN) (Acc)} &= \frac{CA + TN}{CA + FA + MA + TN} \\ \text{Odds Ratio (OR)} &= \frac{CA \cdot TN}{FA \cdot MA} \end{aligned}$$

Nota: Acc y OR pueden inflarse cuando TN es muy grande; por eso el paper recomienda enfocar la comparación con métricas que excluyen d_{11} .

A partir de estas combinaciones se calcularon los indicadores estadísticos de desempeño. Los valores obtenidos para el periodo analizado (abril-junio de 2025) fueron:

Tabla 4. Indicadores de desempeño calculados según el Criterio A.

Métrica	Valor
Eficiencia / Accuracy (sin TN)	6.06×10^{-5}
Tasa de detección (Hit Rate / POD)	0.80
Precisión (PPV)	6.06×10^{-5}
Índice de amenaza (CSI / Threat Score)	6.06×10^{-5}
Tasa de alertas omitidas (Missed Alert Rate)	0.20
Tasa de falsas alarmas (False Alert Rate)	0.9999
Tasa de misclasificación (Misclassification Rate)	0.9999
Accuracy (con TN)	0.4958
Odds Ratio	3.93

Los resultados indican que, durante el periodo analizado, el modelo mostró una alta capacidad para detectar eventos reales ($\text{POD} = 0.8$), pero con una precisión muy baja ($\text{PPV} \approx 0$), lo que sugiere un número considerable de falsas alarmas. Esto se refleja en la tasa de falsas alarmas ($\text{FAR} \approx 0.9999$), que indica que casi todas las horas de alerta no coincidieron con eventos reales. La exactitud general ($\text{Accuracy con TN} = 0.49$) se ve inflada por la gran cantidad de horas sin deslizamientos (TN), típicamente dominantes en este tipo de análisis. El modelo, por tanto, tiende a ser **sensible** (detecta bien cuando ocurre algo) pero poco **preciso** (alerta demasiadas veces sin que ocurra un evento).

Criterio B: grado de corrección (semáforo).

Cada celda de la matriz (excepto d_{11}) se clasifica según su “calidad” (Fig. 25, derecha):

- **Verde (Gre):** mejor respuesta esperable (p.ej., coincidencia razonable entre WE y LE).
- **Amarillo (Yel):** error menor (p.ej., diferencia de un nivel).
- **Rojo (Red):** error significativo (p.ej., FA con $WE \geq 2$, $LE = 1$, o diferencias ≥ 2 niveles).
- **Morado (Pur):** peor error (p.ej., MA con $WE = 1$, $LE \geq 2$).

Sobre el subconjunto informativo $S = \text{Verde} + \text{Amarillo} + \text{Rojo} + \text{Morado}$ se reportan proporciones:

$$\text{ER (errores severos)} = \frac{\text{Rojo} + \text{Morado}}{S}, \quad \text{PSM-NL (solo FA)} = \frac{\text{Rojo}}{S}, \quad \text{PSM-NW (solo MA)} = \frac{\text{Morado}}{S},$$

$$\text{Green share} = \frac{\text{Verde}}{S}, \quad \text{Yellow share} = \frac{\text{Amarillo}}{S}.$$

La regla exacta para asignar los colores puede ajustarse a la política operativa local; la filosofía es: *verde = mejor, morado = peor*.

Normalizaciones (Ecs. 2 y 3 del paper).

Además de las métricas anteriores, el paper propone dos normalizaciones útiles para describir la *estructura* de la matriz (no sustituyen a las métricas de desempeño):

$$d_{ij}^{LE} = \frac{d_{ij}}{\sum_{k=1}^W d_{kj}} N_{LE,j} \quad (\text{distribución, por columnas, relativa a la clase de } LE = j)$$

$$d_{ij}^{WE} = \frac{d_{ij}}{\sum_{k=1}^L d_{ik}} N_{WL,i} \quad (\text{distribución, por filas, relativa a la clase de } WE = i),$$

donde $N_{LE,j}$ es el número de eventos de deslizamiento en la clase j y $N_{WL,i}$ el número de niveles de alerta en la clase i . Estas vistas ayudan a interpretar “cómo se reparte el tiempo” por clase, pero las métricas como POD, PPV o CSI se calculan con los *totales* CA, FA, MA (y opcionalmente TN).

Los resultados para el periodo de análisis fueron los siguientes:

Tabla 5. Indicadores del Criterio B (proporciones relativas al total informativo).

Métrica	Valor
Tasa de errores severos (ER / PSM)	0.9999
Errores severos por alerta omitida (PSM-NW)	1.52×10^{-5}
Errores severos por falsa alarma (PSM-NL)	0.9999
Proporción de aciertos (Green share)	3.25×10^{-5}
Errores menores (Yellow share)	2.60×10^{-5}

Los valores muestran que casi la totalidad del tiempo informativo correspondió a errores severos ($ER \approx 1$), principalmente asociados a falsas alarmas ($PSM-NL \approx 1$). Los aciertos (verde) y errores menores (amarillo) son casi insignificantes, lo que refuerza la necesidad de ajustar los umbrales del modelo. Este comportamiento es similar al observado por Calvello y Piciullo (2016) en zonas del sistema *Alerta Río* donde los umbrales de lluvia eran demasiado bajos, generando exceso de alertas sin eventos reales.

En resumen, la aplicación de los criterios A y B evidencia que el modelo actual presenta un **alto nivel de sensibilidad pero baja precisión**, lo cual es útil para evitar omisiones, aunque incrementa el número de falsas alarmas. Esto sugiere la necesidad de una futura calibración diferenciada por zonas o agregando nuevas variables del terreno, de forma similar a lo planteado en el caso de Río de Janeiro por Calvello and Piciullo (2016).

Conclusiones

Durante el mes de octubre se realizó una primera evaluación de precisión del modelo de alertas por movimientos en masa detonados por lluvias, aplicando la metodología *EDuMaP* propuesta por Calvello and Piciullo (2016).

La metodología se aplicó a la primera temporada de lluvias del año 2025, comprendida entre abril y junio. Para este periodo se recopiló un inventario de 94 movimientos en masa registrados y los mapas de alerta diarios emitidos por el modelo operativo. A partir de esta información se identificaron y agruparon los distintos eventos de deslizamientos, obteniéndose un total de 35 eventos: 20 clasificados como categoría 2, 12 como categoría 3 y 3 como categoría 4. De manera similar, se identificaron los eventos de alerta en los que se emitieron niveles de probabilidad media y alta, alcanzando un total de 2 550 registros de alerta en las distintas veredas y barrios del Valle de Aburrá.

Con estos insumos se construyó la matriz de duración, que permite comparar el tiempo de coincidencia entre los eventos de deslizamiento y los periodos de alerta, sirviendo de base para el cálculo de diversas métricas de desempeño.

Los resultados evidencian que el modelo posee una buena capacidad para emitir alertas cuando ocurren movimientos en masa (alta sensibilidad), aunque también presenta un número elevado de falsas alarmas (baja precisión). Este comportamiento puede deberse a varios factores. En primer lugar, el inventario de deslizamientos utilizado podría no representar la totalidad de los eventos realmente ocurridos durante la temporada de lluvias; por tanto, se recomienda realizar una evaluación con un inventario más robusto y actualizado.

En segundo lugar, se sugiere ampliar el periodo de análisis, incorporando no solo la primera temporada de lluvias, sino también temporadas de menos lluvias y otros periodos en los que se hayan registrado fases variadas del ENSO (La Niña y El Niño), con el fin de evaluar la estabilidad y consistencia del modelo en diferentes condiciones hidrometeorológicas y climáticas.

Asimismo, el modelo fue evaluado dividiendo el Valle de Aburrá en polígonos correspondientes a veredas o barrios, tal como opera actualmente el modelo de alertas. Sin embargo, esta zonificación puede generar sobreestimaciones en áreas extensas o en veredas con pendientes moderadas donde se emiten alertas altas sin que necesariamente ocurran deslizamientos en toda la vereda. En este sentido, se recomienda explorar divisiones más representativas de la dinámica del terreno, como unidades de ladera o zonas de lluvia homogénea.

Finalmente, se plantea la necesidad de ajustar los umbrales de lluvia con los que opera el modelo y considerar la incorporación de variables del terreno, con el fin de reducir el número de falsas alarmas en sectores que, pese a recibir altas precipitaciones, no presentan evidencias de inestabilidad del terreno.

Mes de noviembre

Durante octubre se realizó la primera evaluación del modelo operativo de alertas por movimientos en masa empleando la metodología *Event-Duration Matrix-Performance* (EDuMaP) propuesta por Calvello and Piciullo (2016). Esta evaluación, detallada en los informes anteriores, permitió cuantificar el desempeño del modelo en términos de discriminación y falsas alarmas.

Los resultados mostraron que el modelo tiene una buena capacidad para identificar situaciones de alerta alta cuando efectivamente se presentan movimientos en masa; sin embargo, también exhibe una proporción considerable de falsas alarmas. Esta baja precisión puede estar asociada a varios factores, no

necesariamente relacionados con el modelo en sí.

En primer lugar, el inventario de deslizamientos empleado para la evaluación corresponde únicamente a los eventos más significativos o con reporte de afectaciones durante la primera temporada de lluvias de 2025 (marzo-abril-mayo). Este inventario contiene 94 eventos, pero es probable que hayan ocurrido más deslizamientos no registrados. Por esta razón, se solicitó formalmente al DAGRD un inventario más completo que permita realizar una evaluación más robusta y representativa.

En segundo lugar, esta evaluación inicial constituye un punto de partida para orientar las mejoras del modelo en distintos frentes. La estrategia consiste en implementar ajustes específicos, volver a evaluar el modelo mediante EDuMaP y verificar si las métricas de desempeño presentan mejoras. Esto garantiza que cualquier modificación introducida sea técnicamente justificada y reflejada en beneficios operacionales; y, en caso contrario, permitirá replantear los frentes de mejora necesarios.

El primer frente de mejora, desarrollado durante noviembre, aborda la unidad de análisis del modelo. Actualmente, el modelo opera a nivel de píxel, integrando la lluvia antecedente medida por pluviómetros y radar meteorológico mediante métodos estadísticos para generar un mapa de acumulados con resolución de 500 m. Luego, se evalúan los umbrales definidos para cada celda y se obtiene un mapa de probabilidad, que posteriormente se agrega a escala de vereda o barrio, asignando a cada polígono la probabilidad predominante dentro de su área.

Este enfoque presenta dos limitaciones principales:

- **Sobreestimación espacial:** En veredas extensas (áreas grandes), o con pendientes bajas a moderadas, es común que se generen alertas altas sin que necesariamente existan condiciones reales de inestabilidad en toda la vereda. Esto incrementa la tasa de falsas alarmas y reduce la utilidad operativa del modelo.
- **Representatividad geomorfológica limitada:** El píxel es una unidad arbitraria, rectangular y sin significado geomorfológico. No representa la estructura de las laderas ni la coherencia espacial de los procesos que las gobiernan. Además, los píxeles pueden ser dependientes entre sí, lo que distorsiona la interpretación de la probabilidad a escala local.

Escalar el análisis a cuencas resulta demasiado general, pues una misma cuenca puede contener laderas con morfodinámicas contrastantes. Por otro lado, utilizar veredas o barrios como unidades de análisis tampoco es adecuado, ya que son divisiones político-administrativas que no reflejan la estructura real de las laderas ni sus gradientes de inestabilidad.

Por estas razones, se propone adoptar las **Unidades de Ladera** o **Slope Units** como nueva unidad de análisis del modelo.

Según [?], una Slope Unit es una unidad geomorfológica fundamental del terreno, delimitada por líneas de drenaje y divisorias, que presenta relativa homogeneidad en sus condiciones morfodinámicas. En esencia, representa una porción de ladera donde la topografía, la pendiente, la forma del terreno y los procesos de modelado actúan de manera coherente, convirtiéndose en una unidad funcional para el análisis de susceptibilidad y amenaza.

Estas unidades pueden delimitarse manualmente, pero para áreas extensas el proceso es altamente demandante y subjetivo. Para resolver este problema, Alvioli et al. (2016) desarrollaron el primer algoritmo reproducible y de código abierto para la delimitación automática de Slope Units: el módulo `r.slopeunits` de GRASS GIS. Este algoritmo:

- Delimita automáticamente Slope Units a partir de un DEM.

- Genera polígonos geomorfológicamente coherentes basados en pendiente, curvatura y dirección de drenaje.
- Identifica crestas y cauces y divide el terreno mediante líneas divisorias y líneas de drenaje.
- Agrupa celdas que comparten características topográficas y morfodinámicas sin cruzar límites geomorfológicos.
- Permite ajustar parámetros clave como umbral de acumulación, profundidad mínima de valle y tamaño mínimo de unidad, lo que facilita obtener Slope Units óptimas para modelos de susceptibilidad.

Se adopta la metodología propuesta por Alvioli et al. (2016) para la delimitación automática de las *Slope Units* en todo el Valle de Aburrá. El procedimiento se ejecuta en un entorno Linux con GRASS GIS y el módulo *r.slopeunits* instalado. Este módulo requiere un conjunto de insumos y parámetros que controlan la subdivisión geomorfológica del terreno, los cuales se describen a continuación:

- **Modelo de Elevación Digital (DEM):** Único insumo obligatorio. A partir del DEM se calculan pendiente, curvatura, drenajes, divisorias y acumulación de flujo, que constituyen la base para la partición hidrológica y geomorfológica del terreno.
- **plainsmap (opcional):** Capa binaria que identifica llanuras aluviales o zonas planas donde no se deben delinear Slope Units, dado que en estos sectores no se desarrollan procesos de inestabilidad de ladera.
- **Parámetros numéricos del modelo:**
 - **thresh (t) – Umbral de área de acumulación:** Controla la densidad de la red de drenaje que define los *half basins* (HBs). Valores pequeños generan redes de drenaje densas y HBs pequeños; valores grandes generan redes menos densas y HBs más extensos. Este parámetro determina la estructura hidrológica inicial de la partición. Se da en metros cuadrados.
 - **areamin (a) – Área mínima de una Slope Unit:** Define el tamaño mínimo permitido para cada unidad, evitando polígonos demasiado pequeños o ruidosos. El algoritmo descarta subdivisiones cuya área promedio sea menor que *a*. Se da en metros cuadrados.
 - **cvmin (c) – Varianza circular mínima del aspecto:** Controla la homogeneidad en la orientación del terreno dentro de cada unidad. Valores pequeños (0,1–0,2) producen unidades muy uniformes en su aspecto; valores mayores (0,4–0,6) permiten unidades más heterogéneas. Este parámetro es la restricción geomorfológica central del algoritmo. Adimensional
 - **rf (r) – Factor de reducción del umbral de acumulación:** Regula la disminución progresiva de *t* en cada iteración del algoritmo, siguiendo:

$$t_{i+1} = t_i - \frac{t_i}{r}$$

Valores altos ($r > 10$) producen una disminución lenta y mayor control geomorfológico; valores bajos ($r = 2-5$) generan una reducción rápida y particiones más abruptas. Adimensional.

- **maxarea (opcional) – Área máxima permitida:** Establece un límite superior al tamaño de cada unidad. Cuando un HB supera este valor, deja de subdividirse, evitando la formación de unidades excesivamente grandes en zonas de topografía simple. Metros cuadrados.
- **cleansize – Umbral de limpieza final:** Controla la eliminación o fusión de unidades anómalas al finalizar la delineación preliminar, tales como polígonos demasiado pequeños, alargados o derivados de artefactos del DEM. Metros cuadrados. Existen tres modos de limpieza:

- I. Remover unidades menores al umbral y rellenar con las unidades vecinas.
- II. Eliminar polígonos delgados mediante la opción -m.
- III. Fusionar unidades pequeñas con sus vecinas según similitud de aspecto, mediante la opción -n.

En síntesis, **r.slopeunits** requiere un DEM y un conjunto de parámetros que regulan: (i) la estructura hidrológica inicial (**thresh**, **rf**), (ii) la geometría y coherencia morfodinámica de las unidades (**areamin**, **cvmin**, **maxarea**), (iii) los procesos de corrección y depuración final (**cleansize**), y (iv) la exclusión de zonas planas donde no aplican los procesos de ladera (**plainsmap**).

La adecuada configuración de estos parámetros permite generar unidades de ladera coherentes con la morfología del Valle de Aburrá y más representativas para la mejora del modelo. Por esta razón se realizaron varias ejecuciones del modelo con diferentes combinaciones de parámetros, con el objetivo de identificar no solo las subdivisiones más detalladas, sino también aquellas que mejor reflejaran la organización geomorfológica real de las laderas. En la Tabla 6 se presentan las configuraciones evaluadas.

No. corrida	DEM	thresh (t)	areamin (a)	cvmin (c)	rf	max iter	Observaciones
1	5 m	100 000	5 000	0.5	10	1000	Buena delimitación, pero tiempo de ejecución muy largo.
2	5 m	50 000	5 000	0.5	10	100	Delimitación aceptable; ejecución más rápida, con menor nivel de detalle.
3	5 m	75 000	5 000	0.5	10	100	Buen equilibrio entre detalle y tiempo de ejecución.
4	5 m	75 000	5 000	0.2	10	100	Resultado coherente; menor detalle debido a la restricción del aspecto.
5	5 m	10 000	50 000	0.5	20	100	Regular; ejecución muy lenta y baja coherencia geomorfológica.

Tabla 6. Parámetros utilizados en cada ejecución para la delimitación automática de Slope Units en el Valle de Aburrá.

Como se observa en la Tabla 6, el número de configuraciones evaluadas fue limitado debido al elevado costo computacional asociado al uso de un DEM de alta resolución (5 m), lo que incrementa significativamente el tiempo de procesamiento por parte del algoritmo **r.slopeunits**. Por este motivo se ejecutaron cinco corridas representativas y sus resultados fueron comparados visualmente para seleccionar la delimitación más adecuada.

A partir de esta comparación se determinó que la Corrida 3 ofrece la mejor coherencia geomorfológica: presenta un buen nivel de detalle, las unidades generadas corresponden en la mayoría de los casos con las formas reales de las laderas y el tiempo de ejecución es razonable en comparación con otras configuraciones.

En la Figura 26 se presenta el mapa completo de las Slope Units delimitadas para el Valle de Aburrá utilizando los parámetros de la Corrida 3 de la Tabla 6. Asimismo, se incluye un acercamiento a un sector específico con el fin de resaltar el nivel de detalle alcanzado en la delimitación.

Es importante señalar que se depuraron aquellas unidades con pendientes menores o iguales a 10°, dado que en estos rangos la probabilidad de ocurrencia de movimientos en masa detonados por lluvia es prácticamente nula.

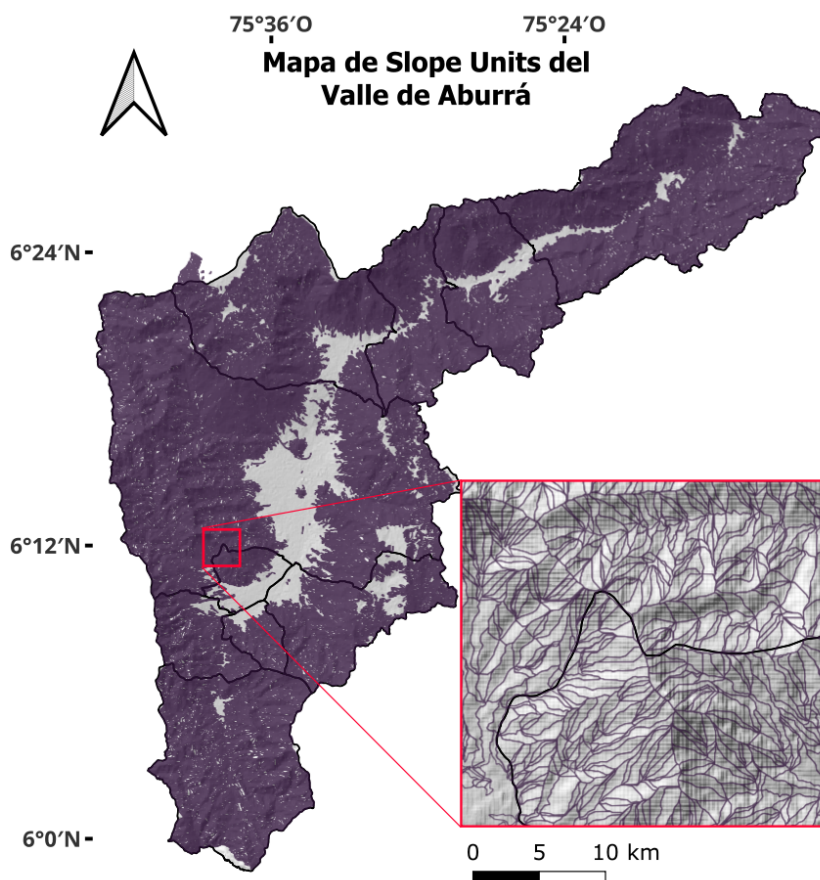


Figura 26. Slope Units delimitadas en el Valle de Aburrá empleando los parámetros de la Corrida 3 de la Tabla 6.

Una vez definida la delimitación de Slope Units candidata, se procedió a ejecutar nuevamente el modelo empleando esta nueva unidad de análisis. El modelo calcula primero los acumulados de lluvia antecedente de corto y largo plazo a partir de la información pluviométrica y del radar meteorológico; posteriormente, ambas fuentes se integran en un único mapa de acumulados, el cual es utilizado para evaluar los umbrales establecidos y estimar la probabilidad asociada.

Con la nueva delimitación, los mapas de lluvia antecedente (corto y largo plazo) se interceptan con la capa de Slope Units para asignar a cada polígono el valor máximo de lluvia presente dentro de su geometría. En consecuencia, cada unidad de ladera obtiene un valor de lluvia acumulada a corto plazo y un valor a largo plazo. Finalmente, cada Slope Unit es evaluada frente a los umbrales operativos, asignando una probabilidad de ocurrencia de movimientos en masa en tres categorías: alta, media o baja.

El siguiente paso consiste en aplicar nuevamente la metodología de validación EDuMaP con el fin de comparar el desempeño del modelo cuando opera a nivel de píxel frente a su ejecución a nivel de Slope Units. Esta evaluación permitirá determinar si las métricas de desempeño mejoran con la nueva unidad de análisis.

Para ello es indispensable contar con un inventario de deslizamientos más robusto para la primera temporada de lluvias del año, tiempo escogido para la validación. La aplicación previa de EDuMaP, realizada el mes pasado para la primera temporada de lluvias, se efectuó con un inventario limitado y con muy pocos eventos, lo que incrementó artificialmente la proporción de falsas alarmas y restringió la capacidad de interpretar adecuadamente los resultados.

Se realizó una solicitud formal al DAGRD y a algunos municipios del Valle de Aburrá para obtener un inventario más completo y actualizado. Se espera que durante el mes de diciembre se disponga de esta información, lo que permitirá llevar a cabo una evaluación más precisa del modelo y comparar de manera confiable si las métricas de validación mejoran bajo la implementación de Slope Units.

Mes de diciembre

Durante el mes de noviembre se realizó la primera propuesta del frente de mejora del modelo de alertas, enfocado en la unidad de análisis. Como propuesta de optimización, se plantea adoptar las **Unidades de Ladera (Slope Units)** como nueva unidad de análisis. No obstante, para validar adecuadamente el desempeño del modelo utilizando **slope units**, es necesario contar con un inventario robusto de movimientos en masa correspondiente a la primera temporada de lluvias del año. En este contexto, durante el mes de noviembre se extendió la solicitud al DAGRD para obtener una base de datos más completa y confiable de los movimientos en masa ocurridos durante la primera temporada lluviosa de 2025.

La respuesta fue recibida el 9 de diciembre, obteniéndose un inventario actualizado de las emergencias asociadas a movimientos en masa registradas en el distrito de Medellín. A partir de este insumo, se pretende depurar y organizar la información con el fin de validar el modelo basado en slope units y compararlo con el modelo anterior basado en celdas o píxeles, evaluando si la nueva unidad de análisis representa una mejora en el desempeño. Se espera que los resultados de esta validación puedan presentarse en los próximos meses, dado que es necesario ejecutar el modelo con slope units para todos los días de la primera temporada de lluvias, lo cual implica un tiempo considerable asociado a las consultas de las bases de datos de precipitación y a la reevaluación del modelo.

Durante el mes de diciembre se decidió abordar un segundo frente de mejora relacionado con las slope units, enfocado en la evaluación óptima de los días de lluvia antecedente de corto y largo plazo, variables fundamentales del modelo. Actualmente, el modelo opera con distintas combinaciones de ventanas temporales de corto y largo plazo. En particular, los reportes semanales del distrito de Medellín se realizan con la combinación de 7 días de corto plazo y 90 días de largo plazo. Se llevó a cabo una evaluación del modelo variando estas combinaciones de días mediante múltiples corridas, utilizando un proceso de Monte Carlo. Para cada combinación de días se ejecutaron múltiples simulaciones, en las cuales se modificaron aleatoriamente los datos correspondientes a no deslizamientos. En total, se evaluaron combinaciones de días de corto plazo entre 1 y 7 días, y de largo plazo entre 8 y 90 días. Cada combinación se repitió 100 veces, con el objetivo de aumentar la robustez de los resultados.

En cada corrida, los eventos de no deslizamiento se asignaron a slope units donde no se han registrado deslizamientos, otorgándoles fechas aleatorias que no coincidieran con un rango de ± 30 días respecto a las fechas del inventario de deslizamientos. Adicionalmente, se exigió que las fechas seleccionadas correspondieran a días lluviosos, con acumulados diarios mayores a 5 mm, acumulados de 7 días superiores a 20 mm y acumulados de 90 días mayores a 100 mm, con el fin de garantizar condiciones representativas de lluvia significativa. La proporción de eventos de no deslizamiento respecto a los deslizamientos se mantuvo constante en una relación 2:1.

Para cada corrida se evaluaron las métricas AUC (Área Bajo la Curva) y F1-Score, ampliamente utilizadas para cuantificar el desempeño de modelos de clasificación. La métrica AUC mide la capacidad del modelo

para discriminar correctamente entre eventos y no eventos, siendo valores cercanos a 1 indicativos de una mayor capacidad predictiva. Por su parte, el F1-Score resume el balance entre la detección correcta de eventos y la reducción de falsas alarmas, integrando la precisión y la sensibilidad del modelo en una única medida. Posteriormente, se calculó la media y la desviación estándar de ambas métricas a partir de las 100 corridas realizadas para cada combinación de días, con el fin de evaluar tanto el desempeño promedio como la estabilidad del modelo.

En las Figuras 27 y 28 se presentan los resultados del análisis de desempeño del modelo para todas las combinaciones evaluadas de ventanas temporales de lluvia de corto plazo (T) y largo plazo (P). Para cada combinación se muestra la mediana de las métricas AUC y F1-Score, junto con su respectiva desviación estándar, calculadas a partir de las corridas del esquema Monte Carlo.

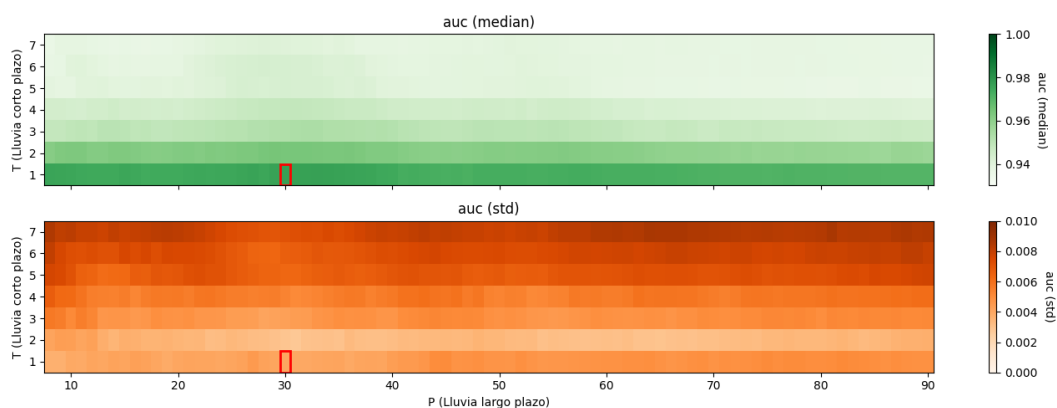


Figura 27. Resultados de la métrica AUC para todas las combinaciones de ventanas temporales de lluvia de corto plazo (T) y largo plazo (P). El panel superior muestra la mediana del AUC obtenida a partir de las corridas Monte Carlo (100), mientras que el panel inferior presenta la desviación estándar asociada. Valores más altos y colores más intensos indican mejor desempeño del modelo, mientras que valores bajos de desviación estándar reflejan mayor estabilidad. El recuadro rojo resalta la combinación de días que presentó el mejor desempeño global.

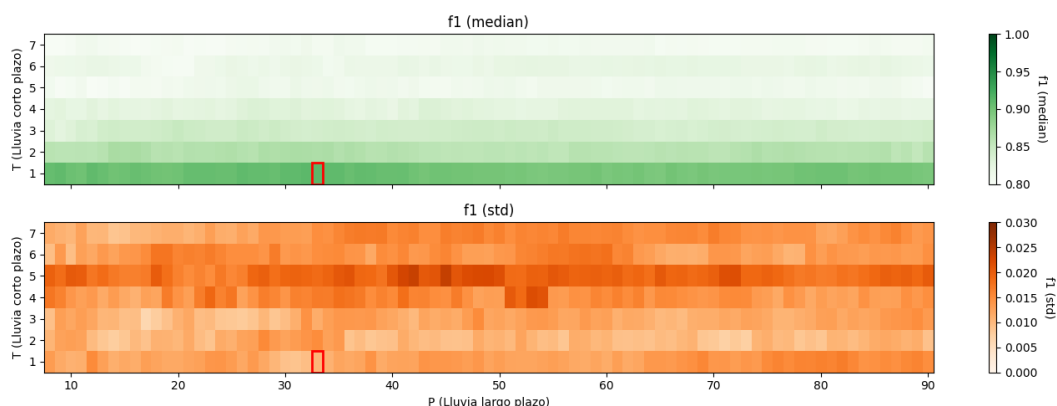


Figura 28. Resultados de la métrica F1-Score para todas las combinaciones de ventanas temporales de lluvia de corto plazo (T) y largo plazo (P). El panel superior corresponde a la mediana del F1-Score, mientras que el panel inferior muestra su desviación estándar, calculadas a partir de las corridas Monte Carlo (100 veces). Las combinaciones con mayor F1-Score y menor dispersión indican un balance óptimo entre detección de eventos y reducción de falsas alarmas. El recuadro rojo identifica la combinación de días con el mejor desempeño relativo.

En general, los resultados evidencian un desempeño alto y estable del modelo en todo el dominio de combinaciones analizadas, con valores de AUC superiores al 94 % y valores de F1-Score mayores al 80 %. No obstante, se observa un patrón claro asociado a la ventana de corto plazo: las combinaciones que utilizan $T = 1$ día presentan los mayores valores medios de AUC y F1-Score, así como las menores desviaciones estándar. En particular, la combinación de $T = 1$ día y $P = 30$ días alcanza el valor máximo de AUC, mientras que la combinación $T = 1$ día y $P = 33$ días muestra el mejor desempeño en términos de F1-Score. Estas combinaciones óptimas se resaltan mediante un recuadro rojo en las figuras correspondientes. Este comportamiento sugiere que la lluvia antecedente de muy corto plazo captura de manera más eficiente el detonante de los movimientos en masa, mientras que la lluvia acumulada en ventanas de largo plazo de 30 días o más aporta información relevante sobre las condiciones de humedad antecedente del terreno.

Adicionalmente, las desviaciones estándar de ambas métricas aumentan de forma progresiva a partir de ventanas de corto plazo de 3 a 4 días, lo que indica una mayor variabilidad en el desempeño del modelo y, por ende, una menor robustez. En contraste, las combinaciones con $T = 1$ y $T = 2$ días presentan una mayor estabilidad, evidenciada por valores de dispersión más bajos tanto en AUC como en F1-Score.

Con base en estos resultados, se concluye que la configuración más adecuada para la evaluación operativa del modelo corresponde al uso de una ventana de lluvia de corto plazo de 1 día y una ventana de largo plazo cercana a 30 días o más. Esta combinación permite maximizar el desempeño predictivo y, al mismo tiempo, reducir la incertidumbre asociada a la variabilidad del modelo, lo que resulta especialmente pertinente para la generación de alertas más coherentes con la dinámica de las lluvias recientes.

Mes de enero de 2026

Durante el mes de **diciembre** se formuló el segundo frente de mejora del modelo de alertas por movimientos en masa, enfocado en buscar la mejor combinación de días de **lluvia antecedente**. Para ello, se llevó a cabo una evaluación sistemática mediante múltiples corridas bajo un esquema de **Monte Carlo**, explorando combinaciones de ventanas de corto plazo entre 1 y 7 días y de largo plazo entre 8 y 90

días, con 100 repeticiones por combinación. Los resultados concluyen que la combinación más adecuada corresponde al uso de una ventana de lluvia antecedente de corto plazo de **1 día** y una ventana de largo plazo cercana a **30 días**, lo que permite capturar de manera más eficiente el efecto detonante de la lluvia reciente y, al mismo tiempo, representar las condiciones de humedad antecedente del terreno.

Sin embargo, quedó pendiente realizar una validación en un horizonte temporal más amplio y evaluar si el desempeño del modelo mejora al aplicar la metodología de validación de EduMap. Esta metodología fue planteada y desarrollada durante los meses de octubre y noviembre de 2025. En una primera aplicación, la metodología se implementó para la primera temporada de lluvias de 2025 (marzo-junio de 2025), evaluando el modelo actualmente operativo. Los resultados indican un desempeño general favorable; no obstante, aunque el modelo identificó adecuadamente varias situaciones de alerta alta asociadas a ocurrencias reales de movimientos en masa, la proporción de falsas alarmas fue elevada. Este comportamiento puede explicarse por varios factores:

- **Limitaciones del inventario para el periodo evaluado:** Para el intervalo marzo-junio de 2025 se disponía únicamente de un inventario de 95 eventos de deslizamiento en todo el Valle de Aburrá. En consecuencia, es posible que existan zonas donde ocurrieron movimientos en masa sin registro en la base de datos; estos casos podrían contabilizarse erróneamente como falsas alarmas, cuando en realidad corresponden a eventos no documentados.
- **Desfase entre el enfoque de EdUmAP y el esquema operativo del modelo:** La metodología EdUmAP está diseñada para trabajar a nivel de eventos de deslizamiento y ha sido probada principalmente en sistemas de alerta temprana que funcionan a escala horaria. En contraste, el modelo actualmente opera a escala diaria y utiliza acumulados de lluvia antecedente de múltiples días. La metodología se describe en detalle en los anexos (particularmente en los avances del mes de octubre) y puede consultarse en el siguiente enlace: Documento de referencia.
- **Condiciones operativas durante el periodo de evaluación:** Entre marzo y junio de 2025, el modelo se encontraba en una fase inicial de operación. En algunas ejecuciones se presentaron fallas asociadas al cálculo de lluvia antecedente (integración de la red pluviométrica y el radar meteorológico). Adicionalmente, el radar registró días con funcionamiento inestable, lo que pudo inducir acumulados artificialmente altos y, en consecuencia, probabilidades/alertas elevadas que no reflejaban condiciones reales.

Considerando las limitaciones anteriores, durante el mes de **enero** se trabajó en replantear la metodología EduMap y adaptarla para que la evaluación del modelo sea consistente con sus condiciones operativas (escala diaria y uso de lluvia antecedente). Además, se identificó como un punto crítico la baja robustez del inventario disponible durante la primera validación (95 eventos), lo cual aumenta la incertidumbre en la estimación de falsas alarmas.

En este contexto, durante el mes de diciembre se recibió respuesta por parte del DAGRD, que suministró un inventario actualizado de emergencias asociadas a movimientos en masa ocurridas en el Distrito de Medellín, con un total de 398 eventos. Si bien este inventario se limita a Medellín y persiste la falta de información equivalente para los demás municipios del Valle de Aburrá, constituye un insumo más robusto para repetir la validación y reducir el sesgo asociado a subregistro de eventos.

La metodología EdUmAP se adaptó para que tanto los eventos de movimientos en masa como las alertas fueran evaluados a escala diaria. Posteriormente, al integrar el inventario actualizado del DAGRD con los 95 eventos previamente disponibles, se consolidó una base de **495 eventos**. Con este insumo se construyó la **matriz de duración**, que cuantifica la frecuencia con la que el modelo emitió alertas **bajas, medias o altas** y su correspondencia con la categoría observada de los eventos de movimientos en masa. Finalmente, se calcularon las distintas métricas de validación bajo el **criterio A** y el **criterio**

B. Todos estos procedimientos se describen en detalle en el documento de referencia: Documento de referencia.

Tabla 7. Matriz de duración obtenida para el periodo comprendido entre el 1 de marzo y el 30 de junio de 2025. Los valores representan horas acumuladas de coexistencia entre niveles de alerta (WE) y clases de deslizamiento (LE).

WE class	LE (1)	LE (2)	LE (3)	LE (4)
1 (Baja)	863,457	84	198	45
2 (Media)	188,128	25	102	1
3 (Alta)	283,098	58	272	156

Tabla 8. Indicadores de desempeño calculados según el Criterio A.

Métrica	Valor
Eficiencia / Accuracy (sin TN)	1.30×10^{-3}
Tasa de detección (Hit Rate / POD)	0.65
Precisión (PPV)	1.30×10^{-3}
Índice de amenaza (CSI / Threat Score)	1.30×10^{-3}
Tasa de alertas omitidas (Missed Alert Rate)	0.35
Tasa de falsas alarmas (False Alert Rate)	0.998
Tasa de misclasificación (Misclassification Rate)	0.998
Accuracy (con TN)	0.65
Odds Ratio	3.44

Tabla 9. Indicadores del Criterio B (proporciones relativas al total informativo).

Métrica	Valor
Tasa de errores severos (ER / PSM)	0.998
Errores severos por alerta omitida (PSM-NW)	6.9×10^{-4}
Errores severos por falsa alarma (PSM-NL)	0.998
Proporción de aciertos (Green share)	6.2×10^{-4}
Errores menores (Yellow share)	6.6×10^{-4}

Los resultados de octubre (resolución horaria, abril-junio 2025, inventario limitado) muestran un modelo **muy sensible** pero **extremadamente poco preciso**. En la matriz de duración, el término dominante fuera de d_{11} es d_{31} (alerta alta sin deslizamientos), con $FA \approx 4,61 \times 10^5$ horas frente a solo $CA = 28$ horas. Esto se refleja en:

$$POD = 0,80, \quad PPV \approx 6,1 \times 10^{-5}, \quad FAR \approx 0,9999, \quad CSI \approx 6,1 \times 10^{-5}.$$

Adicionalmente, cerca del **50 %** del tiempo total analizado quedó en estados con $WE \geq 2$ (en la práctica, casi todo como $WE = 3$), lo cual indica una operación **sobre-alertada**.

En la evaluación ajustada (escala diaria, marzo-junio 2025, inventario consolidado de 495 eventos), el patrón mejora en términos de **precisión y reducción de sobre-alerta**, aunque con una pérdida moderada de sensibilidad. A partir de la matriz:

$$CA = 614, \quad FA = 4,71 \times 10^5, \quad MA = 327, \quad TN = 8,63 \times 10^5,$$

y, por tanto,

$$\text{POD} = 0,65, \quad \text{PPV} = 1,30 \times 10^{-3}, \quad \text{FAR} = 0,998, \quad \text{CSI} = 1,30 \times 10^{-3}.$$

Es decir, **PPV y CSI mejoran ~20 veces** respecto a octubre, y la fracción del tiempo en alerta ($WE \geq 2$) baja de ~50 % a ~35 % (con $WE = 3$ alrededor de 21 % y $WE = 2$ alrededor de 14 %). Esto sugiere una evaluación más coherente con la operación diaria y un menor sesgo por subregistro del inventario.

El *Green share* aumenta de $3,25 \times 10^{-5}$ a $6,2 \times 10^{-4}$ (también ~20 veces), lo cual es consistente con la mejora en PPV/CSI. Sin embargo, el desempeño sigue dominado por **errores severos** ($ER \approx 0,998$), principalmente por **falsas alarmas** ($\text{PSM-NL} \approx 0,998$). Además, el componente por alertas omitidas crece (PSM-NW pasa de $\sim 10^{-5}$ a $\sim 10^{-4}$), coherente con la caída de POD ($0,80 \rightarrow 0,65$) al usar un inventario más robusto y una correspondencia diaria más estricta.

En términos operativos, la segunda evaluación es **mejor** porque:

- reduce la sobre-alerta (menor proporción de tiempo con $WE \geq 2$),
- incrementa de forma clara la precisión (PPV) y el acierto relativo (CSI/Green share),
- y es más representativa del esquema real del modelo (escala diaria + inventario más completo).

La primera evaluación sobreestima la sensibilidad y subestima las falsas alarmas *reales* debido a (i) inventario incompleto y (ii) desajuste entre un método pensado para operación/eventos y un modelo diario con acumulados.

La adaptación diaria de EDuMaP y la ampliación del inventario mejoran la evaluación y muestran que el modelo es **conservador**: detecta una fracción importante de eventos ($\text{POD} \sim 0,65$) pero mantiene una **tasa de falsas alarmas muy alta** ($\text{FAR} \sim 0,998$), indicando que el umbral/criterio de activación aún induce demasiadas horas en alerta sin ocurrencia registrada. Por tanto, el siguiente paso no es “validar más”, sino **recalibrar para reducir FA manteniendo POD aceptable**, por ejemplo: (i) exigir persistencia (k días) para elevar a alerta alta, (ii) ajustar umbrales por zona, y (iv) reforzar control de calidad del radar para evitar acumulados que son reales.

En **febrero** de 2026 se iniciará la exploración de poder direccionar el modelo que actualmente operan y estudiar la viabilidad anexar el pronóstico meteorológico de la precipitación al modelo, con el fin de obtener un modelo del pronóstico en el mediano plazo o en los días siguientes, ya que actualmente el modelo muestra las condiciones actuales pero no las que van ocurriendo en los próximos días. Con este enfoque se busca poder llevar a cabo un modelo que no solo presente las condiciones actuales, si no también un pronóstico que permita mirar cuáles son esas zonas que presentan mayor probabilidad de movimientos en masa dadas unas condiciones de lluvias pronosticadas.

Referencias

- Aleotti, P. (2004). A warning system for rainfall-induced shallow failures. *Engineering Geology*, 73:247–265.
- Alvioli, M., Guzzetti, F., and Rossi, M. (2014). Scaling properties of rainfall induced landslides predicted by a physically based model. *Geomorphology*, 213:38–47.
- Alvioli, M., Marchesini, I., Reichenbach, P., Rossi, M., Ardizzone, F., Fiorucci, F., and Guzzetti, F. (2016). Automatic delineation of geomorphological slope units with r.slopeunits v1.0 and their optimization for landslide susceptibility modeling. *Geoscientific Model Development*, 9:3975–3991.
- Aristizábal, E., Gamboa, F., and Leoz, F. (2010). Sistema de alerta temprana por movimientos en masa inducidos por lluvia para el valle de aburrá, colombia. *Revista EIA*, pages 155–169.
- Balducci, V., Guzzetti, F., Peruccacci, S., and Rossi, M. (2006). Rainfall thresholds for the initiation of landslides.
- Baum, R. L., Savage, W. Z., and Godt, J. W. (2008). Trigrs- a fortran program for transient rainfall infiltration and grid-based regional slope-stability analysis, version 2. 0. Technical report.
- Berti, M., Martina, M. L., Franceschini, S., Pignone, S., Simoni, A., and Pizziolo, M. (2012). Probabilistic rainfall thresholds for landslide occurrence using a bayesian approach. *Journal of Geophysical Research: Earth Surface*, 117.
- Brunetti, M. T., Peruccacci, S., Rossi, M., Luciani, S., Valigi, D., and Guzzetti, F. (2010). Natural hazards and earth system sciences rainfall thresholds for the possible occurrence of landslides in italy. Technical report.
- Calvello, M. and Piciullo, L. (2016). Assessing the performance of regional landslide early warning models: The edumap method. *Natural Hazards and Earth System Sciences*, 16:103–122.
- Chleborad, A. F., Baum, R. L., Godt, J. W., and Leahy, P. P. (2006). Rainfall thresholds for forecasting landslides in the seattle, washington, area-exceedance and probability.
- Crosta, G. B. and Frattini, P. (2001). Rainfall thresholds for soil slip and debris flow triggering. *Proceedings of the 2nd EGS Plinius Conference on Mediterranean Storms*, pages 463–487.
- Gariano, S. L., Brunetti, M. T., Iovine, G., Melillo, M., Peruccacci, S., Terranova, O., Vennari, C., and Guzzetti, F. (2015). Calibration and validation of rainfall thresholds for shallow landslide forecasting in sicily, southern italy. *Geomorphology*, 228:653–665.
- Gianneccchini, R., Galanti, Y., and Avanzi, G. D. (2012). Critical rainfall thresholds for triggering shallow landslides in the serchio river valley (tuscany, italy). *Natural Hazards and Earth System Science*, 12:829–842.

- Guzzetti, F., Cardinali, M., Reichenbach, P., and Carrara, A. (2000). Comparing landslide maps: A case study in the upper tiber river basin, central italy. *Environmental Management*, 25:247–263.
- Guzzetti, F., Peruccacci, S., Rossi, M., and Stark, C. P. (2007). Rainfall thresholds for the initiation of landslides in central and southern europe. *Meteorology and Atmospheric Physics*, 98:239–267.
- Guzzetti, F., Peruccacci, S., Rossi, M., and Stark, C. P. (2008). The rainfall intensity-duration control of shallow landslides and debris flows: An update. *Landslides*, 5:3–17.
- Hidalgo, C. A. and Vega, J. A. (2014). Estimación de la amenaza por deslizamientos detonados por sismos y lluvia (valle de aburrá-colombia). *Revista EIA*, 11:103 – 117.
- IEWP (2005). Dedicated to reducing disasters through effective people-centred early warning systems, world conference on disaster reduction. Technical report.
- Iverson, R. M. (2000). Landslide triggering by rain infiltration. *Water Resources Research*, 36:1897–1910.
- Kirschbaum, D. B., Adler, R., Hong, Y., Hill, S., and Lerner-Lam, A. (2010). A global landslide catalog for hazard applications: method, results, and limitations. *Springer*, pages 561–575.
- Marin, R. J. and Velásquez, M. F. (2020). Influence of hydraulic properties on physically modelling slope stability and the definition of rainfall thresholds for shallow landslides. *Geomorphology*, 351:106976.
- Martelloni, G., Segoni, S., Fanti, R., and Catani, F. (2012). Rainfall thresholds for the forecasting of landslide occurrence at regional scale. *Landslides*, 9:485–495.
- Moreno, H. A., Vélez, M. V., Montoya, J. D., and Rhenals, R. (2006). La lluvia y los deslizamientos de tierra en antioquia: análisis de su ocurrencia en las escalas interanual, intraanual y diaria. *Revista EIA*, 5:59 – 69.
- Peruccacci, S., Brunetti, M. T., Gariano, S. L., Melillo, M., Rossi, M., and Guzzetti, F. (2017). Rainfall thresholds for possible landslide occurrence in italy. *Geomorphology*, 290:39–57.
- Peruccacci, S., Brunetti, M. T., Luciani, S., Vennari, C., and Guzzetti, F. (2012). Lithological and seasonal control on rainfall thresholds for the possible initiation of landslides in central italy. *Geomorphology*, 139-140:79–90.
- Petley, D. (2012). Global patterns of loss of life from landslides. *Geology*, 40:927–930.
- Piciullo, L., Calvello, M., and Cepeda, J. M. (2018). Territorial early warning systems for rainfall-induced landslides.
- Piciullo, L., Dahl, M. P., Devoli, G., Colleuille, H., and Calvello, M. (2017). Adapting the edumap method to test the performance of the norwegian early warning system for weather-induced landslides. *Natural Hazards and Earth System Sciences*, 17:817–831.
- Reichenbach, P., Cardinali, M., De, P., Guzzetti, V. F., Vita, P. D., and Guzzetti, F. (1998). Regional hydrological thresholds for landslides and floods in the tiber river basin (central italy). Technical report.
- Segoni, S., Lagomarsino, D., Fanti, R., Moretti, S., and Casagli, N. (2015). Integration of rainfall thresholds and susceptibility maps in the emilia romagna (italy) regional-scale landslide warning system. *Landslides*, 12:773–785.
- Wu, Y. M., Lan, H. X., Gao, X., Li, L. P., and Yang, Z. H. (2015). A simplified physically based coupled rainfall threshold model for triggering landslides. *Engineering Geology*, 195:63–69.