

**UNIVERSIDAD NACIONAL AGRARIA
LA MOLINA**

**ESCUELA DE POSGRADO
MAESTRÍA EN RECURSOS HÍDRICOS**



**“ESTIMACIÓN DE UMBRALES DE LLUVIA PARA LA
OCURRENCIA DE MOVIMIENTOS EN MASA A PARTIR DE
DATOS DIARIOS GRILLADOS EN EL PERÚ”**

Presentada por:

CARLOS ENRIQUE MILLÁN ARANCIBIA

**TESIS PARA OPTAR EL GRADO DE MAESTRO
MAGISTER SCIENTIAE EN RECURSOS HÍDRICOS**

Lima - Perú

2024

ESTIMACIÓN DE UMBRALES DE LLUVIA PARA LA OCURRENCIA DE MOVIMIENTOS EN MASA A PARTIR DE DATOS DIARIOS GRILLADOS EN PERÚ

INFORME DE ORIGINALIDAD

18%

INDICE DE SIMILITUD

17%

FUENTES DE INTERNET

6%

PUBLICACIONES

6%

TRABAJOS DEL ESTUDIANTE

FUENTES PRIMARIAS

1	repositorio.unheval.edu.pe Fuente de Internet	3%
2	hdl.handle.net Fuente de Internet	2%
3	alicia.concytec.gob.pe Fuente de Internet	1%
4	repositorio.lamolina.edu.pe Fuente de Internet	1%
5	cybertesis.unmsm.edu.pe Fuente de Internet	1%
6	www.researchgate.net Fuente de Internet	1%
7	rstudio-pubs-static.s3.amazonaws.com Fuente de Internet	<1%
8	cdn.www.gob.pe Fuente de Internet	<1%
9	ares.inf.um.es Fuente de Internet	<1%

**UNIVERSIDAD NACIONAL AGRARIA
LA MOLINA**

**ESCUELA DE POSGRADO
MAESTRÍA EN RECURSOS HÍDRICOS**

**“ESTIMACIÓN DE UMBRALES DE LLUVIA PARA LA
OCURRENCIA DE MOVIMIENTOS EN MASA A PARTIR DE
DATOS DIARIOS GRILLADOS EN EL PERÚ”**

**TESIS PARA OPTAR EL GRADO DE MAESTRO
MAGISTER SCIENTIAE**

Presentada por:

CARLOS ENRIQUE MILLÁN ARANCIBIA

Sustentada y aprobada ante el siguiente jurado:

Dr. Néstor Montalvo Arquíñigo
PRESIDENTE

Dr. Waldo Lavado Casimiro
ASESOR

Ph.D. Eduardo Chavarri Velarde
MIEMBRO

Ph.D. Haline Heidinger Abadía
MIEMBRO

DEDICATORIA

A Dios por su simplicidad divina manifestada, cuya esencia da plenitud a la vida y que en medio del caos e incertidumbre permitió los medios y sustentos para iniciar y culminar esta etapa de mi vida.

AGRADECIMIENTOS

A mi familia por su apoyo incondicional e invaluable constancia para conmigo.
Al Dr. Waldo Lavado por su mentoría y asesoría cuyas sugerencias, correcciones y apoyo han contribuido en el desarrollo de esta investigación y sobre todo en mi vida profesional.

ÍNDICE GENERAL

I.	INTRODUCCIÓN	1
1.1.	OBJETIVOS	2
II.	REVISIÓN DE LITERATURA	3
2.1.	MOVIMIENTOS EN MASA	3
2.2.	LLUVIA COMO FACTOR DESENCADENANTE	3
2.3.	UMBRALES DE LLUVIA	4
2.4.	MODELO EMPÍRICO-ESTADÍSTICO	4
2.5.	REGIONALIZACIÓN	5
2.6.	CALIBRACIÓN Y VALIDACIÓN	6
III.	MATERIALES Y METODOLOGÍA	9
3.1.	ÁREA DE ESTUDIO	9
3.2.	MATERIALES	9
3.2.1.	Equipos	9
3.2.2.	Programas de computo	11
3.3.	METODOLOGÍA	11
3.3.1.	Recopilación de datos	11
3.3.2.	Umbrales de lluvia	14
3.3.3.	Regionalización	17
3.3.4.	Calibración y Validación	17
IV.	RESULTADOS Y DISCUSIONES	20
4.1.	UMBRALES DE LLUVIA	20
4.2.	IMPACTO DE LA REGIONALIZACIÓN	22
4.3.	EFFECTO DE LAS CONDICIONES ANTECEDENTES	26
4.4.	EVALUACIÓN DEL DESEMPEÑO DE UMBRALES	26
4.5.	DISCUSIONES	27
V.	CONCLUSIONES	31
VI.	RECOMENDACIONES	33
VII.	REFERENCIAS BIBLIOGRÁFICAS	35
VIII.	ANEXOS	42

ÍNDICE DE FIGURAS

Figura 1	Matriz de confusión para el modelo de clasificación binaria de Umbrales.	6
Figura 2	Curva ROC. El umbral óptimo se muestra como distancia mínima, d , hasta el punto (0, 1) donde la sensibilidad (s_e o S_N) y la especificidad (s_p o S_p) son 1. La tasa de falsos positivos (FPR) es igual a 1-especificidad.	8
Figura 3	Area de estudio. Distribución espacial de los catalogos de MM: Global Landslide Catalog (rojo) y SENAMHI landslide inventory (amarillo).	10
Figura 4	Producto PISCO de P (mm). Valor total para el año 2000.	13
Figura 5	Metodología	14
Figura 6	(a) Extracto de la serie temporal de lluvias diarias (período lluvioso 2019) para una cuenca de ejemplo, donde se observan eventos de lluvia (cada color es un evento de lluvia y los de color plomo son días no lluviosos). (b) Un ejemplo de evento de lluvia asociado a la ocurrencia de un movimiento de masas, en este caso, el evento de lluvia N° 5, donde se muestran las variables analizadas para la estimación de umbrales: la intensidad máxima diaria I_{max} (mm por día), la lluvia acumulada E (mm), la duración D (día) y la intensidad media diaria $I_{mean} = E/D$ (mm por día).	16
Figura 7	Área de estudio subdividido en once regiones de susceptibilidad a movimientos en masa para el Perú, en la cual se incluye la distribución de movimientos en masa usados para la calibración (azul) y validación (amarillo).	18
Figura 8	Diagrama de caja de la precipitación total desencadenante (amarilla) y no-desencadenante (azul) E para las 11 regiones establecidas en este estudio para Perú. Los gráficos de diagrama de caja incluyen valores atípicos y muestran el potencial predictivo de la variable E para separar los eventos desencadenantes y no-desencadenantes de MM poco profundos. Además, el gráfico muestra la variabilidad regional de los eventos desencadenantes de lluvia.	19
Figura 9	Distribución espacial a nivel regional del número de eventos de movimientos en masa (izquierda), eventos de lluvia (centro) y la probabilidad (derecha) de que los eventos de movimientos en masa sean desencadenados por eventos de lluvia.	24

Figura 10 Gráficos de Intensidad-media vs. Duración ($I_{mean} - D$) con curvas de umbral regional en escala logarítmica. El fondo, representado por puntos de colores en una escala de verde-azul-negro, muestra la densidad de eventos de lluvia no-desencadenantes. Los eventos de lluvia desencadenantes se representaron con el mismo color de umbral regional. 25

ÍNDICE DE TABLAS

Tabla 1	Comparación de atributs entre PISCOp y PISCOpd_Op.	12
Tabla 2	Umbrales de precipitación de variables independientes (Th: umbral, Rad: distancia radial mínima, Cal: calibración, Val: validación).	21
Tabla 3	Umbrales de precipitación de dos variables (Th: umbral, Cal: calibración, Val: validación).)	22
Tabla 4	Número de eventos de movimientos en masa y umbrales más óptimos para una y dos variables en cada región (Th: umbral, SL: número de MM por región, Cal: calibración, Val: validación).	23
Tabla 5	Comparación de TSS entre enfoques de validación (1-año: selección de 1 año de la infomación disponible y Aleatoria: selección aleatoria).	30

ÍNDICE DE ANEXOS

Anexo 1	Información de las cuencas de nivel 12 del servicio GEOGloWS ECMWF.	42
Anexo 2	Información de umbrales de una sola variable para cada región.	43
Anexo 3	Probabilidad de ocurrencia de movimientos en masa basado en la presencia de lluvias intensas.	44
Anexo 4	Probabilidad de ocurrencia de movimientos en masa basado en la presencia de lluvias intensas.	45
Anexo 5	Relación de correspondencia entre variables y el número de eventos de movimientos en masa.	46
Anexo 6	Implementación de la investigación en R.	47

LISTA DE ABREVIATURAS

INDECI: Instituto Nacional de Defensa Civil

MM: Movimientos en masa

SENAMHI: Servicio Nacional de Meteorología de Hidrología del Perú

PISCO: Datos Interpolados peruanos de las observaciones climatológicas e hidrológicas del
SENAMHI

PISCO_{pd}_Op: PISCO Operativo de precipitación diaria

ROC: Curva Característica Operativa del Receptor

TSS: True Skill Statistics

RESUMEN

El objetivo de este trabajo fue generar y evaluar umbrales de lluvia regionales obtenidos a partir de la combinación de datos de precipitación grillada de alta resolución (PISCOpd_Op) desarrollado por el Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI) y la información de eventos observados de movimientos en masa desencadenados por lluvias (ML) para obtener umbrales regionales de lluvia. Los datos de eventos de ML se relacionaron con las lluvias desencadenantes y no desencadenantes de estos eventos, obteniéndose propiedades de los eventos de lluvia a partir de los cuales se determinaron umbrales de lluvia. Se realizó una validación del desempeño de los umbrales con eventos ocurridos durante el año 2020 enfocado en evaluar la operatividad de estos umbrales en sistemas de alerta frente a ML en Perú. La determinación de umbrales se realizó para 11 regiones homogéneas determinados en base a las precipitaciones máximas para todo el Perú. El método de determinación de los umbrales fue bajo un enfoque empírico-estadístico; y el rendimiento predictivo de los umbrales se evaluó a partir del “True Skill Statistic” (TSS) y la distancia radial en el espacio de la Curva ROC. El mejor rendimiento predictivo se obtuvo mediante la curva de umbral intensidad media-duración $I_{mean} - D$ seguido por la lluvia acumulada E . Este trabajo, es la primera aproximación de estimación de umbrales regionales a escala del Perú para entender los movimientos en masa (MM) y los resultados obtenidos serán utilizados para el monitoreo y pronóstico de MM causados por lluvias intensas para soportar las acciones de Gestión de Riesgos de Desastres.

Palabras clave: Movimientos en masa, umbrales de lluvia, regiones, huaicos, flujo de detritos, Peru.

ABSTRACT

This work aims to generate and evaluate regional rainfall thresholds obtained from a combination of high-resolution gridded rainfall data, developed by the National Service of Meteorology and Hydrology of Peru, and information from observed shallow landslide events. The landslide data were associated with rainfall data, determining triggering and non-triggering rainfall events with rainfall properties from which rainfall thresholds are determined. The validation of the performance of the thresholds is carried out with events that occurred during 2020 and focuses on evaluating the operability of these thresholds in landslide warning systems in Peru. The thresholds are determined for 11 rainfall regions. The method of determining the thresholds is based on an empirical–statistical approach, and the predictive performance of the thresholds is evaluated with true skill statistics. The best predictive performance is the mean daily intensity-duration ($I_{mean} - D$) threshold curve, followed by accumulated rainfall E . This work is the first estimation of regional thresholds on a country scale to better understand landslides in Peru, and the results obtained reveal the potential of using thresholds in the monitoring and forecasting of shallow landslides caused by intense rainfall and in supporting the actions of disaster risk management.

Key words: landslides, rainfall thresholds, regions, huaicos, debris flow, Peru.

I. INTRODUCCIÓN

Los movimientos en masa (MM) es uno de los peligros de mayor impacto a nivel global causando víctimas y daños a propiedades públicas y privadas, y es responsables de al menos el 17% de todas las muertes por peligros naturales en el mundo (Chae et al. 2017; Segoni et al. 2018). De los diferentes tipos de MM, los provocados por lluvias son los más frecuentes, además que son responsables de víctimas y pérdidas económicas en todo el mundo (Petley 2012). Con respecto a Perú, los MM son el quinto peligro natural que más emergencias han generado en los últimos 16 años (INDECI 2019), junto con las lluvias intensas, bajas temperaturas, vientos fuertes e inundaciones. Sin embargo, en el Perú aún no se ha investigado de forma sistemática la relación de las lluvias con los MM, que considere al entorno fisiográfico y climático del país. Por lo que conocer y entender la interrelación entre los MM y su principal desencadenante, las lluvias; ayudará a proponer herramientas útiles en el monitoreo de las zonas más propensas a MM.

Este estudio tiene como objetivo estimar umbrales regionales de lluvia a escala de todo el país para el monitoreo de las condiciones favorables para la ocurrencia de MM generados por lluvias. La base de datos utilizada para el estudio fue uno de los subproductos derivados de la base de datos de precipitación PISCO (Aybar et al. 2020), denominada PISCO Operativo a paso diario (PISCOpd_Op) el cual es utilizado en los servicios operativos de SENAMHI. Estos umbrales se desarrollaron, en primer lugar, a través de la caracterización de los eventos de lluvias asociadas a eventos de MM desencadenados por lluvias recolectados de dos diferentes catálogos, el primero corresponde a un catálogo de eventos de MM observados y registrados por el SENAMHI y el segundo catalogo corresponde al Catálogo Global Landslides filtrado para Perú (Kirschbaum et al. 2015a). Posteriormente se determinaron los umbrales para las regiones y se evaluó su rendimiento predictivo para la ocurrencia de MM potenciales en Perú. Por último, se realizó un proceso de validación de los umbrales, en un escenario que simule su incorporación en un sistema de alerta temprana con los eventos

registrados durante el año 2020. El enfoque metodológico utilizado puede aplicarse a escala local o de ladera generando umbrales locales utilizando datos locales, lo cual está supeditado al conocimiento de la geomorfología, el tipo de suelo y sus características, por lo que se sugiere que en estudios más localizados se realice un análisis exhaustivo de muestreo en campo para obtener los parámetros de fricción y variables reológicas que caractericen los MM (e.g., Millan y Chavarri-Velarde 2019).

1.1. OBJETIVOS

El objetivo principal de este estudio es la de estimar umbrales regionales de lluvia para la ocurrencia de movimientos en masa desencadenados por lluvias a partir de datos de lluvia grillada y eventos de movimientos en masa. Los objetivos específicos son:

- Caracterizar las lluvias asociadas a movimientos en masa observados en todo el país a partir de regiones de precipitación máxima.
- Analizar el impacto de la regionalización de umbrales en el desempeño de la respuesta de los umbrales de lluvia.
- Evaluar los rendimientos predictivos de los umbrales estimados.

II. REVISIÓN DE LITERATURA

2.1. MOVIMIENTOS EN MASA

Los MM son movimientos de roca y material no consolidados como respuesta a la atracción de la gravedad (Cruden 1991). Aunque los medios de transporte son variados, entre los principales figuran los ríos. Existen diferentes tipos de movimientos de masas utilizados en la literatura. Basados en la clásica clasificación de movimientos de masa de Varnes (1978), actualizada por Hungr et al. (2014) donde se especifica hasta seis tipos de movimientos en masa: caídas o desprendimientos, volcamiento, deslizamiento (rotacional, traslacional o planar), flujos, propagación lateral y complejos (combinación de uno o más tipos de movimiento). Para el contexto de sudamérica, el Proyecto Multinacional Andino: Geociencias para las Comunidades Andinas (Grupo de estándares para movimientos en masa-GEMMA 2007) conformado por los diferentes servicios geológicos nacionales, estableció la traducción del vocablo inglés *landslide* como movimientos en masa.

2.2. LLUVIA COMO FACTOR DESENCADENANTE

Los MM encuentran en la saturación del terreno la causa principal de deslizamiento, este efecto de saturación puede ocurrir de diferentes formas, por ejemplo, las lluvias intensas, deshielo, cambio en el nivel de las aguas subterráneas, y desembalse de agua en lagos, lagunas y reservorios, y un incremento del flujo en canales, quebradas y ríos (Highland 2004). Con estas formas de saturar y alterar las condiciones de estabilidad del terreno, las lluvias es el factor más frecuente e importante para desencadenar los diferentes tipos de MM (Prenner et al. 2018; Segoni et al. 2018). Sin embargo, un punto interesante es que la alta probabilidad de ocurrencia de MM no siempre está asociada a las condiciones extremas de las lluvias intensas y humedad del suelo, sino que también existe una influencia de las condiciones antecedentes

de los días lluviosos previos a la ocurrencia de los MM (Abraham et al. 2020; Leonarduzzi et al. 2017).

2.3. UMBRALES DE LLUVIA

La estimación de umbrales es una de las técnicas para estudiar la interrelación de las lluvias como factor desencadenante de los MM. Dicha técnica ha sido ampliamente aplicada a nivel mundial, mediante diferentes métodos (por ejemplo, métodos empíricos, estadísticos, manuales, probabilísticos y modelos físicamente-basados) (Guzzetti et al. 2007; Segoni et al. 2018; Tang et al. 2019; Berti et al. 2020). Para los MM inducidos por lluvia, un umbral puede definirse como la lluvia, la humedad del suelo o las condiciones hidrológicas que, cuando se alcanzan o se exceden, es probable que se presente un MM. Se han desarrollado umbrales a diferentes escalas temporales (sub horario, horario, diario, mensual) y espaciales (locales, cuenca, regionales, nacionales y globales) dependiendo de la información con la que se cuenta (Segoni et al. 2018). Por ejemplo, se han desarrollado umbrales globales basados en índices de precipitación antecedente (Caine 1980; Guzzetti et al. 2008; Kirschbaum y Stanley 2018) y umbrales nacionales bajo un enfoque empírico-estadístico (Leonarduzzi et al. 2017; Peruccacci et al. 2017; Uwihirwe et al. 2020). En el Perú, los antecedentes de esta línea de investigación se encuentran en los umbrales basados en percentiles del Índice de Lluvia a Antecedente (*Antecedent Rainfall Indice ARI*) del Sistema de Monitoreo de Movimientos en Masa de SENAMHI denominado SILVIA (Millan 2020), instituido con la Resolución Ministerial N.º 049-20020-PCM. Para la cuenca del río Rimác se desarrolló un umbral basado en un análisis de frecuentista, definiendo una curva potencial (Castro 2022).

2.4. MODELO EMPÍRICO-ESTADÍSTICO

Los enfoques empíricos para pronosticar la ocurrencia de MM superficiales inducidos por la lluvia dependen de la definición de los umbrales de lluvia que se vinculen a las variables hidrometeorológicas (Gariano et al. 2015; Segoni et al. 2018). Existen una gran cantidad de variables de análisis que se podrían definir para estudiar los umbrales de lluvia (difer-

entes estudios reportaron hasta veintidós parámetros) (Guzzetti et al. 2007, 2008). Bajo este enfoque, cuando se dispone de información sobre eventos de lluvia que no resultaron en MM, los umbrales tiene como objetivo separar eventos de lluvia desencadenantes de los no-desencadenantes. Este enfoque empírico es ampliamente aplicado pues su análisis e implementación no requiere el monitoreo constante de otras variables físicas en los cuales se basan otros tipos de enfoque, además, esta metodología tiene la ventaja de abarcar diferentes escalas (locales, regionales e incluso a escala global).

Los umbrales pueden ser establecidos para diferentes escalas espaciales dependiendo de la extensión de análisis, los cuales se pueden categorizar en cinco clases: global, nacional, regional, cuenca, local y a escala de ladera. Entendiéndose como escala regional a la subdivisión administrativa de una nación, que típicamente se extiende por miles de kilómetros cuadrados (Segoni et al. 2018).

2.5. REGIONALIZACIÓN

En un estudio a escala nacional, es necesario tomar en cuenta la alta variabilidad meteorológica y fisiográfica espacial del área de estudio, para obtener umbrales de lluvia más precisos y confiables, ya que un solo umbral global o nacional no puede representar dicha variabilidad. Esto se logra a través de la regionalización del área de estudio, en zonas con condiciones meteorológicas homogéneas (Segoni et al. 2014). La regionalización de las lluvias en el análisis de umbrales asociados a MM ha sido usada bajo diferentes enfoques, por ejemplo, se han utilizado índices de precipitación como la media anual, máxima diaria, media mensual, máxima diaria mensual de precipitación, entre otros (Augusto Filho et al. 2020; Segoni et al. 2014), la erodabilidad, y climatología representada por la intensidad máxima diaria de un evento de precipitación (Leonarduzzi et al. 2017).

2.6. CALIBRACIÓN Y VALIDACIÓN

La calibración y validación es un proceso fundamental para definir de forma objetiva los umbrales. La calibración tiene como finalidad la estimación de umbrales a partir de la maximización de la capacidad del rendimiento predictivo o clasificador del modelo de umbral analizado. Y la validación, tiene como finalidad mostrar el potencial de la capacidad de predicción los eventos de lluvias desencadenantes. Entre los enfoques de calibración y validación, se tiene que el más usado y recomendado es dividir los conjuntos de datos para la estimación de umbrales y otro conjunto independiente para la validación Segoni et al. (2018). Para la evaluación del modelo de umbral, se utilizó la matriz de confusión o tabla de contingencia, el cual es una herramienta para determinar la precisión de modelos de clasificación binaria (eventos de lluvias desencadenantes y no desencadenantes) y evaluar el análisis de concordancia entre los resultados del modelo y los datos observados. A partir del cual se determina la cantidad de acierto verdaderos o verdaderos positivos (TP), la cantidad de falsos positivos (FP), la cantidad de verdaderos negativos (TN) y la cantidad de falsos negativos (FN) (Figura 1).

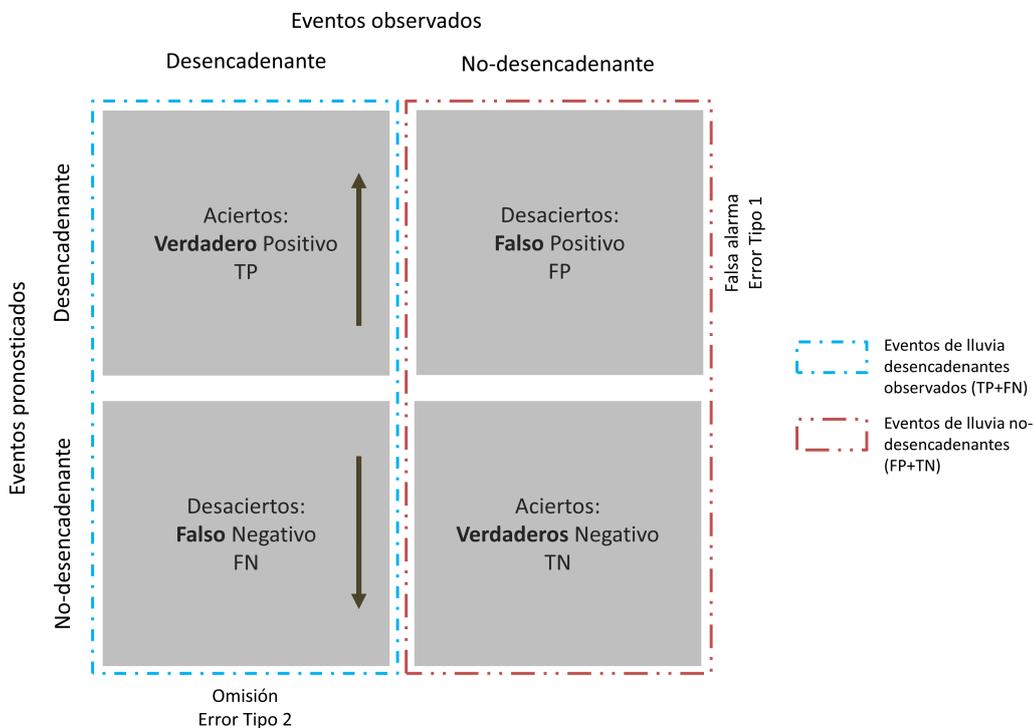


Figura 1: Matriz de confusión para el modelo de clasificación binaria de Umbrales.

FUENTE: Elaboración propia.

Algunas de las medidas más comunes para el pronóstico de movimientos en masa son la sensibilidad ($s_e = TP/(TP + FN)$), especificidad ($s_p = 1 - FP/(FP + TN)$) y la estadística de habilidades verdaderas o *true skill statistic* por sus siglas en inglés ($TSS = s_e + s_p - 1$) (e.g., Staley et al. 2013; Gariano et al. 2015; Leonarduzzi et al. 2017; Mirus et al. 2018; Leonarduzzi y Molnar 2020; Hirschberg et al. 2021). El TSS es una estadística de eficiencia que ayuda en la medición de la bondad de los modelos de umbrales, ya que es una medida integradora del rendimiento predictivo del modelo. El TSS es más objetivo que simplemente una estimación manual aleatoria (Frattini et al. 2010). Varía entre 1 y -1 , con su puntuación óptima igual a 1, que indica el rendimiento máximo del modelo. $TSS = s_e - (1 - s_p)$ es la diferencia entre la tasa de verdaderos positivos (sensibilidad s_e) y la tasa de falsas alarmas ($1 - s_p$), que son los dos componentes más importantes para proporcionar alertas tempranas (Leonarduzzi et al. 2017). El TSS también se conoce como el puntaje de habilidad de Peirce (Peirce 1884), el índice de Youden (Youden 1950) o el puntaje de habilidad de Hanssen-Kuipers (Hanssen y Kuipers 1965). El beneficio de usar la especificidad sobre la tasa de falsos positivos ($FPR = FP/(FP + TN)$) es que en un modelo perfecto TSS, sensibilidad y especificidad son iguales a 1 (Hirschberg et al. 2021). Otra métrica es la distancia radial mínima a la prueba de clasificación perfecta ($TSS=1$, con $s_e=1$ y $1-s_p=0$) de la llamada curva operativa del receptor (ROC) (Fawcett 2006) (Figura 2) ampliamente usada en la definición de umbrales (e.g., Uwihirwe et al.; Gariano et al.; Postance et al.).

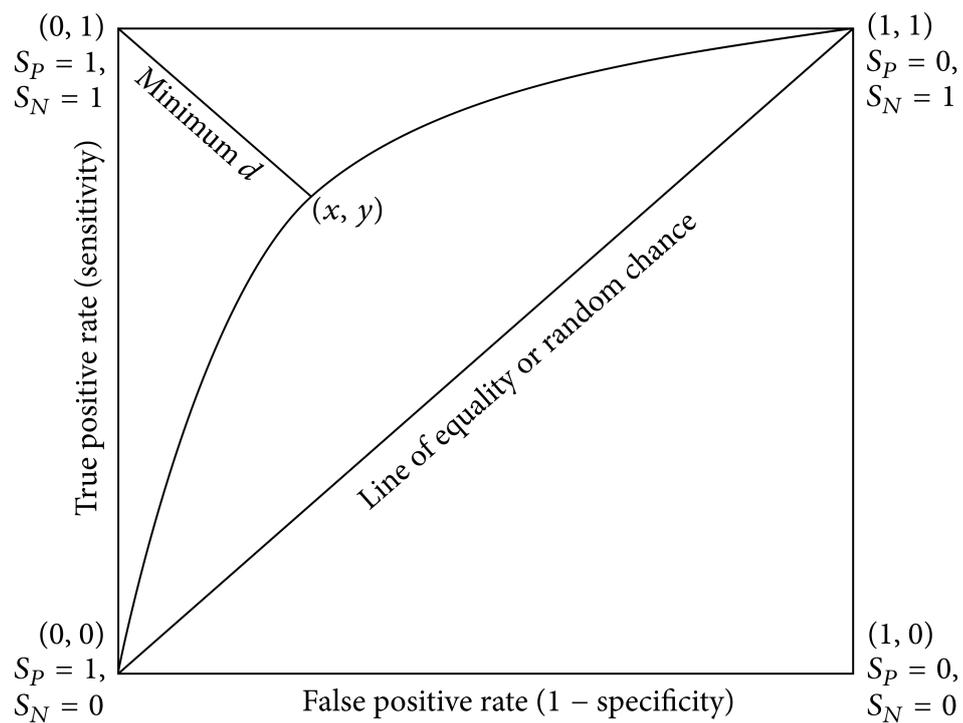


Figura 2: Curva ROC. El umbral óptimo se muestra como distancia mínima, d , hasta el punto $(0, 1)$ donde la sensibilidad (s_e o S_N) y la especificidad (s_p o S_p) son 1. La tasa de falsos positivos (FPR) es igual a 1-especificidad.

FUENTE: (Huynh et al. 2015).

III. MATERIALES Y METODOLOGÍA

3.1. ÁREA DE ESTUDIO

Este trabajo adopta el dominio definido para el Sistema de Monitoreo de Movimientos en Masa Potenciales generados por Lluvias Intensas - SILVIA (Millan 2020; Millan et al. 2021) del Servicio Nacional de Meteorología e Hidrología (SENAMHI) para Perú. Esta área de estudio fue obtenido a partir de la unión de dos bases de datos, el primero es el mapa susceptibilidad a MM del Instituto Geológico, Minero y Metalúrgico del Perú (Villacorta et al. 2012), el cual cuenta con cinco (5) categorías de susceptibilidad. Para el sistema SILVIA no se incluyeron las zonas con susceptibilidad muy baja. La segunda base de datos, fueron las cuencas usadas en el servicio de caudales GEOGloWS ECMWF (David et al. 2011; Qiao et al. 2019; Souffront Alcantara et al. 2019), a partir del cual el dominio de este estudio fue discretizado contando con 5373 cuencas con áreas cuya mediana es de aproximadamente 105 km². El área de estudio y la distribución del área de las cuencas en el área de estudio se muestra en la Figura 03.

3.2. MATERIALES

3.2.1. Equipos

- 01 Workstation - HP Z8 G4 (2 procesadores Intel Xeon silver, 24 núcleos, 88 RAM, 16 GB GPU).

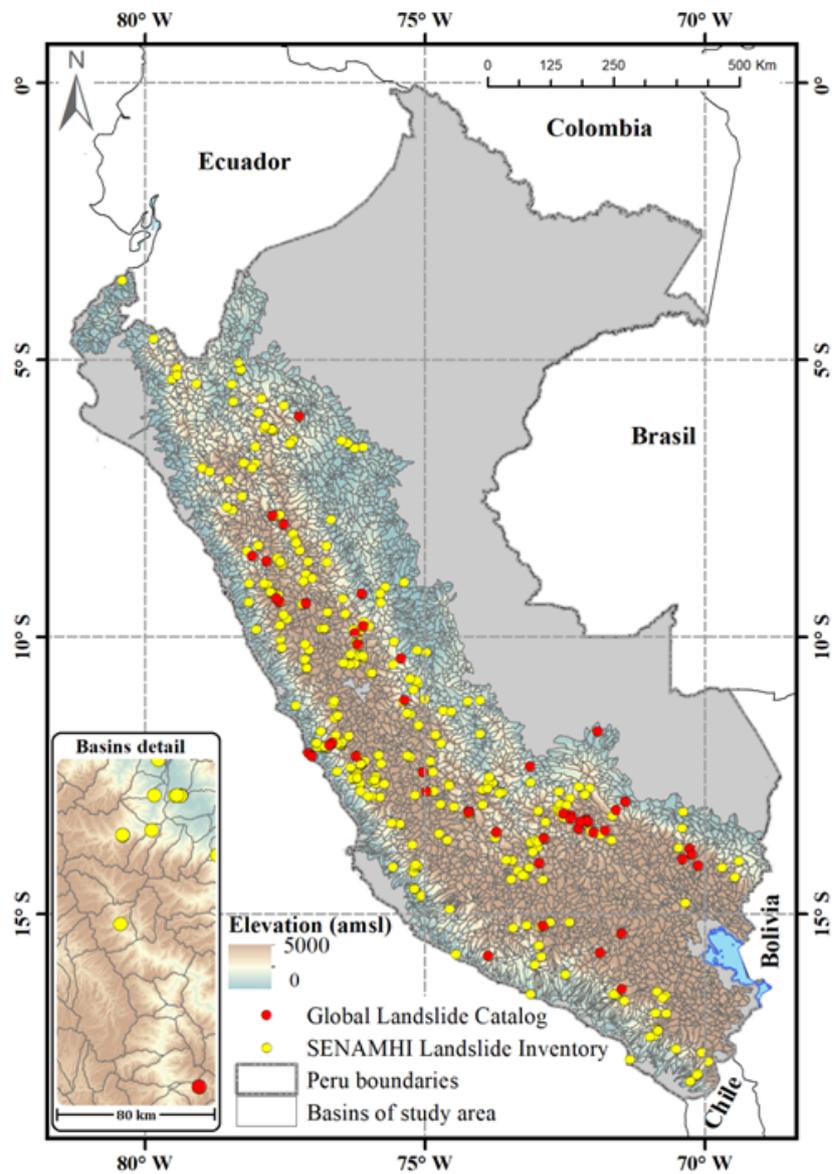


Figura 3: Area de estudio. Distribución espacial de los catalogos de MM: Global Landslide Catalog (rojo) y SENAMHI landslide inventory (amarillo).

FUENTE: Elaboración propia.

3.2.2. Programas de computo

- R 4.0.3 - <https://www.r-project.org/>.
- RStudio 2022.07.1 - <https://www.r-project.org/>.
- ArcMap 10.6 - <https://www.esri.com/>.
- L^AT_EX 2020 - <https://www.overleaf.com>.

3.3. METODOLOGÍA

3.3.1. Recopilación de datos

Precipitación El área de estudio es a escala de todo el Perú, se requiere información que cubra todo el territorio nacional. Por lo que, la principal fuente de información de este estudio, fue la base de datos de precipitaciones diarias grillada PISCOpd_Op (Gridded Daily Rainfall Operative data of PISCO). De entre todas los datos grillados disponibles a tiempo real a escala nacional y global, PISCOpd_Op representa la mayor asimilación de estaciones convencionales, y hace uso de otras fuentes de información para estimar los valores de precipitación, adicionalmente, presenta un rango temporal adecuado para la investigación. PISCOpd_Op es un conjunto de datos operacionales de precipitaciones que forma parte de los datos interpolados peruanos de las Observaciones Climatológicas e Hidrológicas derivado del producto de precipitación del SENAMHI (Peruvian Interpolated data of SENAMHI's Climatological and Hydrological Observations), conocido como PISCO, con datos grillados de precipitación (Aybar et al. 2020), temperatura del aire (Huerta et al. 2018), evapotranspiración de referencia (Huerta et al. 2022) y caudales diarios y mensuales (Llauca et al. 2021, 2023) a escala de todo el Perú. PISCOpd_Op tiene una resolución espacial de 0,1° y una resolución temporal diaria. PISCOpd_Op tiene datos desde 1981 y se actualiza diariamente, acumulando precipitaciones diarias (de 07:00 a 07:00 hora local, LT), generadas a partir de 416 pluviómetros de la red de estaciones convencionales de SENAMHI (ver Fig. 7). PISCOpd_Op se genera en base a un método de interpolación genRE (van Osnabrugge et al. 2017), que consiste en una

interpolación de distancia inversa ponderada (IDW) que incluye multiplicadores que se basan en la climatología mensual de PISCOp. Una comparación entre PISCOp y PISCOpd_Op se encuentra en la Tabla 1 y Figura 4.

Tabla 1: Comparación de atributos entre PISCOp y PISCOpd_Op.

Característica	PISCOp v2.1	PISCOpd_Op
Escala temporal	Daily	Daily
Resolución espacial	10km x 10km	10km x 10km
Periodo	1981-2016	1981-present
Dominio	Peru	Peru
Actualización	Esporádica	Diaría
Referencia	Aybar et al. (2020)	-

Movimientos en masa La otra fuente de información que se utilizó para el desarrollo de esta investigación son dos inventarios de eventos de MM: el primero es el Catálogo de Movimientos en Masa Generados por Lluvias del SENAMHI (LCS) y el segundo es el Catálogo Global de Movimientos en Masa (GLC) de la NASA (Kirschbaum et al. 2015b). Estos catálogos fueron desarrollados con el objetivo de identificar eventos de MM provocados por lluvias en el Perú y todo el mundo respectivamente. Ambos catálogos consideran todo tipo de MM provocados por la lluvia, que han sido reportados en los medios de comunicación, bases de datos de agencias asociadas a desastres, informes científicos u otras fuentes disponibles.

El LCS se implementó en enero de 2019 y tiene 330 registros del período 2014-2020. El GLC tiene 6788 registros para todo el mundo, mientras que para Perú se han registrado 49 eventos de movimientos en masa, que se distribuyeron temporalmente entre 2007 y 2014. Para utilizar estos datos, se realizaron análisis exploratorios para evitar inconsistencias en el registro de los eventos. La correspondencia espacial de los datos se evaluó a través de un subconjunto espacial entre las ubicaciones del evento y el área de estudio. También se evaluó la consistencia de los datos con respecto a los errores tipográficos. Como resultado, se determinaron dos eventos incongruentes: el primero fue reportado en un lugar sin condiciones de ocurrencia de movimientos en masa (fuera del área de estudio) y, por lo tanto, no fue

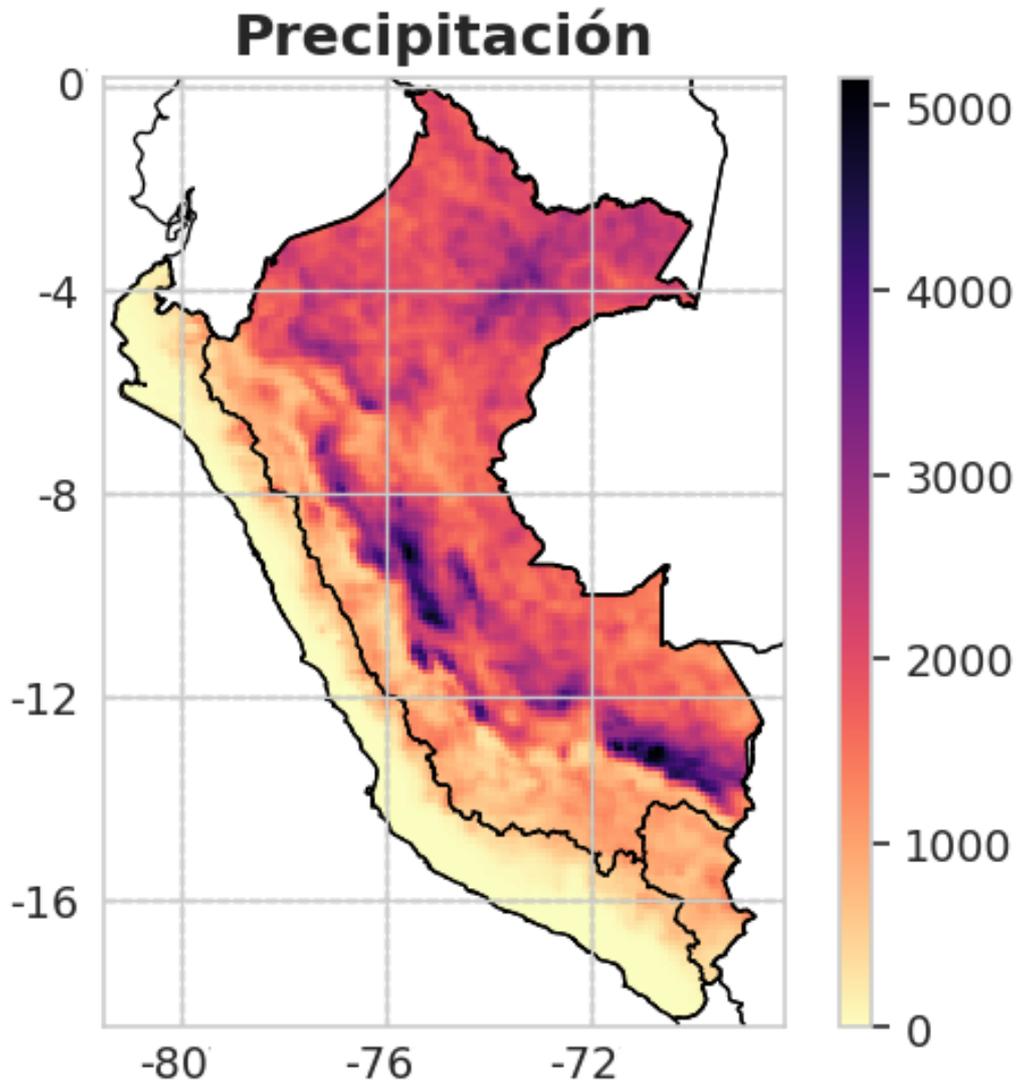


Figura 4: Producto PISCO de P (mm). Valor total para el año 2000.

FUENTE: (Huerta Julca 2020)

considerado en el análisis. En el segundo evento, se determinó un error en su tabulación; Este error fue corregido y el evento fue incluido en el análisis. El número total de registros de movimientos en masa es 377, y la distribución espacial de estos eventos se muestra en la Fig. 3.

3.3.2. Umbrales de lluvia

Para definir los umbrales de lluvia en zonas susceptibles a MM, se empleó un enfoque empírico-estadístico que consistió en los siguientes pasos: (1) la determinación de eventos de lluvia a partir de series históricas de precipitación, (2) la definición de variables de los eventos de lluvias, (3) la identificación de eventos de lluvia desencadenantes y no-desencadenantes de movimientos en masa; y (4) la calibración de umbrales para las propiedades de los eventos de lluvia basados en la maximización objetiva del rendimiento predictivo, se resume en (5). A continuación, se detalla los pasos tomados en cuenta para la estimación de los umbrales.

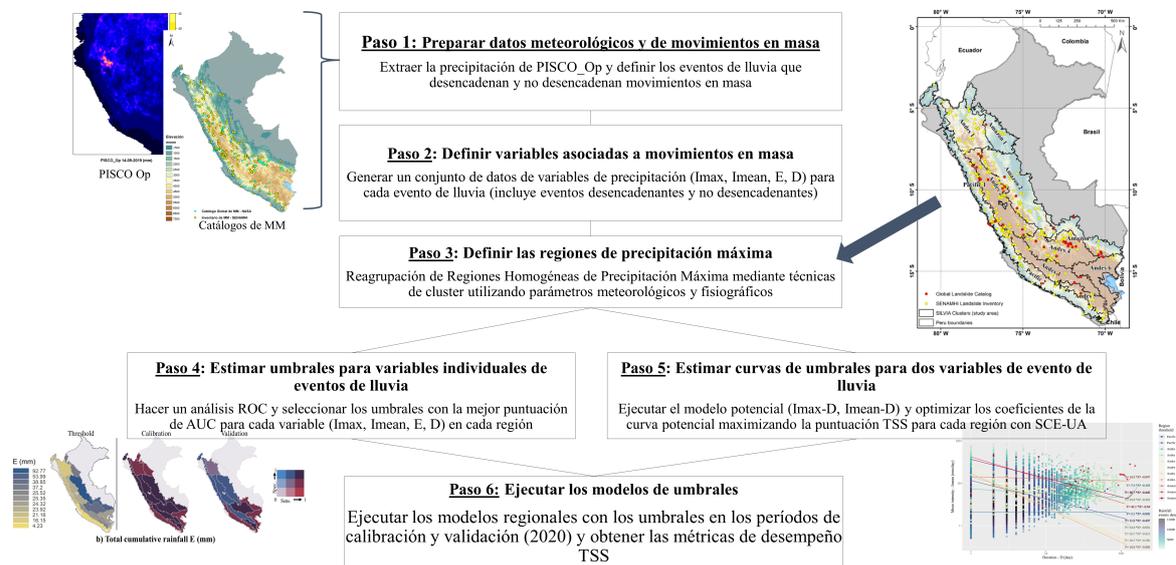


Figura 5: Metodología

FUENTE: Elaboración propia.

El primer paso fue la construcción de una serie histórica de precipitaciones a partir de datos de precipitación cuadrículados (PISCOpd_Op) para cada cuenca que tuvo un mínimo de un evento de movimiento en masa. Después de obtener la serie de lluvias, se definieron los eventos de lluvia junto con una serie histórica para cada cuenca seleccionada. Para este trabajo, definimos un evento de lluvia independiente como una serie de días lluviosos consecutivos donde ha llovido por encima de un umbral mínimo de lluvia (Fig. 6). Diversidad de autores utilizan umbrales mínimos de 1 mm para definir los días lluviosos (Dai 2006; Dai et al. 2007; Han et al. 2016; Leonarduzzi et al. 2017; Shen et al. 2021; Tian et al. 2007; Yong et al. 2010). Sin embargo, dada la gran variabilidad espacial climatológica en el área de estudio, se de-

terminó que no había un único umbral mínimo para todo el territorio, sino que se discretizó un umbral mínimo a partir del sesgo de PISCOpd_Op para los días no lluviosos. El sesgo de PISCOpd_Op se determinó cuando los pluviómetros no informaron lluvia (0 mm), y el umbral mínimo discretizado (U_{min}) de lluvia se definió de acuerdo con la siguiente Ec. 1:

$$U_{min} = \begin{cases} U_0 & \text{if } s \leq U_0 \\ s & \text{if } s > U_0 \end{cases} \quad (1)$$

donde s es el promedio de sesgo simple cuando las estaciones pluviométricas registraron un valor 0 de precipitaciones en comparación con la estimación en PISCOpd_Op. Y U_0 es el umbral mínimo inicial de precipitación, y se establece en 1 mm para todas las regiones, con excepción de las regiones costeras del Pacífico en las que se considera 0,5 mm. Una vez que se definieron los eventos de lluvia, se clasificaron en eventos desencadenantes o no-desencadenantes, es decir, si se produjo un movimiento en masa durante el evento de lluvia.

El segundo paso consistió en definir las variables asociadas a los eventos de lluvia desencadenantes y no-desencadenantes, para lo cual se calculó la intensidad diaria máxima I_{max} (mm/d), la lluvia acumulada E (mm), la duración D (d), y la intensidad media diaria $I_{mean} = E/D$ (mm/d). Estas variables, se determinaron para dos escenarios, la primera considera la duración total del evento (Evento completo-EC) hasta el día de ocurrencia del MM, y el segundo escenario solo considera el evento hasta un día antes de la ocurrencia del MM (Evento Antecedentes-EA). El motivo de estudiar ambos escenarios es permitir evaluar de forma separada la incidencia de las condiciones antecedentes para la ocurrencia de los MM, lo que nos permitirá establecer las relaciones de las condiciones favorables mínimas antecedentes para la ocurrencia de MM. La evolución temporal de las variables hidrometeorológicas proveen una idea de como están las condiciones críticas para el desarrollo de la activación de MM (Prenner et al. 2018; Segoni et al. 2018).

El tercer paso consistió en dividir el área de estudio en regiones basados en técnicas de cluster (este paso se explica con más detalle en la Sección 3.3.3.). El cuarto paso y quinto paso consistió en la definición y selección objetiva de umbrales de lluvia que separen eventos

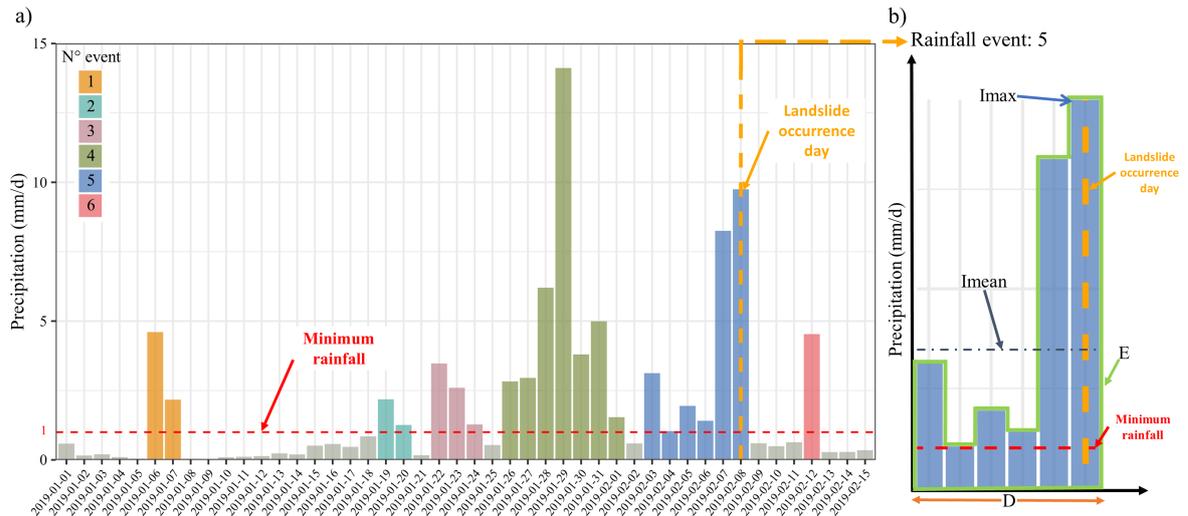


Figura 6: (a) Extracto de la serie temporal de lluvias diarias (período lluvioso 2019) para una cuenca de ejemplo, donde se observan eventos de lluvia (cada color es un evento de lluvia y los de color plomo son días no lluviosos). (b) Un ejemplo de evento de lluvia asociado a la ocurrencia de un movimiento de masas, en este caso, el evento de lluvia N° 5, donde se muestran las variables analizadas para la estimación de umbrales: la intensidad máxima diaria I_{max} (mm por día), la lluvia acumulada E (mm), la duración D (día) y la intensidad media diaria $I_{mean} = E/D$ (mm por día).

FUENTE: Elaboración propia.

de lluvia desencadenantes de los no-desencadenantes basado en rendimientos predictivos. Los umbrales fueron establecidos maximizando la capacidad predictiva en dos enfoques: el primero incluye cada variable de forma independiente (I_{max} , E , D , I_{mean}), y el segundo enfoque fue mediante curvas potenciales que relacionó dos variables ($I_{max} - D$ y $I_{mean} - D$) en la forma de $V = a \cdot D^{-b}$, donde V representa las variables I_{max} y I_{mean} ; a y b son los parámetros de escala y forma de la curva (mientras que para el espacio logarítmico, a es el parámetro de intersección y b denota la pendiente de la curva linealizada). Finalmente el sexto paso consistió en la aplicación del modelo para obtener las métricas de rendimiento y comparar los resultados para cada región tanto en las fases de calibración y validación.

3.3.3. Regionalización

En este estudio se utiliza las Regiones Homogéneas de Precipitación Máxima del SENAMHI (Yupanqui et al. 2017) como insumo para la regionalización del área de estudio. Estas regiones se determinaron basado en técnicas de “clustering” a partir de la información de precipitación de 535 estaciones automáticas. En la cual nativamente se identificaron 10 macrorregiones y 30 subregiones de precipitación máxima. Las regiones climáticas establecidas para el presente estudio, consistió en un agrupamiento de las 30 regiones de precipitación máximas. El reagrupamiento consistió en un análisis multicriterio basado principalmente en que las regiones agrupadas no superen un umbral del valor 10 en el test de heterogeneidad (Hosking y Wallis 1997), que incluyeran eventos registrados en las bases de datos además de compartir similitud de las covariables de relieve (altitud) y climatología (precipitación media). Si bien es cierto que este valor de 10 supera el nivel de heterogeneidad recomendados en 2, se contempla esta tolerancia pues se tratan de regiones obtenidas de un reagrupamiento. A partir de este análisis, se obtuvo 11 regiones para el área de estudio, ver Figura 7. Se definieron 4 umbrales de variables independientes (I_{max} , E , D , I_{mean}) y 3 umbrales curva ($I_{max} - D$ y $I_{mean} - D$) para cada región. Siendo en total 77 umbrales por región. La Figura 8 presenta un diagrama de caja E de lluvia acumulada que muestra su poder predictivo para discriminar entre eventos de lluvia desencadenantes y no desencadenantes.

3.3.4. Calibración y Validación

Para la fase de calibración, se seleccionaron todos los eventos ocurridos antes del 2020, lo que representa alrededor del 70% de los eventos registrados. Mientras que el proceso de validación de los umbrales consistió en evaluar los umbrales para los eventos de movimientos en masa registrados en el año 2020, los cuales representan alrededor del 30% de los eventos registrados. Se realizó este proceso para el año 2020, pues se deseaba conocer cómo sería el desempeño de los umbrales cuando se asimilen a un sistema de alerta temprana regional. Este método de calibración/validación que establece 1 año del conjunto de datos en el procedimiento de validación es un método que se ha utilizado en otras investigaciones (e.g.,

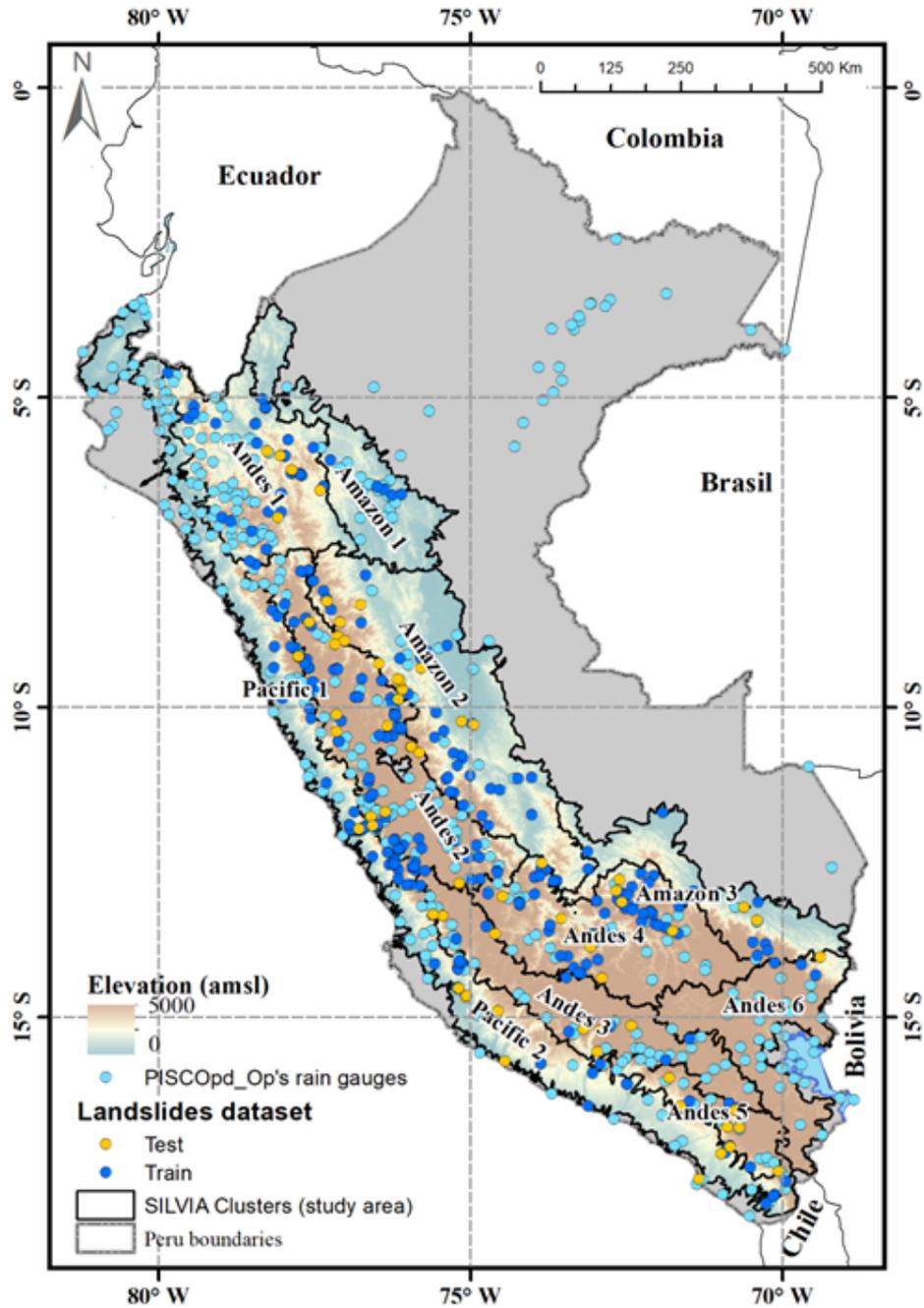


Figura 7: Área de estudio subdividido en once regiones de susceptibilidad a movimientos en masa para el Perú, en la cual se incluye la distribución de movimientos en masa usados para la calibración (azul) y validación (amarillo).

FUENTE: Elaboración propia.

(Kirschbaum et al. 2015b; Dikshit et al. 2019).

Para los umbrales basados en variables independientes (I_{max} , E , D o I_{mean}), la impresión general del poder predictivo se estimó con la denominada curva característica operativa del

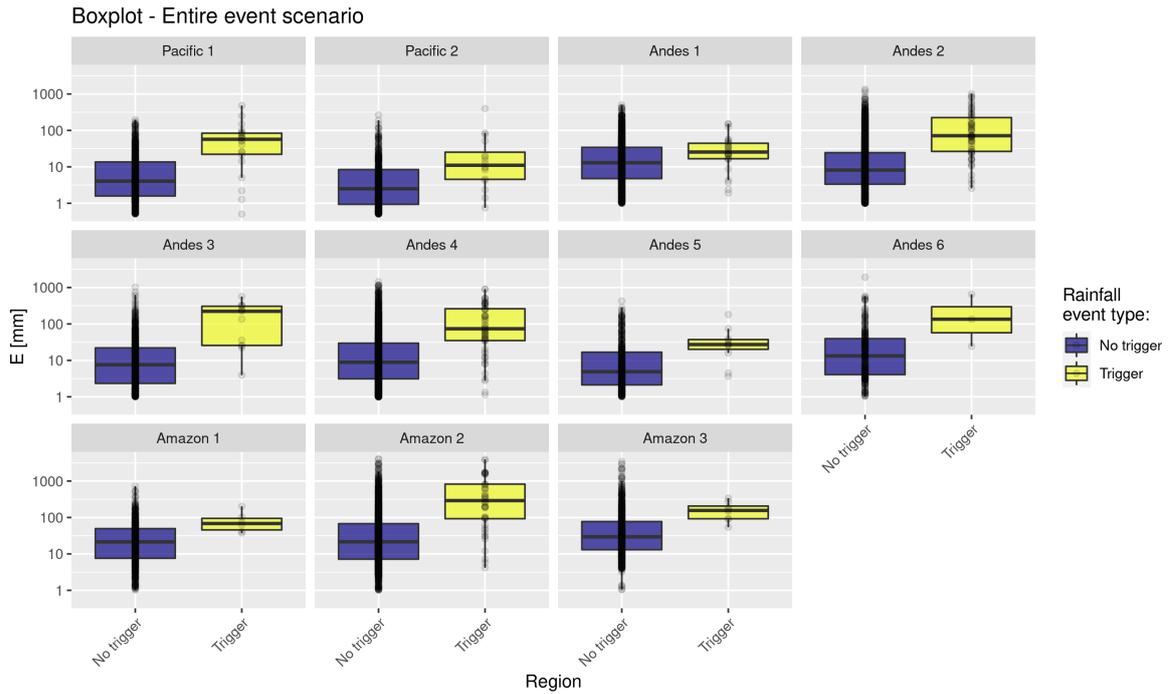


Figura 8: Diagrama de caja de la precipitación total desencadenante (amarilla) y no-desencadenante (azul) E para las 11 regiones establecidas en este estudio para Perú. Los gráficos de diagrama de caja incluyen valores atípicos y muestran el potencial predictivo de la variable E para separar los eventos desencadenantes y no-desencadenantes de MM poco profundos. Además, el gráfico muestra la variabilidad regional de los eventos desencadenantes de lluvia.

receptor(ROC) (Fawcett 2006), a partir del cual se usó la mínima distancia radial del test de clasificación perfecta ($TSS = 1$, con $s_e=1$ y $1-s_p=0$) para seleccionar los umbrales de variables individuales (e.g., Uwihirwe et al. 2020; Gariano et al. 2015; Postance et al. 2018); mientras que para los umbrales de tipo curva ($I_{max} - D$ y $I_{mean} - D$), el factor de escala a y el factor de forma b de la curva de tipo $V = a \cdot D^{-b}$ fueron configurados simultáneamente para maximizar el TSS (e.g., Leonarduzzi et al. 2017; Hirschberg et al. 2021), con una aproximación inicial de la curva con a =promedio de la variable V de los eventos de lluvia desencadenantes y $b = 0.5$. Esta maximización fue desarrollada de forma automática usando el Shuffled Complex Evolutionary Algorithm (SCEA-UA) (Duan et al. 1993), considerando a TSS como la función objetivo. Esta metodología fue ampliada para cada región del área de estudio encontrando diferentes umbrales bajo cada enfoque diferente.

IV. RESULTADOS Y DISCUSIONES

4.1. UMBRALES DE LLUVIA

Los umbrales calibrados, para las propiedades individuales de los eventos (I_{max} , E , D , I_{mean}) se muestran en la Tabla 2 y los umbrales curva ($I_{max} - D$ y $I_{mean} - D$) se muestran en la Tabla 3. Los resultados se presentan para dos escenarios, el primero describe los eventos de lluvia que incluyen el mismo día de ocurrencia del MM, denominado Entire Event (EE) y el segundo que solo incluye las condiciones antecedentes hasta un día antes de la ocurrencia de MM, denominado Antecedent Event (AE), esto dado que se tiene el interés de analizar eventos de MM bajo un enfoque que incluya la capacidad predictiva de las condiciones antecedentes y su influencia en la ocurrencia de eventos futuros para la operatividad de servicios de avisos de alerta temprana.

A partir de los resultados se observa que los umbrales con mejores rendimientos promedios para los eventos enteros fueron E (TSS = 0,59) para las Propiedades Individuales y $I_{mean} - D$ (TSS = 0,65) para las Curvas Combinadas. Según lo esperado, la integración de propiedades en curvas produjo un mejor rendimiento general en comparación con las propiedades de eventos individuales. Entre las dos curvas ($I_{max} - D$ y $I_{mean} - D$), la curva $I_{mean} - D$ tuvo el mejor rendimiento, con TSS=0,65 para calibración y TSS=0,42 para validación. En la Figura 10, se presentan los umbrales $I_{mean} - D$ para cada región.

Los resultados muestran que los componentes con menor rendimiento para la determinación de umbrales fueron la duración (D) tanto para el periodo de calibración como para la validación, seguida por la tasa promedio de lluvia (I_{mean}). En el caso de las curvas combinadas, se observa una menor diferencia en sus rendimientos, siendo la $I_{max} - D$ la de menor rendimiento. Estos umbrales no tienen una buena capacidad de discriminar eventos de lluvia desencadenantes y no-desencadenantes de MM.

Tabla 2: Umbrales de precipitación de variables independientes (Th: umbral, Rad: distancia radial mínima, Cal: calibración, Val: validación).

Scenario	Region	$E(mm)$			$I_{mean}(mm/d)$			$I_{max}(mm/d)$			$D(d)$						
		Th	Rad	TSS	Th	Rad	TSS	Th	Rad	TSS	Th	Rad	TSS				
			Cal	Val		Cal	Val		Cal	Val		Cal	Val				
	Pacific 1	21.16	0.38	0.66	0.20	5.62	0.33	0.54	0.60	10.11	0.43	0.68	0.58	8	0.49	0.56	-0.09
	Pacific 2	4.23	0.48	0.44	0.39	2.12	0.35	0.61	0.20	4.55	0.34	0.51	0.27	7	0.48	0.30	0.39
	Andes 1	16.15	0.49	0.39	0.12	6.20	0.41	0.43	0.18	11.84	0.43	0.38	0.23	2	0.60	0.19	-0.16
	Andes 2	23.92	0.29	0.58	0.41	5.17	0.30	0.51	0.28	8.59	0.29	0.58	0.47	8	0.31	0.54	0.33
	Andes 3	25.35	0.21	0.78	0.41	6.01	0.19	0.83	0.22	16.72	0.08	0.92	0.34	21	0.29	0.69	0.28
	Andes 4	38.85	0.35	0.51	0.61	6.17	0.42	0.45	0.61	8.44	0.40	0.43	0.69	9	0.43	0.45	0.33
	Andes 5	25.52	0.27	0.67	0.39	4.25	0.35	0.52	0.37	9.75	0.29	0.61	0.51	4	0.33	0.54	0.03
	Andes 6	24.32	0.36	0.64	0.66	4.05	0.57	0.40	0.44	5.56	0.54	0.45	0.46	6	0.31	0.68	0.69
	Amazon 1	37.20	0.36	0.64	-	12.68	0.26	0.74	-	20.73	0.34	0.66	-	3	0.57	0.29	-
	Amazon 2	92.77	0.31	0.57	0.52	8.88	0.44	0.41	0.34	16.15	0.41	0.46	0.38	5	0.40	0.51	0.37
	Amazon 3	53.99	0.32	0.68	0.66	11.14	0.59	0.41	0.39	17.74	0.48	0.52	0.55	12	0.50	0.44	-0.10
	Pacific 1	19.01	0.18	0.63	0.35	4.87	0.30	0.51	0.90	10.11	0.22	0.65	0.91	7	0.28	0.60	-0.11
	Pacific 2	18.60	0.27	0.53	-0.17	2.98	0.33	0.43	-0.21	10.56	0.35	0.43	-0.11	6	0.28	0.49	0.55
	Andes 1	7.57	0.65	0.14	0.42	5.70	0.47	0.35	0.63	7.57	0.54	0.30	0.57	7	0.73	0.04	-0.11
	Andes 2	40.03	0.28	0.59	0.42	5.26	0.29	0.51	0.30	9.74	0.27	0.59	0.33	7	0.32	0.54	0.38
	Andes 3	127.47	0.30	0.69	0.53	6.08	0.23	0.55	0.33	16.72	0.16	0.77	0.44	20	0.29	0.69	0.34
	Andes 4	31.73	0.33	0.53	0.57	5.77	0.37	0.51	0.59	8.44	0.35	0.50	0.60	9	0.40	0.46	0.24
	Andes 5	15.77	0.31	0.57	0.32	2.22	0.56	0.43	0.28	8.25	0.31	0.57	0.31	3	0.41	0.44	0.01
	Andes 6	18.76	0.43	0.55	0.60	3.75	0.60	-0.14	0.41	4.75	0.64	0.33	0.44	5	0.38	0.61	0.67
	Amazon 1	70.79	0.54	0.31	-	10.18	0.48	0.32	-	13.26	0.52	0.34	-	15	0.83	0.14	-
	Amazon 2	175.88	0.36	0.53	0.64	8.81	0.44	0.44	0.40	16.15	0.40	0.48	0.45	17	0.39	0.51	0.47
	Amazon 3	137.64	0.52	0.35	0.30	11.05	0.59	0.41	0.39	16.90	0.49	0.51	0.53	11	0.50	0.10	-0.10

Tabla 3: Umbrales de precipitación de dos variables (Th: umbral, Cal: calibración, Val: validación).

Scenario	Region	$I_{mean}(mm/d)-D(d)$				$I_{max}(mm)-D(d)$			
		Thresh		TSS		Thresh		TSS	
		a	b	Cal	Val	a	b	Cal	Val
Entire event	Pacific 1	11.55	-0.44	0.68	0.26	16.73	-0.17	0.71	0.28
	Pacific 2	2.10	-0.00	0.61	0.20	4.58	-0.00	0.51	0.27
	Andes 1	7.34	-0.10	0.44	0.19	20.97	-0.98	0.36	0.09
	Andes 2	14.28	-0.53	0.62	0.28	13.62	-0.17	0.64	0.34
	Andes 3	10.84	-0.25	0.89	0.33	16.77	-0.01	0.91	0.34
	Andes 4	25.69	-0.81	0.52	0.68	44.51	-0.66	0.49	0.70
	Andes 5	16.68	-0.77	0.66	0.39	15.08	-0.25	0.64	0.38
	Andes 6	16.93	-0.81	0.62	0.63	19.25	-0.69	0.56	0.63
	Amazon 1	14.25	-0.05	0.77	-	20.91	-0.02	0.66	-
	Amazon 2	42.06	-0.54	0.57	0.53	66.35	-0.56	0.57	0.48
	Amazon 3	36.74	-0.45	0.73	0.68	49.54	-0.42	0.73	0.70
	Antecedent event	Pacific 1	8.50	-0.50	0.68	0.84	18.60	-0.28	0.67
Pacific 2		14.85	-0.88	0.53	-0.17	25.15	-0.31	0.47	-0.08
Andes 1		6.45	-0.08	0.36	0.66	7.52	0.00	0.30	0.56
Andes 2		11.54	-0.39	0.65	0.39	19.37	-0.54	0.60	0.43
Andes 3		13.98	-0.26	0.80	0.48	16.01	-0.49	0.73	0.48
Andes 4		19.29	-0.72	0.56	0.66	34.81	-0.69	0.51	0.66
Andes 5		8.59	-0.63	0.53	0.41	23.61	-0.66	0.62	0.22
Andes 6		16.39	-0.92	0.55	0.59	18.54	-0.89	0.51	0.57
Amazon 1		51.63	-0.56	0.43	-	49.46	-0.17	0.37	-
Amazon 2		22.41	-0.42	0.53	0.51	33.70	-0.32	0.54	0.49
Amazon 3		16.81	-0.14	0.55	0.55	16.83	-0.01	0.50	0.53

4.2. IMPACTO DE LA REGIONALIZACIÓN

El área de estudio se regionalizó en 11 regiones con base en la información de precipitación máxima diaria. Los resultados estimados muestran que la variabilidad de las precipitaciones del Perú en las magnitudes de los umbrales para cada región se presenta en la Tabla 2. A nivel regional, el umbral de mejor desempeño de una sola variable, la precipitación acumulada E , con un promedio de 33 mm, varió de 4,23 mm (Pacific 2) a 92,77 mm (Amazon 2). I_{max} tiene un rango de 4,55 mm/d (Pacific 2) a 20,73 mm/d (Amazon 1) con un promedio de 11.83 mm/d. La región con mejor desempeño predictivo fue Andes 3, con un TSS de 0,8 para la media de los umbrales de variables individuales y un TSS de 0,89 para la media de la curva tipo umbral

Tabla 4: Número de eventos de movimientos en masa y umbrales más óptimos para una y dos variables en cada región (Th: umbral, SL: número de MM por región, Cal: calibración, Val: validación).

Region	SL total	SL Cal	SL Val	Best Th - 1 variable	TSS	Best Th - 2 variables	TSS
Pacific 1	46	43	3	I_{max}	0.68	$I_{max} - D$	0.71
Pacific 2	27	20	7	I_{mean}	0.61	$I_{mean} - D$	0.61
Andes 1	34	28	6	I_{mean}	0.43	$I_{mean} - D$	0.44
Andes 2	98	83	15	E and I_{mean}	0.58	$I_{max} - D$	0.64
Andes 3	17	10	7	I_{max}	0.92	$I_{max} - D$	0.91
Andes 4	65	54	11	E	0.51	$I_{mean} - D$	0.52
Andes 5	14	7	7	E	0.67	$I_{mean} - D$	0.66
Andes 6	4	3	1	D	0.68	$I_{mean} - D$	0.62
Amazon 1	6	6	-	I_{mean}	0.74	$I_{mean} - D$	0.77
Amazon 2	54	41	13	E	0.57	$I_{mean} - D$ and $I_{max} - D$	0.57
Amazon 3	12	10	2	E	0.68	$I_{mean} - D$ and $I_{max} - D$	0.73

en el escenario 2. El umbral con el mejor desempeño para esta región fue $I_{max}=16,72$ mm/d (TSS=0,92), que separó correctamente el 100 % de los eventos desencadenantes de lluvia y solo tuvo una tasa del 8 % de falsas alarmas. Del mismo modo, el $I_{max} - D$ (TSS=0,91) separó correctamente el 100 % de los eventos desencadenantes de lluvia y solo tuvo una tasa de falsas alarmas del 9 %. En la Tabla 4 se presenta un resumen de los mejores umbrales de variable única y curvos basados en el TSS para los resultados de calibración para cada región.

La regionalización logra obtener una mejor separación de las distribuciones desencadenantes y no desencadenantes. Los umbrales calibrados tuvieron un mejor desempeño general en las áreas de Andes 3 (TSS=0,83) en comparación con las regiones de Andes 1 (TSS=0,4), Andes 4 (TSS=0,47) y Amazonía 1 (TSS=0,5), que fueron las regiones con menor desempeño. Con respecto a las dos regiones del Pacífico, la región del Pacífico 1 (TSS=0,66) tuvo un mejor desempeño que la región del Pacífico 2 (TSS=0,51). En las regiones más húmedas de la Amazonía, la región Amazon 1 fue la de mejor desempeño, seguida de las regiones Amazon 3 y Amazon 2. Esta región amazónica y la región del Altiplano (Andes 6) fueron las regiones con menos eventos de calibración.

Los resultados no muestran que alguna vertiente (Pacific, Andes o Amazon) se resalte en poder separar eventos de lluvia desencadenantes de los no- desencadenantes, por el contrario, hay regiones con buen desempeño y regular desempeño a lo largo del Pacifico, Andes

y Amazonas. Las regiones Andes 6 (4 eventos de MM), Amazon 1 (6 eventos de MM) y Amazon 3 (12 eventos de MM) son las que tuvieron menor cantidad de eventos para la calibración y validación, las demás regiones incluyeron más de 10 eventos (Figura 9), resaltando las regiones Andes 2 (98 eventos), Andes 4 (65 eventos), Amazon 2 (54 eventos) y Pacific 1 (46 eventos de MM).

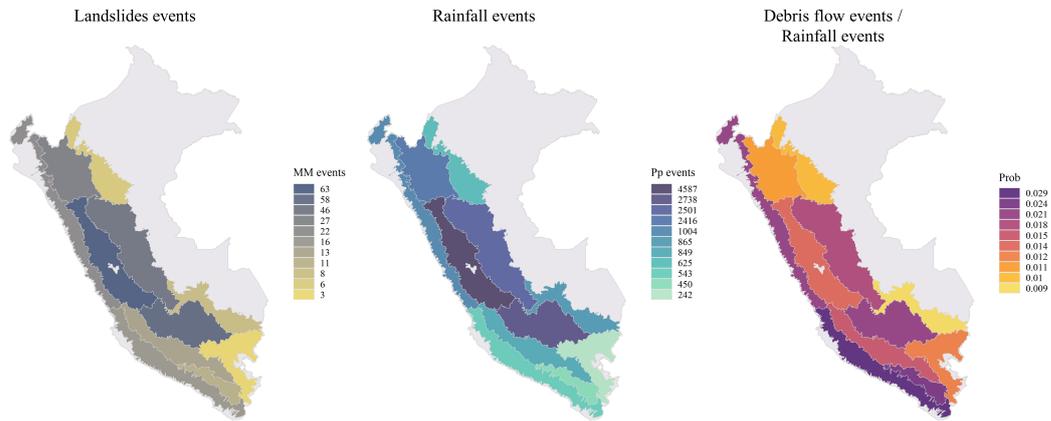


Figura 9: Distribución espacial a nivel regional del número de eventos de movimientos en masa (izquierda), eventos de lluvia (centro) y la probabilidad (derecha) de que los eventos de movimientos en masa sean desencadenados por eventos de lluvia.

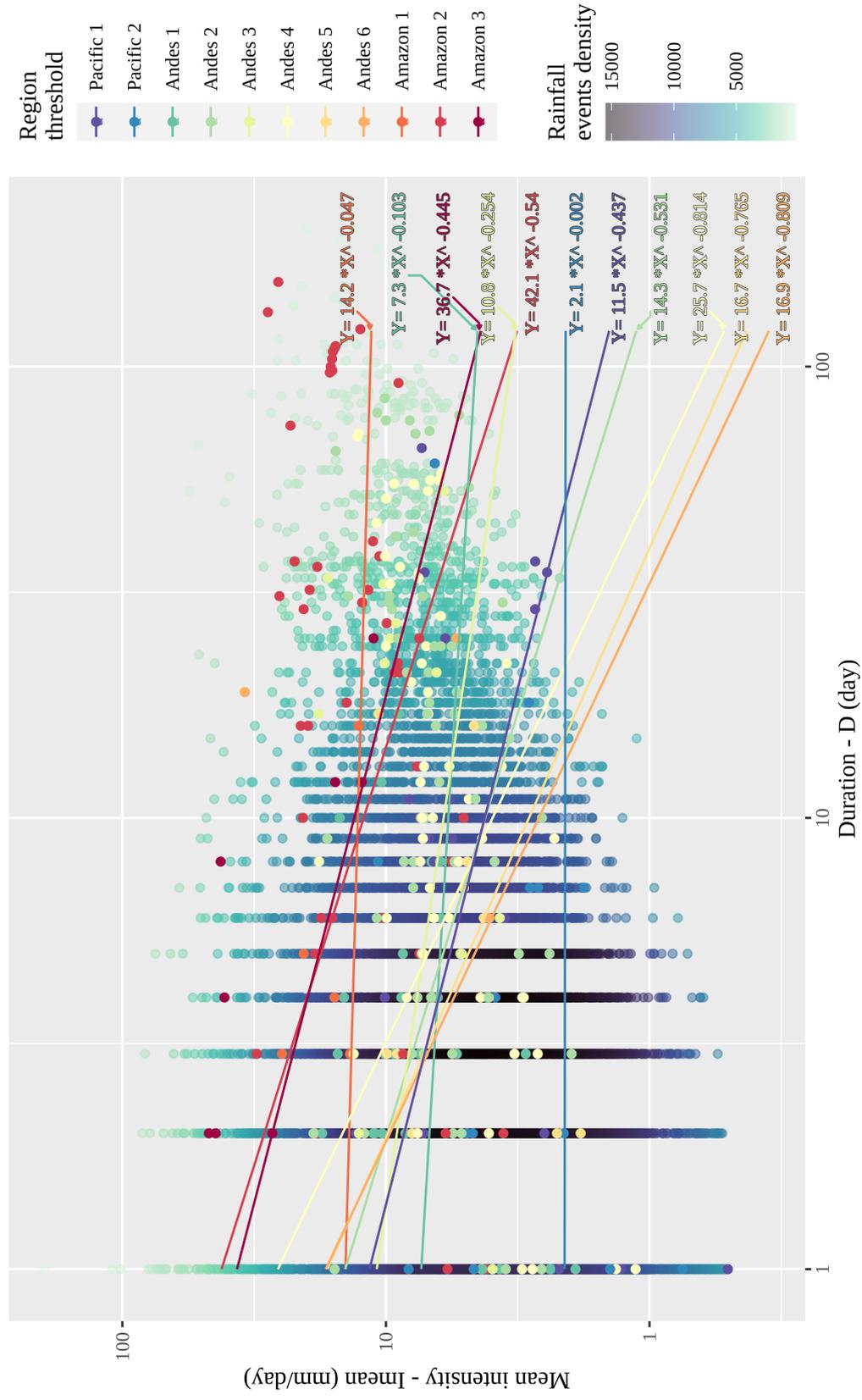


Figura 10: Gráficos de Intensidad-media vs. Duración ($I_{mean} - D$) con curvas de umbral regional en escala logarítmica. El fondo, representado por puntos de colores en una escala de verde-azul-negro, muestra la densidad de eventos de lluvia no-desencadenantes. Los eventos de lluvia desencadenantes se representaron con el mismo color de umbral regional.

4.3. EFECTO DE LAS CONDICIONES ANTECEDENTES

Es conocido que las condiciones antecedentes del terreno juegan un rol importante en la ocurrencia de lo MM y sobre todo en su magnitud, esta es la razón por la cual se analizó este escenario, que incluya la separación de eventos de lluvia que solo consideran la tasa de lluvia hasta un día previo del día de ocurrencia. Se observa que, en la fase de calibración, el escenario de condiciones antecedentes (antecedent event) obtuvo rendimientos menores que el escenario de eventos enteros. Sin embargo, en la etapa de validación para el año 2020, se observó que para algunos umbrales de forma aislada su rendimiento fue superior, por ejemplo para la región Pacific 1, los umbrales I_{max} y I_{mean} obtuvieron rendimientos superiores que el escenario de evento entero (incluyendo la tasa de precipitación del día del evento de MM), esto significa que días previos al día de ocurrencia hubo un día con una lluvia intensa mayor al del día de ocurrencia que permite separar ese evento como evento desencadenante y capaz de alterar la tasa de lluvia promedio asociada a dicho evento.

4.4. EVALUACIÓN DEL DESEMPEÑO DE UMBRALES

La validación se llevó a cabo para los eventos ocurridos en 2020 simulando la operatividad de los umbrales calibrados en un sistema de alerta regional. La región Amazon 1 no contempló eventos de MM para ese año, por lo que no entró en esta evaluación. La validación muestra que en la mayoría de los umbrales hubo una clara disminución de la magnitud para este año (Tablas 2 y 3). Por ejemplo, el umbral I_{max} , que obtuvo el mejor desempeño en calibración, disminuyó para este período, excepto para las regiones Andes 4, Andes 6 y Amazon 3, que mejoraron en esta validación; esto significa que el umbral permitió la separación de los eventos de lluvia de 2020 mejor de lo esperado en la calibración.

La variable D se confirma que es, por si sola, un mal umbral separador de eventos de lluvia desencadenantes de los que no, incluso teniendo rendimientos negativos (Pacific 1, Andes 1 y Amazon 3), esta negatividad está asociada a que la sensibilidad (predicción correcta de movimientos en masa) del modelo para estas regiones es 0 es decir que el umbral estimado en

la calibración no ha sido capaz de separar los eventos de lluvias. Sin embargo, esta variable muestra que podemos asociar los movimientos en masa a eventos de lluvia continua con una duración antecedente de 8 días.

En cuanto a la variabilidad de los umbrales (Fig. 10), podemos explicarlo principalmente a la climatología de las precipitaciones en el Perú. Se puede observar que las magnitudes tienen una relación con la distribución espacial de las precipitaciones en el Perú, es decir, umbrales bajos relacionados con precipitaciones de menor magnitud en las zonas áridas de la parte occidental del Perú (Pacífico), umbrales intermedios relacionados con el aumento de la magnitud de las precipitaciones en la parte media o región montañosa (Andes) y los umbrales más altos relacionados con las regiones húmedas (Amazonía). Sin embargo, las regiones Andes 1, Andes 3 y Andes 6 no tienen esta relación, por lo que esta discusión no es concluyente y se considera que está relacionada con datos limitados, por lo que se sugiere que esta variabilidad se confirme en futuras investigaciones que incluyan datos de eventos de MM superficiales.

Con respecto al periodo de validación, se utilizaron 61 eventos en total, por lo que el estadístico TSS es más sensible, principalmente debido a la alta sensibilidad del modelo (es decir, la probabilidad de predecir correctamente los eventos de lluvia desencadenantes de MM), mientras que la especificidad se mantuvo aproximadamente igual (es decir, la probabilidad de predecir correctamente los eventos de lluvias no-desencadenantes de MM). Este efecto apunta a la importancia de obtener inventarios de MM amplios.

4.5. DISCUSIONES

En esta investigación, se determinaron umbrales de precipitación que permiten separar eventos de lluvias desencadenantes y no-desencadenantes de movimientos en masa para dos escenarios, a partir de parámetros de eventos de lluvias asociados a MM, este tipo de análisis ya se desarrollaron objetivamente en estudios previos (Peruccacci et al. 2017, 2012; Segoni et al. 2014; Rosi et al. 2012; Leonarduzzi et al. 2017; Uwihirwe et al. 2020; Abraham et al. 2019). Pero, este trabajo es la primera aproximación de umbrales regionales a escala del

Perú, lo cual servirá de punto de partida y referencia para continuar el desarrollo de este tipo de investigaciones en el Perú.

Los umbrales estimados se muestran en la Tabla 2 para variables independientes y en la Tabla 3 para umbrales de curva. Los umbrales con el mejor desempeño fueron E para las variables individuales y $I_{mean} - D$ para umbrales tipo curva potencial. La variable que tuvo el rendimiento más bajo fue la duración del evento, D , por lo que no debe usarse de forma independiente, sino combinada con otras variables de evento. Sin embargo, nos permite asociar los eventos de movimientos en masa con las condiciones de lluvia antecedentes de los últimos 8 días, una asociación que puede usarse para futuras investigaciones.

En cuanto a los umbrales tipo curva, el TSS tuvo un desempeño ligera mejor, todos superando 0,5 en la calibración de la $I_{mean} - D$ (el umbral con el mejor rendimiento para umbrales tipo curva), excepto para Andes 1. La selección de estos umbrales se basa en un modelo de optimización (maximizando el TSS), a través del cual se busca una alta tasa de detección de MM (sensibilidad), manteniendo, en la medida de lo posible, una baja tasa de detección de falsos positivos (especificidad). Sin embargo, se observó que para buscar esta optimización, se sacrifica la detección de MM (dando falsos negativos), aunque se reducen las falsas alarmas, y este es un dilema en términos de sistemas de alerta, pero TSS es un buen equilibrio entre la detección de MM y falsas alarmas.

La región Pacific 1 se ve constantemente afectada por MM poco profundos y también contiene la mayoría de las ciudades con mayor densidad de población en Perú, por lo que su evaluación es altamente relevante. En esta región, se observó que el I_{max} (TSS=0,68) y $I_{max} - D$ (TSS=0,71) fueron los mejores umbrales para el escenario de evento-entero (entire event), lo que indica que las cuencas en esta región son altamente susceptibles a eventos de máxima intensidad. Mientras que los umbrales $I_{max} - D$ (TSS=0,65) y $I_{mean} - D$ (TSS=0,68) fueron los mejores umbrales para el escenario de evento-antecedente (antecedent event). La variable I_{max} tuvo el mejor desempeño, lo que sugiere que las lluvias de alta intensidad tienen un alto impacto condicionante en el desarrollo de MM. Con respecto al hecho de que los rendimientos de validación en el escenario anterior fueron mayores en los rendimientos de calibración, puede deberse a que el conjunto de validación es demasiado pequeño.

Es necesario indicar que esta investigación ha considerado los datos de precipitación y de eventos de MM a escala nacional. Sin embargo, el entendimiento de los umbrales a una escala local, va a estar supeditado al conocimiento de la geomorfología, el tipo de suelo y sus características, por lo que se sugiere que en estudios a escala local se realice un análisis exhaustivo de muestreo en campo para obtener los parámetros de fricción y variables reológicas que caractericen los MM (e.g., Millan y Chavarri-Velarde 2019). Por lo que los resultados presentados en esta tesis que es a una escala regional/nacional, se sugiere adaptar los resultados de esta tesis a unas escalas más locales.

La regionalización era necesaria dada la alta variabilidad climática en el Perú, evidenciada por las diferencias de magnitud entre los umbrales. Esta regionalización nos ayudó a observar las regiones del Perú donde hay mayor ocurrencia de MM y la respuesta a este tipo de umbral diario. Por ejemplo, observamos que los Andes 2 (la región con el mayor número de eventos) tuvo una mejor respuesta para I_{max} en el proceso de calibración y validación. Peruccacci et al. (2012) encontraron que el número de eventos debe ser mayor que 175 para limitar la incertidumbre relativa por debajo del 10 %, pero esta cifra puede cambiar para un conjunto de datos diferente. Con base en esto, se observa que solo cuatro regiones (Andes 2, Andes 4, Pacific 1 y Amazon 2) tienen una serie de eventos que son aceptables. Las otras regiones tienen una mayor fuente de incertidumbre debido a la cantidad de datos. En la Tabla 4 se presenta un resumen del número de eventos de MM poco profundos utilizados para la investigación y los umbrales con los mejores rendimientos por región.

La evaluación del desempeño de los umbrales se llevó a cabo a través de la validación con los eventos de 2020. Sin embargo, se observó que los rendimientos disminuyeron, lo que puede deberse a que, en el año 2020, no hubo eventos de lluvias extremas como en otros años, y el número de MM fue menor que en otros años. Incluso la región Amazon 1 no tenía registro de eventos de activación; por lo tanto, podemos afirmar que el bajo rendimiento se debió a que los umbrales no representan eventos de deslizamiento con magnitud de bajo impacto, y esto está asociado con uno de los enfoques del modelo, que es reducir la tasa de falsas alarmas.

La metodología de calibración/validación, basada en tomar 1 año de observaciones para el conjunto de validación, que se utilizó en otros trabajos de investigación (e.g., Kirschbaum

Tabla 5: Comparación de TSS entre enfoques de validación (1-año: selección de 1 año de la información disponible y Aleatoria: selección aleatoria).

TSS comparison between validation approaches									
Procedure	<i>I_{mean} - D</i>			<i>I_{max} - D</i>			<i>E - D</i>		
	1 year	<i>Random</i>	ΔTSS	1year	<i>Random</i>	ΔTSS	1year	<i>Random</i>	ΔTSS
Calibration	0.65	0.61	-0.04	0.62	0.59	-0.03	0.59	0.58	-0.01
Validation	0.42	0.50	0.08	0.42	0.45	0.03	0.43	0.40	-0.02

et al. 2015b; Dikshit et al. 2019), es bastante corto, y existe el riesgo de sobreinterpretación. Por esta razón, este método se comparó con otro método de validación basado en una selección aleatoria del conjunto de datos (e.g., Brunetti et al. 2021; Gariano et al. 2020). De acuerdo con este método, los datos se dividieron aleatoriamente en 70% para calibración y 30% para validación. La comparación de ambos enfoques de validación se muestra en la Tabla 5. En este sentido, la comparación entre los métodos de validación no indicó cambios significativos entre cada método. Los resultados son muy similares, probablemente porque el tamaño de los datos no es lo suficientemente grande como para observar las variaciones entre los métodos. Se recomienda altamente que las investigaciones futuras se centren en la expansión del conjunto de datos y luego comparen la eficiencia del método de validación.

Todavía hay muchas limitaciones para el estudio del umbrales de lluvia a escala regional en Perú. Principalmente, los registros cortos de MM no son suficientes para limitar la incertidumbre en la definición del umbral (Peruccacci et al. 2012; Hirschberg et al. 2021). Otra fuente importante de incertidumbre fue el uso de la resolución de datos de lluvia temporal gruesa que causan una subestimación sistemática de los umbrales (Marra 2019; Gariano et al. 2020). Otra es la resolución espacial de los datos de lluvia porque una celda de 10 km puede cubrir varias quebradas o cuencas pequeñas. Y finalmente, la regionalización puede no ser suficientemente representativa de la alta variabilidad de las variables descriptoras de MM. Estas limitaciones deben tenerse en cuenta en futuras investigaciones.

V. CONCLUSIONES

Este estudio es la primera aproximación de umbrales regionales desencadenantes de movimientos en masa a escala del Perú para el análisis de la interrelación entre las lluvias y su efecto desencadenador de movimientos en masa. Como conclusión principal de la tesis es que es factible la estimación de umbrales de lluvia desencadenante de movimientos en masa a partir de la información de precipitación grillada diaria y eventos de movimientos en masa, de una forma objetiva bajo un enfoque metodológico empírico-estadístico. La calibración fue realizada mediante minimización de la distancia radial al punto óptimo de la Curva ROC y la maximización del estadístico TSS, mediante la cual se confirma su capacidad para la selección de umbrales con mejor rendimiento predictivo. El mejor rendimiento predictivo se obtuvo mediante la curva umbral intensidad media-duración $I_{mean} - D$, seguida de la intensidad diaria acumulada E . La duración del evento de forma independiente tiene un poder predictivo muy bajo.

Las conclusiones abordadas son:

- Se caracterizó las lluvias desencadenantes de movimientos en masa mediante la estimación de umbrales como medida de la interrelación de las lluvias y los movimientos en masa. Se evaluaron 4 variables independientes I_{max} , E , D , I_{mean} y 3 curvas potenciales $I_{max} - D$, $E - D$, $I_{mean} - D$.
- El rendimiento de los umbrales calibrados tuvo una alta variabilidad entre regiones. Estas diferencias en el desempeño están asociadas con la alta variabilidad de los eventos de lluvia en cada región, donde los mejores resultados ocurren en áreas donde es más fácil separar los eventos de lluvia desencadenantes y no-desencadenantes para la ocurrencia de MM superficiales (por ejemplo, regiones Andes 3, Amazon 1, Amazon 3 y Pacific 1). Sin embargo, en otras regiones, esta separación entre eventos de

lluvia es más compleja de llevar a cabo, ya que hay muchos eventos de lluvia no-desencadenantes con altas magnitudes, lo que se refleja en rendimientos más bajos (por ejemplo, Andes 1, Andes 4 y Amazon 2). Por lo tanto, la regionalización muestra que existen regiones donde el componente climático tuvo más predominio en la ocurrencia de movimientos en masa poco profundos en comparación con otras regiones donde la litología podría tener más influencia en la ocurrencia de MM superficiales que solo las lluvias. Los estudios futuros pueden explorar la regionalización basada en la litología.

- A través de las bases de datos de precipitaciones y MM, es posible generar umbrales diarios de lluvia para la ocurrencia de MM superficiales. Sin embargo, las incertidumbres asociadas con estas bases de datos son la principal fuente de incertidumbre para los umbrales. Los pocos MM registrados hicieron que el rendimiento de validación fuera muy sensible a los pocos datos (es decir, un solo evento podría conducir a un valor alto o bajo de las estadísticas de rendimiento). Por lo tanto, solo cuatro regiones (Andes 2, Andes 4, Pacific 1 y Amazon 2) tienen suficientes eventos para limitar estas incertidumbres. A pesar de estas incertidumbres, el marco metodológico establecido de este trabajo permite actualizaciones sistemáticas de los umbrales a medida que crecen los registros.

VI. RECOMENDACIONES

Los resultados de este trabajo demuestran el potencial de los umbrales de lluvia basados en las características de los eventos de lluvia asociados con MM para su implementación en el monitoreo de MM en Perú. La labor futura debería centrarse en tres perspectivas principales basadas en las limitaciones y fuentes de incertidumbre: i) mejora de la resolución espacio-temporal de las precipitaciones cuadrículadas; ii) mejora de la discretización espacial de las regiones en las que se produce el mayor número de movimientos en masa, que depende en primer lugar de la mejora de la resolución espacio-temporal de las precipitaciones; y iii) la asimilación de nuevas bases de datos sobre movimientos en masa para mejorar la certeza de los umbrales y reducir su sensibilidad.

Asimismo, las nuevas perspectivas de investigación y recomendaciones del presente trabajo proponen lo siguiente:

- Es necesario priorizar las regiones con alta ocurrencia de movimientos en masa determinadas en este estudio (Pacific 1, Amazon 2, Andes 2 y Andes 4).
- Implementar enfoques empírico-estadísticos en estudios de análisis de eventos extremos para la alerta temprana a escala local.
- Hacer uso de una mayor cantidad de productos de lluvia, no solo provenientes de información satelital y/o reanálisis, si no de pluviómetros. De igual manera, adicionar otros métodos de estimación de umbrales.
- Implementar datos de lluvia a escala horaria o subhoraria para disminuir a subestimación de los umbrales.
- Aplicar el modelo con mayor cantidad de eventos y realizar un análisis de incertidumbre de los umbrales.

- Analizar otras variables como la humedad de suelo (u otros) en la generación de umbrales de lluvia.

VII. REFERENCIAS BIBLIOGRÁFICAS

- Abraham, M. T., Pothuraju, D., y Satyam, N. 2019. Rainfall thresholds for prediction of landslides in Idukki, India: An empirical approach. *Water (Switzerland)*, 11(10):1–16.
- Abraham, M. T., Satyam, N., Pradhan, B., y Alamri, A. M. 2020. Forecasting of Landslides Using Rainfall Severity and Soil Wetness: A Probabilistic Approach for Darjeeling Himalayas. *Water*, 12(3):804.
- Augusto Filho, O., da Silva Júnior, P. S., y Eiras, C. G. S. 2020. Simple rainfall indices for forecasting hazardous events of hydrologic and geologic nature. *Natural Hazards*, (0123456789).
- Aybar, C., Fernández, C., Huerta, A., Lavado, W., Vega, F., y Felipe-Obando, O. 2020. Construction of a high-resolution gridded rainfall dataset for Peru from 1981 to the present day. *Hydrological Sciences Journal*, 65(5):770–785.
- Berti, M., Bernard, M., Gregoretti, C., y Simoni, A. 2020. Physical Interpretation of Rainfall Thresholds or Runoff-Generated Debris Flows. *Journal of Geophysical Research: Earth Surface*, 125(6).
- Brunetti, M. T., Melillo, M., Gariano, S. L., Ciabatta, L., Brocca, L., Amarnath, G., y Perrucci, S. 2021. Satellite rainfall products outperform ground observations for landslide prediction in India. *Hydrology and Earth System Sciences*, 25(6):3267–3279.
- Caine, N. 1980. The rainfall intensity-duration control of shallow landslides and debris flows. *Geografiska annaler A*, 62(1-2):23–27.
- Castro, D. 2022. Umbral de precipitación para deslizamientos de tierra en la cuenca del río rímac. *Universidad Nacional Agraria La Molina. Escuela de Posgrado. Maestría en Gestión Integral de Cuencas Hidrográficas*.

- Chae, B. G., Park, H. J., Catani, F., Simoni, A., y Berti, M. 2017. Landslide prediction, monitoring and early warning: a concise review of state-of-the-art. *Geosciences Journal*, 21(6):1033–1070.
- Cruden, D. M. 1991. A simple definition of a landslide. *Bulletin of the International Association of Engineering Geology - Bulletin de l'Association Internationale de Géologie de l'Ingénieur*, 43(1):27–29.
- Dai, A. 2006. Precipitation Characteristics in Eighteen Coupled Climate Models. *Journal of Climate*, 19(18):4605–4630.
- Dai, A., Lin, X., y Hsu, K.-L. 2007. The frequency, intensity, and diurnal cycle of precipitation in surface and satellite observations over low- and mid-latitudes. *Climate Dynamics*, 29:727–744.
- David, C. H., Maidment, D. R., Niu, G. Y., Yang, Z. L., Habets, F., y Eijkhout, V. 2011. River network routing on the NHDPlus dataset. *Journal of Hydrometeorology*, 12(5):913–934.
- Dikshit, A., Sarkar, R., Pradhan, B., Acharya, S., y Dorji, K. 2019. Estimating Rainfall Thresholds for Landslide Occurrence in the Bhutan Himalayas. *Water*, 11(8):1616.
- Duan, Q., Gupta, V., y Sorooshian, S. 1993. Shuffled complex evolution approach for effective and efficient global minimization. *J Optim Theory Appl*, 76(76):501–521.
- Fawcett, T. 2006. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874.
- Frattoni, P., Crosta, G., y Carrara, A. 2010. Techniques for evaluating the performance of landslide susceptibility models. *Engineering Geology*, 111(1-4):62–72.
- Gariano, S. L., Brunetti, M. T., Iovine, G., Melillo, M., Peruccacci, S., Terranova, O., Venari, C., y Guzzetti, F. 2015. Calibration and validation of rainfall thresholds for shallow landslide forecasting in Sicily, southern Italy. *Geomorphology*, 228:653–665.
- Gariano, S. L., Melillo, M., Peruccacci, S., y Brunetti, M. T. 2020. How much does the rainfall temporal resolution affect rainfall thresholds for landslide triggering? *Natural Hazards*, 100(2):655–670.

- Grupo de estándares para movimientos en masa-GEMMA 2007. *Movimientos en masa en la región Andina: Una guía para la evaluación de amenazas*. 2007. Number 2.
- Guzzetti, F., Peruccacci, S., Rossi, M., y Stark, C. P. 2007. Rainfall thresholds for the initiation of landslides in central and southern Europe. *Meteorology and Atmospheric Physics*, 98(3-4):239–267.
- Guzzetti, F., Peruccacci, S., Rossi, M., y Stark, C. P. 2008. The rainfall intensity-duration control of shallow landslides and debris flows: An update. *Landslides*, 5(1):3–17.
- Han, Z., Chen, G., Li, Y., Xu, L., y Fan, F. 2016. Geotechnical Hazards from Large Earthquakes and Heavy Rainfalls. *Geotechnical Hazards from Large Earthquakes and Heavy Rainfalls*.
- Hanssen, A. W. y Kuipers, W. J. A. 1965. *On the relationship between the frequency of rain and various meteorological parameters*.
- Highland, L. 2004. Fact Sheet 2004-3072: Landslide Types and Processes. Technical report.
- Hirschberg, J., Badoux, A., McArdell, B. W., Leonarduzzi, E., y Molnar, P. 2021. Evaluating methods for debris-flow prediction based on rainfall in an Alpine catchment. *Natural Hazards and Earth System Sciences*, 21(9):2773–2789.
- Hosking, J. R. M. y Wallis, J. R. 1997. *Regional Frequency Analysis*. Cambridge University Press.
- Huerta, A., Aybar, C., y Lavado-Casimiro, W. 2018. Pisco temperatura versión 1.1 (piscot v1. 1). *Lima, Peru: National Meteorology and Hydrology Service of Peru (SENAMHI)*.
- Huerta, A., Bonnesoeur, V., Cuadros-Adriazola, J., Gutierrez, L., Ochoa-Tocachi, B. F., Román-Dañobeytia, F., y Lavado-Casimiro, W. 2022. PISCOeo_pm, a reference evapotranspiration gridded database based on FAO Penman-Monteith in Peru. *Scientific Data*, 9(1):1–18.
- Huerta Julca, A. 2020. Vulnerabilidad de la disponibilidad de los recursos hidricos en el Perú frente al cambio climático: Analisis probabilistico de budyko.
- Hungr, O., Leroueil, S., y Picarelli, L. 2014. The Varnes classification of landslide types, an update. *Landslides*, 11(2):167–194.

- Huynh, Q. T., Nguyen, U. D., Irazabal, L. B., Ghassemian, N., y Tran, B. Q. 2015. Optimization of an accelerometer and gyroscope-based fall detection algorithm. *Journal of Sensors*.
- INDECI 2019. *Compendio Estadístico Del INDECI 2019. En La Preparación, respuesta y rehabilitación de la GRD*. Lima.
- Kirschbaum, D. y Stanley, T. 2018. Satellite-Based Assessment of Rainfall-Triggered Landslide Hazard for Situational Awareness. *Earth's Future*, 6(3):505–523.
- Kirschbaum, D., Stanley, T., y Zhou, Y. 2015a. Spatial and temporal analysis of a global landslide catalog. *Geomorphology*, 249:4–15.
- Kirschbaum, D. B., Stanley, T., y Simmons, J. 2015b. A dynamic landslide hazard assessment system for Central America and Hispaniola. *Natural Hazards and Earth System Sciences*, 15(10):2257–2272.
- Leonarduzzi, E. y Molnar, P. 2020. Deriving rainfall thresholds for landsliding at the regional scale: daily and hourly resolutions, normalisation, and antecedent rainfall. *Natural Hazards and Earth System Sciences*, 20(11):2905–2919.
- Leonarduzzi, E., Molnar, P., y McArdell, B. W. 2017. Predictive performance of rainfall thresholds for shallow landslides in Switzerland from gridded daily data. *Water Resources Research*, 53(8):6612–6625.
- Llauca, H., Lavado-Casimiro, W., Montesinos, C., Santini, W., y Rau, P. 2021. PISCO_HyM_GR2M: A model of monthly water balance in Peru (1981–2020). *Water (Switzerland)*, 13(8):1–19.
- Llauca, H., Leon, K., y Lavado-Casimiro, W. 2023. Construction of a daily streamflow dataset for Peru using a similarity-based regionalization approach and a hybrid hydrological modeling framework. *Journal of Hydrology: Regional Studies*, 47:101381.
- Marra, F. 2019. Rainfall thresholds for landslide occurrence: systematic underestimation using coarse temporal resolution data. *Natural Hazards*, 95(3):883–890.
- Millan, C. 2020. Sistema de Monitoreo de movimientos en masa potenciales generados por lluvias intensas del SENAMHI (SILVIA). In Lavado-Casimiro, W., editor, *Estudios Hidrológicos del SENAMHI Resúmenes Ejecutivos - 2020*, pages 17–21. Lima, i edition.

- Millan, C. y Chavarri-Velarde, E. 2019. DEBRIS FLOW MODELING FOR RISK ANALYSIS IN LOS CONDORES CATCHMENT, CHACLACAYO, LIMA, PERU. 38:3525–3533.
- Millan, C., Lavado, W., Vega, F., Felipe, O., Acuña, J., y Takahashi, K. 2021. SILVIA : An operational system to monitoring landslides forced by heavy precipitations at national scale in Peru. In *EGU General Assembly 20220*, volume 1, page 10970.
- Mirus, B. B., Morphew, M. D., y Smith, J. B. 2018. Developing hydro-meteorological thresholds for shallow landslide initiation and early warning. *Water*, 10(9).
- Peirce, C. 1884. The Numerical Measure of Success in Predictions. *Science*, pages 453–454.
- Peruccacci, S., Brunetti, M. T., Gariano, S. L., Melillo, M., Rossi, M., y Guzzetti, F. 2017. Rainfall thresholds for possible landslide occurrence in Italy. *Geomorphology*, 290(April):39–57.
- Peruccacci, S., Brunetti, M. T., Luciani, S., Vennari, C., y 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(10):927–930.
- Postance, B., Hillier, J., Dijkstra, T., y Dixon, N. 2018. Comparing threshold definition techniques for rainfall-induced landslides: A national assessment using radar rainfall. *Earth Surface Processes and Landforms*, 43(2):553–560.
- Prenner, D., Kaitna, R., Mostbauer, K., y Hrachowitz, M. 2018. The Value of Using Multiple Hydrometeorological Variables to Predict Temporal Debris Flow Susceptibility in an Alpine Environment. *Water Resources Research*, 54(9):6822–6843.
- Qiao, X., Nelson, E. J., Ames, D. P., Li, Z., David, C. H., Williams, G. P., Roberts, W., Sánchez Lozano, J. L., Edwards, C., Souffront, M., y Matin, M. A. 2019. A systems approach to routing global gridded runoff through local high-resolution stream networks for flood early warning systems. *Environmental Modelling and Software*, 120(July).
- Rosi, A., Segoni, S., Catani, F., y Casagli, N. 2012. Statistical and environmental analyses for the definition of a regional rainfall threshold system for landslide triggering in Tuscany (Italy). *Journal of Geographical Sciences*, 22(4):617–629.

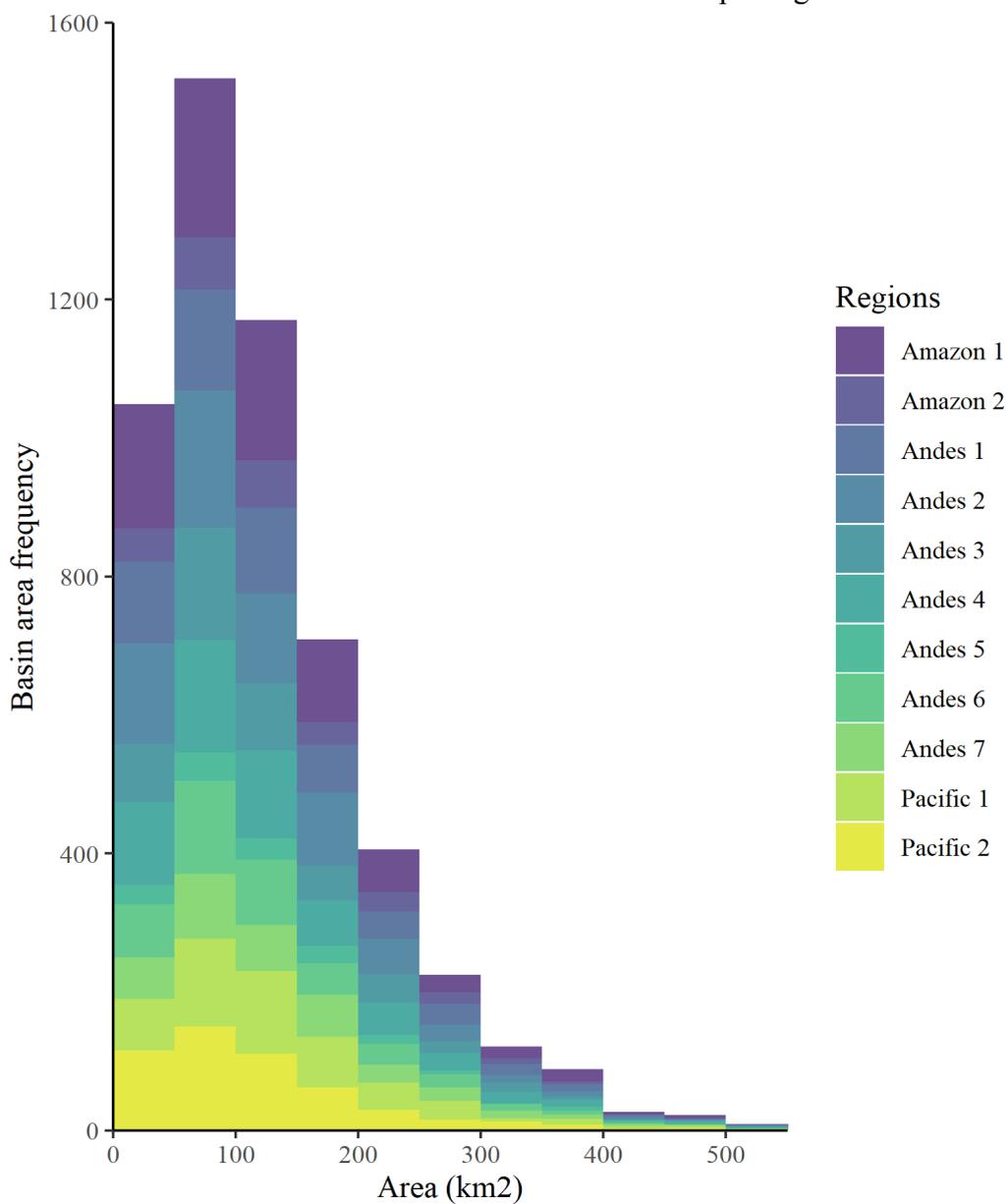
- Segoni, S., Piciullo, L., y Gariano, S. L. 2018. A review of the recent literature on rainfall thresholds for landslide occurrence. *Landslides*, 15(8):1483–1501.
- Segoni, S., Rosi, A., Rossi, G., Catani, F., y Casagli, N. 2014. Analysing the relationship between rainfalls and landslides to define a mosaic of triggering thresholds for regional-scale warning systems. *Natural Hazards and Earth System Sciences*, 14(9):2637–2648.
- Shen, Z., Yong, B., Gourley, J. J., y Qi, W. 2021. Real-time bias adjustment for satellite-based precipitation estimates over Mainland China. *Journal of Hydrology*, 596.
- Souffront Alcantara, M. A., Nelson, E. J., Shakya, K., Edwards, C., Roberts, W., Krewson, C., Ames, D. P., Jones, N. L., y Gutierrez, A. 2019. Hydrologic Modeling as a Service (HMaaS): A New Approach to Address Hydroinformatic Challenges in Developing Countries. *Frontiers in Environmental Science*, 7(October).
- Staley, D. M., Kean, J. W., Cannon, S. H., Schmidt, K. M., y Laber, J. L. 2013. Objective definition of rainfall intensity-duration thresholds for the initiation of post-fire debris flows in southern California. *Landslides*, 10(5):547–562.
- Tang, H., McGuire, L. A., Rengers, F. K., Kean, J. W., Staley, D. M., y Smith, J. B. 2019. Developing and Testing Physically Based Triggering Thresholds for Runoff-Generated Debris Flows. *Geophysical Research Letters*, 46(15):8830–8839.
- Tian, Y., Peters-Lidard, C. D., Choudhury, B. J., y Garcia, M. 2007. Multitemporal Analysis of TRMM-Based Satellite Precipitation Products for Land Data Assimilation Applications. *Hydrometeorology*, 8(6):1165–1183.
- Uwihirwe, J., Hrachowitz, M., y Bogaard, T. A. 2020. Landslide precipitation thresholds in Rwanda. *Landslides*, 17(10):2469–2481.
- van Osnabrugge, B., Weerts, A. H., y Uijlenhoet, R. 2017. genRE: A Method to Extend Gridded Precipitation Climatology Data Sets in Near Real-Time for Hydrological Forecasting Purposes. *Water Resources Research*, 53(11):9284–9303.
- Varnes, D. 1978. SLOPE MOVEMENT TYPES AND PROCESSES. Technical Report Special Report 176, Transportation Research Board, Washington, DC United States 20001.

- Villacorta, S., Fidel, L., y Zavala Carrión, B. 2012. Mapa de susceptibilidad por movimientos en masa del Perú. *Revista de la Asociación Geológica Argentina*, 69(3):393–399.
- Yong, B., Ren, L.-L., Hong, Y., Wang, J.-H., Gourley, J. J., Jiang, S.-H., Chen, X., y Wang, W. 2010. Hydrologic evaluation of Multisatellite Precipitation Analysis standard precipitation products in basins beyond its inclined latitude band: A case study in Laohahe basin, China. *Water Resources Research*.
- Youden, W. J. 1950. Index for rating diagnostic tests. *Cancer*, 3(1):32–35.
- Yupanqui, R. S., Lavado-Casimiro, W. S., y Felipe-Obando, O. S. 2017. Regionalización de las precipitaciones máximas en el Perú.

VIII. ANEXOS

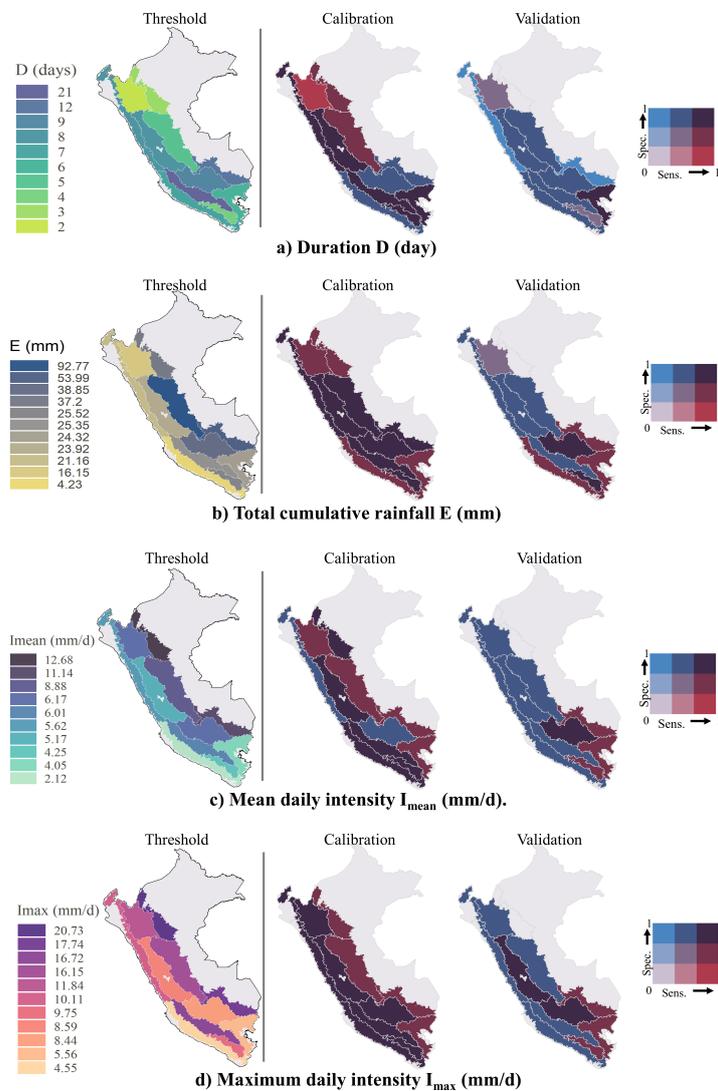
Anexo 1: Información de las cuencas de nivel 12 del servicio GEOGloWS ECMWF

Frecuencia del área de cuencas de nivel 12 por región.



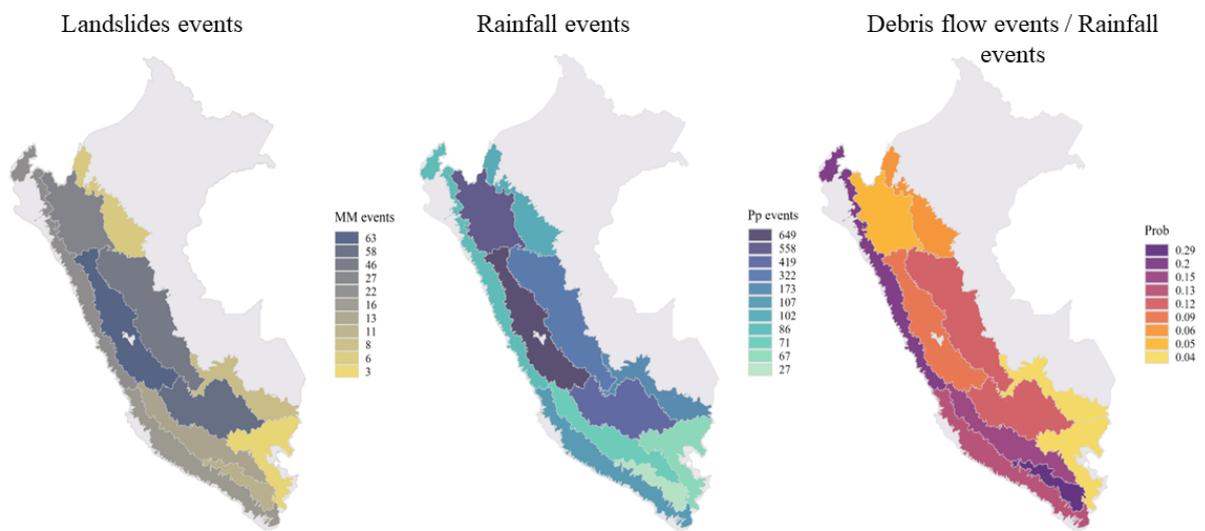
Anexo 2: Información de umbrales de una sola variable para cada región

La primera columna muestra la distribución espacial de los umbrales de lluvia para magnitudes de variables independientes para Perú: a) D (días), b) precipitación acumulada total E (mm), c) intensidad media diaria I_{mean} (mm/d), y d) intensidad máxima diaria I_{max} (mm/d). La segunda y tercera columnas muestran mapas bivariados, los cuales indican la distribución espacial de la sensibilidad (probabilidad de predecir correctamente eventos de lluvia que desencadenan movimientos en masa) y la especificidad (probabilidad de predecir correctamente eventos de lluvia no desencadenantes de movimientos en masa) de los umbrales para las fases de calibración y validación.



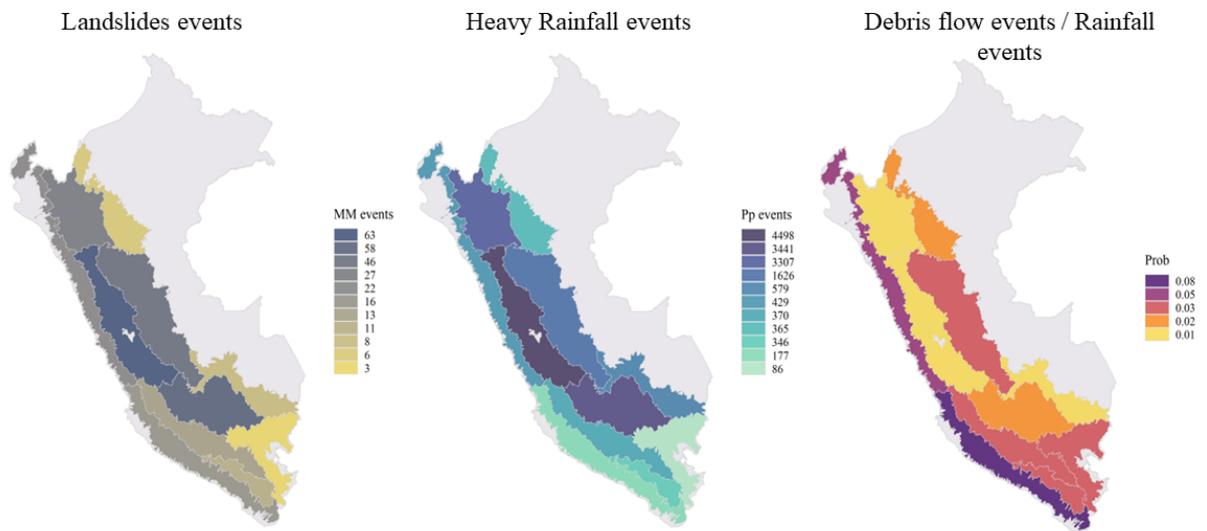
Anexo 3: Probabilidad de ocurrencia de movimientos en masa basado en la presencia de lluvias intensas

Probabilidad de ocurrencia de movimientos en masa basado solo en eventos de lluvia que superan el umbral $I_{mean} - D$. Se observa la correspondencia entre el número de eventos de MM y de lluvias, obteniendo una probabilidad de la ocurrencia de movimientos en masa generados por lluvias intensas. Las regiones Andes 5 y Pacific 1 (donde se encuentran las quebradas de alto impacto) superan el 0.2 (20%) de probabilidad de que ocurrencia de un MM dada una lluvia intensa que supera el umbral $I_{mean} - D$.



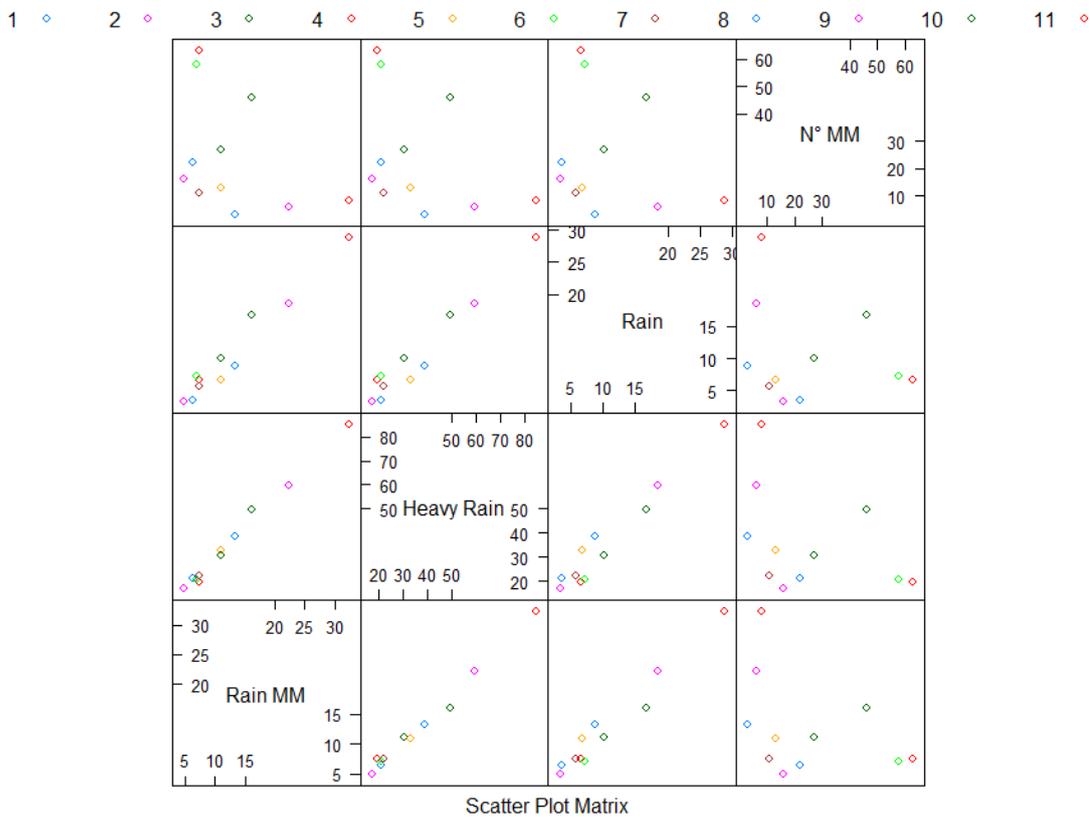
Anexo 4: Probabilidad de ocurrencia de movimientos en masa basado en la presencia de lluvias intensas

Probabilidad de ocurrencia de movimientos en masa basado solo en eventos de lluvia que superan el valor medio más la desviación estándar de las lluvias desencadenantes. Se observa la correspondencia entre el número de eventos de MM y de lluvias, obteniendo una probabilidad de la ocurrencia de movimientos en masa generados por lluvias intensas. Las regiones Pacific 1 y 2 (donde se encuentran las quebradas de alto impacto) son las de mayor probabilidad de ocurrencia de un MM dada una lluvia que supera el valor medio de las lluvias desencadenante sumado a su desviación estándar .



Anexo 5: Relación de correspondencia entre variables y el número de eventos de movimientos en masa

Gráfico de dispersión entre el promedio de eventos de lluvia desencadenantes de MM (Rain MM) vs el promedio de eventos de lluvias intensas (Heavy Rain) vs el promedio de todos los eventos diarios de lluvia desencadenantes y no-desencadenantes (Rain) para las 11 regiones. Las lluvias intensas se establecieron a partir de eventos diarios de lluvia que superan el promedio de lluvia desencadenantes sumado a su desviación estándar. No se observa una correlación entre el número de eventos y las intensidades de lluvia, sin embargo, se observa la correspondencia entre los eventos de lluvia de MM y los de lluvia intensas evidenciando la variabilidad espacial de las lluvias.



Anexo 6: Implementación de la investigación en R.

Los códigos realizados (scripts) en R para esta tesis se pueden acceder en:

https://github.com/caemillan/Rainfall_thresholds_for_shallow_landslide

Código para la generación de Umbrales de lluvia asociado a movimientos en masa

Desarrollado por: @Carlos Millan [cmillan@senamhi.gob.pe]

Version 1.0

```
rm(list = ls())
```

Paquetes requeridos

```
library(dplyr)
```

```
library(ggplot2)
```

```
library(rtop)
```

```
library(rgdal)
```

```
library(rtop)
```

```
library(hydroGOF)
```

```
library(devEMF)
```

```
library(extrafont)
```

```
library(pROC)
```

```
library(ROCR)
```

```
if (!require('future.apply')) install.packages('future.apply')
```

```
wd <- "C:/Users/usuario/SENAMHI_DHI_SEH/2019_SENAMHI-SILVIA/R/"
```

```
setwd(wd)
```

```
hydroshed <- raster::shapefile("5Regiones/shp/GEOGLOWS_SilviaV3.shp")
```

SECCION 1: PREPARACION DE DATOS DE ENTRADA

1.1 Lectura y formato de datos

```
huaycosEvents <- read.csv("7Analisis_huaycos/Output/
                          Huaycos273_param_ajustados_2020.csv")
huaycosEvents <- huaycosEvents[,-1]
huaycosEvents$class <- as.factor(1)
huaycosEvents$Fecha <- as.Date(huaycosEvents$Fecha)
head(huaycosEvents)
PpEvents <- read.csv("7Analisis_huaycos/Output/EventosPp_param_2020.csv")
PpEvents <- PpEvents[,-1]
PpEvents <- PpEvents[,-3]
PpEvents$class <- as.factor(2)
PpEvents$Fecha.Inicio <- as.Date(PpEvents$Fecha.Inicio)#,"%m/%d/%y")
head(PpEvents)
```

1.2 Uniformizacion de datos

Seleccionamos las variables y datos importantes.

En modelos de clasificacion se requieren:

Variables predictoras y factores de clases.

```
names(huaycosEvents)
names(PpEvents)
#####
escenario <- "Esc2" # elegir entre 1 o 2 (este ultimo incluye el ultimo evento)
#Escenario 1: Antes del evento
xx <- dplyr::select(huaycosEvents,"Fecha","ID_Hydro", "N.Evento",
                   "Imax_PD","E_PD",
                   "D_PD", "I_mean_PD","Class") %>%
  na.omit()
#Escenario 2: Incluye el evento
xx <- dplyr::select(huaycosEvents,"Fecha","ID_Hydro", "N.Evento",
```

```

      "Imax_SD", "E_SD",
        "D_SD", "I_mean_SD", "Class")
#####
yy <- dplyr::select(PpEvents, "Fecha.Inicio", "Pixel.Station", "N.Event",
  "Imax..mm.d.", "E..mm.",
  "D..day.", "Imean..mm.d.", "Class") %>%
  .[which(format(as.Date(.$Fecha, "%d%b%Y"), "%Y") %in% c(2018,2019,2020)),]
#####

names(yy) <- names(xx)
BaseDatos <- bind_rows(yy,xx)

#Seleccion de Cluster o region
regiones <- hydroshed@data %>% select(ID_Hydro=HydroID, Cluster)
regiones$ID_Hydro <- as.integer(regiones$ID_Hydro)
BaseDatos <- left_join(BaseDatos, regiones, by = "ID_Hydro" )

names(BaseDatos) <- c("Fecha", "ID_Hydro", "N.Evento", "Imax [mm/d]", "E
  [mm/d]", "D [d]", "I_mean [mm/d]", "Class", "Cluster")
BaseDatos$Class <- as.factor(BaseDatos$Class)
BaseDatos$Cluster <- as.factor(BaseDatos$Cluster)
# BaseDatos <- BaseDatos[~which(BaseDatos$D [d]'==0),] #Filtramos los eventos
  que no tienen dias previos
head(BaseDatos)
table(BaseDatos$Class)
BaseDatos$Class <- as.numeric(BaseDatos$Class)-1
BaseDatos$Class <- as.factor(BaseDatos$Class)
table(BaseDatos$Class)

#Seleccionamos el umbral Minimo Zero (Columna HydroID y Zero)
a=hydroshed@data[,c("HydroID", "Zero_mod")] #15,29 en Hydroshed
a$HydroID=as.numeric(a$HydroID)

```

```

BaseDatos <- inner_join(x=BaseDatos,y=a,by=c("ID_Hydro"="HydroID"))
# BaseDatos <- BaseDatos[which((BaseDatos$Cluster) %in% c(8,9)),]

# 1.3 Seleccion del set de entrenamiento #####
#Opcion 1: si el periodo de lluvias es diferente a los MM observados
DB.train <- BaseDatos[~which(format(as.Date(BaseDatos$Fecha, "%d%b%Y"), "%Y")
  == 2020),]
DB.test <- BaseDatos[which(format(as.Date(BaseDatos$Fecha, "%d%b%Y"), "%Y")
  == 2020),]
table(DB.train$Class)
table(DB.test$Class)

#===== FIN SECCION 2

#===== Inicio SECCION 3
# SECCION 3: Optimizacion de parametros del modelo para Umbrales univariables
# A nivel de una sola variable discriminante
# esta seccion optimiza el umbral para cada parametro de lluvia analizado

nclusters <- sort(as.numeric(unique(BaseDatos$Cluster)))

# Llamamos a las funciones
source("8UmbralesLluvias/funciones/ModeloUnivariado.R")
Model1 # Modelo Univariado
source("8UmbralesLluvias/funciones/Funcion de optimizacion Univariado.R")
opt_model # Funcion de optimizacion

# Hacemos uso de la computacion en paralelo explicito
# if (!require('future.apply')) install.packages('future.apply')
# library('future.apply')

availableCores()-2

```

```

plan(multisession, workers = availableCores()-2)

StartTime <- Sys.time()
tablas <- future_lapply(c(1:11), function(j) {
  region = nclusters[j]

  variables <- filter(DB.train, Cluster==region)
  variables.test <- filter(DB.test, Cluster==region)
  # head(variables)
  # table(variables$class)

  tabla_Umbrales <- data.frame()
  for (i in 4:7) {
    # 3.1 Seleccionamos parametros de variables
    =====

    # 4 <- 'Imax [mm/d]'
    # 5 <- 'E [mm/d]'
    # 6 <- 'D [d]'
    # 7 <- 'I_mean [mm/d]'

    var <- variables[,i]
    # Parametros del modelo
    param <- data.frame(Imax=mean(1,mean(var[which(variables$class==1)])),
                       c=5)

    # ggplot()+
    # geom_density(aes(x=var, color=variables$class))+
    # geom_vline(aes(xintercept=mean(var), color="darkblue", linetype="dashed",
    size=1))+
    # scale_color_discrete(name="Tipo de evento")+
    # geom_vline(aes(xintercept=mean(var[which(variables$class==1)])),
    color="green", linetype="dashed", size=1)+

```

```

# theme_bw()
param_min <- c(1,1)
param_max <- c(mean(var[which(variables$Class==1)])+1,10)

Observed_class <- as.numeric(variables$Class)-1
# Observed_class[which(Observed_class==2)]=0
# table(Observed_class)

# Ejecutar optimizacion con SCE-UA #####
# (Shuffled Complex Evolution Method Developed at The University of Arizona)
  Global Optimization Method
ans2 <- sceua(OFUN=opt_model,
             pars=as.numeric(param),
             lower=param_min,
             upper=param_max,
             maxn=100000,
             c(T, T),
             var=var,
             Observed_class=Observed_class)
param_tuw <- ans2$par # Parametros optimos
# param_tuw <- c(6.32,5)

#Grafico 1: Densidades #####
den1 <- density(as.numeric(var[which(Observed_class==0)]))
den2 <- density(as.numeric(var[which(Observed_class==1)]))
xmin <- min(c(min(den1$x),min(den2$x)))
xmax <- max(c(max(den1$x),max(den2$x)))
ymin <- min(c(min(den1$y),min(den2$y)))
ymax <- max(c(max(den1$y),max(den2$y)))

plot(NULL,xlim=c(0,xmax),ylim=c(0,ymax*1.25), type="n",xlab="Variable",

```

```

    ylab="Densidades")
axis(1, tck=1, col.ticks="light gray")
axis(2, tck=1, col.ticks="light gray")
lines(den1,col="darkgray",lwd=2)
lines(den2,col="#ca5268",lwd=2)
abline(v=param_tuw[1],col="#38b2a3",lwd=2,lty = 2:6)
# abline(v=opt.cut,col="red",lwd=2,lty = 2:6)
text(xmax*0.7,ymax*0.9, "Lluvia no desencadenante", col="darkgray")
text(xmax*0.7,ymax*0.8, "Lluvia desencadenante", col="#ca5268")
if (i==4) {#Imax
    text(xmax*0.7,ymax*0.7, paste("Umbral=",round(param_tuw[1],2),"mm/d"),
        col="#38b2a3")
    title("Curva densidad – Imax")}
if (i==5) { #E
    text(xmax*0.7,ymax*0.7, paste("Umbral=",round(param_tuw[1],2),"mm"),
        col="#38b2a3")
    title("Curva densidad – E [mm]")}
if (i==6) { # Day
    text(xmax*0.8,ymax*0.7, paste("Umbral=",round(param_tuw[1]),"d"),
        col="#38b2a3")
    title("Curva densidad – D")}
if (i==7) { #Imean
    text(xmax*0.7,ymax*0.7, paste("Umbral=",round(param_tuw[1],2),"mm/d"),
        col="#38b2a3")
    title("Curva densidad – Imean [mm/d]")}
dev.off()
#Fin Grafico 1: Densidades #####

# Metricas de calibracion con la Curva ROC para umbrales univariado
#####
#Curva ROC (Receiver operating characteristic)
# representacion grafica de la sensibilidad frente a la especificidad

```

```

# para un sistema clasificador binario segun se varia el umbral de discriminacion
pred <- prediction(as.numeric(var), as.factor(Observed_class)) # Variable como
vector numerico y su clasificacion

perf <- performance(pred,measure="tpr",x.measure="fpr") # Evaluacion del
rendimiento ROC

# Area bajo la curva
AUC <- performance(pred,measure="auc")
AUCaltura <- AUC@y.values

# Punto de corte optimo
TSS_list <- perf@y.values[[1]]-perf@x.values[[1]] #TSS=TPR-FPR or
TSS=sen-(1-spec)
opt.cut <- pred@cutoffs[[1]][which.max(TSS_list)]
cat("AUC:", AUCaltura[[1]])
cat("Punto de corte optimo:",opt.cut)

objroc <- roc(as.factor(Observed_class), as.numeric(var),auc=T,ci=T)
objroc
TSS <-objroc$sensitivities + objroc$specificities -1
maximo <- max(TSS)
numOrdenCutoff <- which(TSS==maximo)
cat("Punto de corte optimo:",objroc$thresholds[numOrdenCutoff])

#Grafico 2: ROC #####
png(paste("Graficos/Esc1_Univariados/",escenario,"_Region",region,"-Curva
ROC",gsub("/","x",names(variables)[i]),Sys.Date(),"v2.png",sep = ""))

plot(perf,colorize=T,type="l",xlab="1-Especificidad",ylab="Sensibilidad",lwd=2)
abline(a=0,b=1)

#coordenadas del punto de corte optimo

```

```

x<-perf@x.values[[1]][which.max(TSS_list)]
y<-perf@y.values[[1]][which.max(TSS_list)]
points(x,y, pch=20, col="red")
text(0.7,0.5, paste("AUC=",round(AUCaltura[[1]],2)), col="gray10")
if (i==4) {#Imax
  text(0.7,0.4, paste("Umbral=",round(opt.cut,2),"mm"), col="gray10")}
if (i==5) { #E
  text(0.7,0.4, paste("Umbral=",round(opt.cut,2),"mm"), col="gray10")}
if (i==6) { #D
  text(0.7,0.4, paste("Umbral=",round(opt.cut)," dias"), col="gray10")}
if (i==7) { #Imean
  text(0.7,0.4, paste("Umbral=",round(opt.cut,2),"mm/d"), col="gray10")}
dev.off()
#Fin Grafico 2: ROC #####

# 3.2 Ejecutar Modelo de Umbral con parametros optimos
=====

# param_opt=param_tuw #Parametros obtenidos del modelo de optimizacion
param_opt=c(opt.cut,param_tuw[2]) #Parametro obtenido del analisis ROC

# Ejecutar modelo calibrado #####

ans3 <- Model1(x=var,
               A=as.numeric(param_opt[1]),
               B=as.numeric(param_opt[2]))

# Extraer observados y simulados
ConfusionMatrix <- data.frame(clase_predic=ans3,
                              clase_obs=as.logical(Observed_class)) #Pisac

#Matriz de Confusion
positive <- sum(ConfusionMatrix$clase_obs==T)

```

```

negative <- sum(ConfussionMatrix$class_obs==F)
predicted_positive <- sum(ConfussionMatrix$class_predic==T)
predicted_negative <- sum(ConfussionMatrix$class_predic==F)
total <- nrow(ConfussionMatrix)

tp<-sum(ConfussionMatrix$class_obs==T & ConfussionMatrix$class_predic==T)
tn<-sum(ConfussionMatrix$class_obs==F & ConfussionMatrix$class_predic==F)
fp<-sum(ConfussionMatrix$class_obs==F & ConfussionMatrix$class_predic==T)
fn<-sum(ConfussionMatrix$class_obs==T & ConfussionMatrix$class_predic==F)

# Calcular metricas
accuracy <- (tp+tn)/total
error_rate <- (fp+fn)/total
sensitivity <- tp/positive
especificity <- tn/negative
precision <- tp/predicted_positive
npv <- tn / predicted_negative
TSS <- sensitivity-(1-especificity) #True Skill statistic

# data.frame(positive, negative,predicted_positive,predicted_negative)
# data.frame(tp,tn,fp,fn)
# data.frame(accuracy,error_rate,sensitivity,especificity,precision,npv,TSS)
# cat("Umbrales:", param_tuw[1],opt.cut,objroc$thresholds[numOrdenCutoff])
# cat("AUC:", AUCaltura[[1]])

# Ejecutar modelo validado #####
var_test <- variables.test[,i]
Obs_test_class <- as.numeric(variables.test$Class)-1

ans32 <- Model1(x=var_test,
                A=as.numeric(param_opt[1]),
                B=as.numeric(param_opt[2]))

```

```
# Extraer observados y simulados
```

```
ConfussionMatrix <- data.frame(clase_predic=ans32,  
  clase_obs=as.logical(Obs_test_class)) #Pisac
```

```
#Matriz de Confusion
```

```
positive.test <- sum(ConfussionMatrix$clase_obs==T)
```

```
negative.test <- sum(ConfussionMatrix$clase_obs==F)
```

```
predicted_positive.test <- sum(ConfussionMatrix$clase_predic==T)
```

```
predicted_negative.test <- sum(ConfussionMatrix$clase_predic==F)
```

```
total.test <- nrow(ConfussionMatrix)
```

```
tp.test<-sum(ConfussionMatrix$clase_obs==T &  
  ConfussionMatrix$clase_predic==T)
```

```
tn.test<-sum(ConfussionMatrix$clase_obs==F &  
  ConfussionMatrix$clase_predic==F)
```

```
fp.test<-sum(ConfussionMatrix$clase_obs==F &  
  ConfussionMatrix$clase_predic==T)
```

```
fn.test<-sum(ConfussionMatrix$clase_obs==T &  
  ConfussionMatrix$clase_predic==F)
```

```
# Calcular metricas
```

```
accuracy.test <- (tp.test+tn.test)/total.test
```

```
error_rate.test <- (fp.test+fn.test)/total.test
```

```
sensitivity.test <- tp.test/positive.test
```

```
especificity.test <- tn.test/negative.test
```

```
precision.test <- tp.test/predicted_positive.test
```

```
npv.test <- tn.test / predicted_negative.test
```

```
TSS.test <- sensitivity.test-(1- especificity.test) #True Skill statistic
```

```
# data.frame(positive, negative,predicted_positive,predicted_negative)
```

```
# data.frame(tp,tn,fp,fn)
```

```

# data.frame(accuracy,error_rate,sensitivity,especificity,precision,npv,TSS)
# cat("Umbrales:", param_tuw[1],opt.cut,objroc$thresholds[numOrdenCutoff])
# cat("AUC:", AUCaltura[[1]])

#Salvamos los umbrales y las metricas #####
tabla_Umbrales[i-3,"Region"] <- region
tabla_Umbrales[i-3,"Variable"] <- names(variables)[i]
tabla_Umbrales[i-3,"Umbral-Modelo"] <- param_tuw[1]
tabla_Umbrales[i-3,"Umbral-ROC"] <- opt.cut
tabla_Umbrales[i-3,"Positivos"] <- positive
tabla_Umbrales[i-3,"Negativos"] <- negative
tabla_Umbrales[i-3,"Predicted_P"] <- predicted_positive
tabla_Umbrales[i-3,"Predicted_N"] <- predicted_negative
tabla_Umbrales[i-3,"T"] <- total
tabla_Umbrales[i-3,"TP"] <- tp
tabla_Umbrales[i-3,"TN"] <- tn
tabla_Umbrales[i-3,"FP"] <- fp
tabla_Umbrales[i-3,"FN"] <- fn
tabla_Umbrales[i-3,"Espec"] <- especificity
tabla_Umbrales[i-3,"Sens"] <- sensitivity
tabla_Umbrales[i-3,"TSS"] <- TSS
tabla_Umbrales[i-3,"AUC"] <- AUCaltura[[1]]

tabla_Umbrales[i-3,"Positivos_test"] <- positive.test
tabla_Umbrales[i-3,"Negativos_test"] <- negative.test
tabla_Umbrales[i-3,"Predicted_P_test"] <- predicted_positive.test
tabla_Umbrales[i-3,"Predicted_N_test"] <- predicted_negative.test
tabla_Umbrales[i-3,"T_test"] <- total.test
tabla_Umbrales[i-3,"TP_test"] <- tp.test
tabla_Umbrales[i-3,"TN_test"] <- tn.test
tabla_Umbrales[i-3,"FP_test"] <- fp.test
tabla_Umbrales[i-3,"FN_test"] <- fn.test

```

```

    tabla_Umbrales[i-3,"Espec_test"] <- especificity.test
    tabla_Umbrales[i-3,"Sens_test"] <- sensitivity.test
    tabla_Umbrales[i-3,"TSS_test"] <- TSS.test
  }
  tabla_Umbrales
})

EndTime <- Sys.time()-StartTime
EndTime

plan(sequential) # returns to sequential processing

tabla_Umbrales2 <- do.call("rbind",tablas)
write.csv(tabla_Umbrales2,paste("Output/",escenario,"_Tabla de Umbrales Univariado
  DB Train",Sys.Date(),".csv",sep = ""),row.names = F)

##### Informacion de valores de la Curva ROC #####
CurvasROC <- future_lapply(c(1:11), function(j) {
  region = nclusters[j]

  # Variables:
  # 4 <- 'Imax [mm/d]'
  # 5 <- 'E [mm/d]'
  # 6 <- 'D [d]'
  # 7 <- 'I_mean [mm/d]'

  variables <- filter(DB.train, Cluster==region)
  variables.test <- filter(DB.test, Cluster==region)
  # head(variables)
  # table(variables$class)

```

```

pred <- prediction(variables[4:7], data.frame(variables$Class, variables$Class,
variables$Class, variables$Class))
perf <- performance(pred,measure="tpr",x.measure="fpr") # Evaluacion del
rendimiento ROC

# Area bajo la curva
AUC <- performance(pred,measure="auc")
AUCaltura <- AUC@y.values

# Punto de corte optimo
TSS_list <- mapply('-',perf@y.values,perf@x.values,SIMPLIFY=FALSE)
#TSS=TPR-FPR or TSS=sen-(1-spec)
max.TSS <- lapply(TSS_list,which.max)
opt.cut <- list()
for (i in 1:length(pred@cutoffs)) {
  opt.cut[i] <- list(pred@cutoffs[[i]][max.TSS[[i]])
}
# cat("AUC:", AUCaltura[[2]])
# cat("Punto de corte optimo:",opt.cut[[2]])

list(fpr= do.call(qpcR:::cbind.na,perf@x.values), tpr=
do.call(qpcR:::cbind.na,perf@y.values))
})

# CurvasROC.sub <- list()
CurvasROC.fpr <- list()
CurvasROC.tpr <- list()
for (i in 1:length(CurvasROC)){
  CurvasROC.fpr[[i]] <- CurvasROC[[i]][[1]] #Los valores de la lista 1, contienen los
fpr
  CurvasROC.tpr[[i]] <- CurvasROC[[i]][[2]] #Los valores de la lista 2, contienen los

```

```

    tpr
}
CurvasROC.fpr <- do.call(qpcR:::cbind.na, CurvasROC.fpr) #Join data by columns
CurvasROC.tpr <- do.call(qpcR:::cbind.na, CurvasROC.tpr) #Join data by columns
colnames(CurvasROC.fpr) <-
  paste(rep(paste(names(variables[4:7]), "fpr", sep="-"), 11), rep(1:11, 4), sep = " ")
colnames(CurvasROC.tpr) <-
  paste(rep(paste(names(variables[4:7]), "tpr", sep="-"), 11), rep(1:11, 4), sep = " ")

# write.csv(CurvasROC.sub, "Output/Valores Curvas ROC.csv")
#####

#===== FIN SECCION 3

# SECCION 4: Optimizacion de parametros del modelo Potencial para Umbrales

# A nivel de curvas discriminantes
# esta seccion optimiza el umbral para parametros reacionados ID, ED, IE

nclusters <- sort(as.numeric(unique(BaseDatos$Cluster)))
# tablas <- list()

# Funcion modelo
exponencialModel <- function(x,A,B){A*x^B}
# Funcion de optimizacion
source("~/R/R_projects/UmbralesDiarios/8UmbralesLluvias/funciones/Funcion de
  optimizacion bivariado.R")
opt_model

plan(multisession, workers = availableCores()-2)

StartTime <- Sys.time()

```

```

tablas_2var <- future_lapply(c(1:11), function(j) {
  # 4.0 Seleccionamos region para modelaci?n#####
  region = nclusters[j]
  variables <- filter(DB.train, Cluster==region)
  # head(variables)
  # table(variables$class)
  variables <- variables[,c(1,2,3,4,5,7,6,8,9,10)] #Reordenamos las variables
  c <- c(4,5,6) # columnas con las variables predecidas
  # 4 <- 'Imax [mm/d]'
  # 5 <- 'E [mm/d]'
  # 6 <- 'I_mean [mm/d]'
  # 7 <- 'D [d]' #Variable estatica

  variables.test <- filter(DB.test, Cluster==region)
  # head(variables)
  # table(variables.test$class)
  variables.test <- variables.test[,c(1,2,3,4,5,7,6,8,9,10)] #Reordenamos las variables

  tabla_Umbrales <- list() #Lista para guardar los umbrales

  day <- variables[, "D [d]" ] #Variable predictora
  Observed_class <- as.numeric(variables$class)-1 #Clase observada (la clase factor se
    convierte en numerico para posteriormente pasarlo a valores logicos)

  day_test <- variables.test[, "D [d]" ] #Variable predictora
  Obs_test_class <- as.numeric(variables.test$class)-1 #Clase observada (la clase
    factor se convierte en numerico para posteriormente pasarlo a valores logicos)

  ## Funcion modelo
  # exponencialModel <- function(x,A,B){A*x^B}
  ## Funcion de optimizacion

```

```

# source("~/8UmbralesLluvias/funciones/Funcion de optimizacion bivariado.R")
## opt_model

for (i in c) {
  # 4.1 Seleccionamos parametros de variables

  obs <- variables[,i] #Variable predecida
  obs_test <- variables.test[,i] #Variable predecida

  # Parametros del modelo
  param_expmodel <-
    data.frame(A=mean(c(1,mean(obs[which(variables$Class==1)]))),
              B=-0.5)

  # Ploteo de los datos para determinar y restringir los valores de los parametros
  =====
  y <- exponencialModel(day,
    as.numeric(param_expmodel[1]),as.numeric(param_expmodel[2]))

  # ggplot()+
  # geom_point(aes(x=day, y=obs,color=variables$Class))+
  # geom_line(aes(x=day, y,color="Umbral"))+
  # geom_hline(aes(yintercept=mean(obs)), color="darkblue", linetype="dashed",
    size=1)+
  # scale_color_discrete(name="Tipo de evento:")+
  # geom_hline(aes(yintercept=mean(obs[which(variables$Class==1)])),
    color="green", linetype="dashed", size=1)+
  # theme_bw()

  #Rango minimos y maximo de los parametros =====
  param_min <- c(1,-1)
  param_max <- c(max(obs[which(variables$Class==1)])+1,0)

```

```

param_max2 <- c(2*max(obs[which(variables$Class==1)])+1,0)
param_max3 <- c(3*max(obs[which(variables$Class==1)])+1,0)

# Ejecutar optimizacion con SCE-UA =====
# (Shuffled Complex Evolution Method Developed at The University of Arizona)
# Global Optimization Method

#Curva 1: El valor de corte sera maximo 1 veces el parametro maximo analizado
ans <- sceua(opt_model,
             pars=as.numeric(param_expmodel),
             lower=param_min,
             upper=param_max,
             maxn=100000,#50 mil iteraciones
             day=day,
             obs=obs,
             Observed_class=Observed_class)

#Curva 2: El valor de corte sera maximo 2 veces el parametro maximo analizado
ans2 <- sceua(opt_model,
              pars=as.numeric(param_expmodel),
              lower=param_min,
              upper=param_max2,
              maxn=100000,#50 mil iteraciones
              day=day,
              obs=obs,
              Observed_class=Observed_class)

#Curva 3: El valor de corte sera maximo 3 veces el parametro maximo analizado
ans3 <- sceua(opt_model,
              pars=as.numeric(param_expmodel),
              lower=param_min,
              upper=param_max3,
              maxn=100000,#50 mil iteraciones
              day=day,

```

```

        obs=obs,
        Observed_class=Observed_class)
param_tuw <- ans$par # Parametros optimos
param_tuw2 <- ans2$par # Parametros optimos
param_tuw3 <- ans3$par # Parametros optimos
# param_tuw=c(14.75,-0.79)
params <- list(param_tuw,param_tuw2,param_tuw3)

# 4.2 Ejecutar Modelo de Umbral Curva Intensidad-Duracion o Modelo de la Curva
# Potencial I=D con parametros optimos=====
# Ejecucion y obtencion de metricas para 3 diferentes umbrales por parametro
Sub_TablaUmbrales <- list()
Ans_Modelo <- list()
Ans_Modelo_ <- list()
for (k in 1:3) {
  # Ejecutar modelo CALIBRADO =====
  # param_opt= c(15.60,-0.81)
  param_opt=params[[k]]
  ans4 <- exponencialModel(day,
                           as.numeric(param_opt[1]),
                           as.numeric(param_opt[2]))
  Ans_Modelo[[k]] <- as.data.frame(ans4,)

  # Extraer observados y simulados
  ConfussionMatrix <- data.frame(sim=ans4,
                                obs=obs,clase_obs=as.logical(Observed_class)) #Pisac
  ConfussionMatrix <- mutate(ConfussionMatrix, clase_predic = (obs>=sim))

  #Matriz de Confusion
  positive <- sum(ConfussionMatrix$clase_obs==T)
  negative <- sum(ConfussionMatrix$clase_obs==F)
  predicted_positive <- sum(ConfussionMatrix$clase_predic==T)

```

```

predicted_negative <- sum(ConfussionMatrix$class_predic==F)
total <- nrow(ConfussionMatrix)

tp<-sum(ConfussionMatrix$class_obs==T &
        ConfussionMatrix$class_predic==T)
tn<-sum(ConfussionMatrix$class_obs==F &
        ConfussionMatrix$class_predic==F)
fp<-sum(ConfussionMatrix$class_obs==F &
        ConfussionMatrix$class_predic==T)
fn<-sum(ConfussionMatrix$class_obs==T &
        ConfussionMatrix$class_predic==F)

# Calcular metricas
accuracy <- (tp+tn)/total
error_rate <- (fp+fn)/total
sensitivity <- tp/positive
especificity <- tn/negative
precision <- tp/predicted_positive
npv <- tn / predicted_negative
TSS <- sensitivity-(1-especificity) #True Skill statistic

# Ejecutar modelo VALIDACION =====
# param_opt= c(15.60,-0.81)
param_opt=params[[k]]
ans42 <- exponencialModel(day_test,
                           as.numeric(param_opt[1]),
                           as.numeric(param_opt[2]))
# Ans_Modelo[[k]] <- as.data.frame(ans42,)

# Extraer observados y simulados
ConfussionMatrix.test <- data.frame(sim=ans42,
                                     obs=obs_test,clase_obs=as.logical(Obs_test_class))

```

```
ConfussionMatrix.test <- mutate(ConfussionMatrix.test, clase_predic =  
  (obs>=sim))
```

#Matriz de Confusion

```
positive.test <- sum(ConfussionMatrix.test$clase_obs==T)  
negative.test <- sum(ConfussionMatrix.test$clase_obs==F)  
predicted_positive.test <- sum(ConfussionMatrix.test$clase_predic==T)  
predicted_negative.test <- sum(ConfussionMatrix.test$clase_predic==F)  
total.test <- nrow(ConfussionMatrix.test)
```

```
var
```

```
tp.test <-sum(ConfussionMatrix.test$clase_obs==T &  
  ConfussionMatrix.test$clase_predic==T)  
tn.test <-sum(ConfussionMatrix.test$clase_obs==F &  
  ConfussionMatrix.test$clase_predic==F)  
fp.test <-sum(ConfussionMatrix.test$clase_obs==F &  
  ConfussionMatrix.test$clase_predic==T)  
fn.test <-sum(ConfussionMatrix.test$clase_obs==T &  
  ConfussionMatrix.test$clase_predic==F)
```

Calcular metricas

```
accuracy.test <- (tp.test+tn.test)/total.test  
error_rate.test <- (fp.test+fn.test)/total.test  
sensitivity.test <- tp.test/positive.test  
especificity.test <- tn.test/negative.test  
precision.test <- tp.test/predicted_positive.test  
npv.test <- tn.test / predicted_negative.test  
TSS.test <- sensitivity.test-(1- especificity.test) #True Skill statistic
```

Guardamos las metricas =====

```
Sub_TablaUmbrales[[k]] <- data.frame(region,  
  paste(names(variables)[i],names(variables[\"D [d]\"])),sep =
```

```

    "--"),param_opt[1],param_opt[2],
    positive, negative,predicted_positive,predicted_negative,
    tp,tn,fp,fn,
    accuracy,error_rate,sensitivity,especificity,precision,npv, TSS,
    positive.test, negative.test, predicted_positive.test, predicted_negative.test,
    tp.test,tn.test, fp.test, fn.test,
    accuracy.test, error_rate.test, sensitivity.test, especificity.test,
    precision.test, npv.test, TSS.test)

# #Salvamos los umbrales y las metricas
# Sub_TablaUmbrales[i-3+k,"Region"] <- region
# Sub_TablaUmbrales[i-3+k,"Curva"] <-
    paste(names(variables)[i],names(variables["D [d]"]),sep = "--")
# Sub_TablaUmbrales[i-3+k,"Cut_Y"] <- param_tuw[1]
# Sub_TablaUmbrales[i-3+k,"Expon"] <- param_tuw[2]
# Sub_TablaUmbrales[i-3+k,"Espec"] <- especificity
# Sub_TablaUmbrales[i-3+k,"Sens"] <- sensitivity
# Sub_TablaUmbrales[i-3+k,"TSS"] <- TSS
}

#===== FIN SECCION 4.2
tabla_Umbrales[[i-3]] <- do.call("rbind",Sub_TablaUmbrales)

#Grafico 1: Umbral #####
x <- c(1:max(day))
y <- as.numeric(param_opt[1])*(x^as.numeric(param_opt[2]))

titulo <- c("Imax (mm/d)", "E (mm/d)", "Imean (mm/d)")
ylab <- c("Pp max [Imax] (mm/d)", "Pp acumulada [E] (mm/d)", "Ppmean [Imean]
    (mm/d)")

# ploteamos los resultados

```

```

fig <- ggplot()+
  geom_point(aes(day[which(Observed_class==0)],
    obs[which(Observed_class==0)]), colour="darkgrey")+
  geom_point(aes(day[which(Observed_class==1)],
    obs[which(Observed_class==1)], shape="Desencadenante"),
    colour="#38b2a3") + #, log="xy"
  geom_line(aes(x=day,y=unlist(Ans_Modelo[[1]]), color="Umbral1"),
    size=1,colour="yellow")+ #
  geom_line(aes(x=day,y=unlist(Ans_Modelo[[2]]), color="Umbral2"),
    size=1,colour="orange")+
  geom_line(aes(x=day,y=unlist(Ans_Modelo[[3]]), color="Umbral3"),
    size=1,colour="#ca5268")+
  scale_shape(name="Evento:")
  labs(title = paste(titulo[i-3], "vs Duracion (d)", "- Reg.",region), x ="Duracion
    (d)",y=ylab[i-3])
  geom_hline(aes(yintercept=mean(obs)), color="darkgrey", linetype="dashed",
    size=1)+
  geom_hline(aes(yintercept=mean(obs[which(variables$Class==1)])),
    colour="#38b2a3", linetype="dashed", size=1)+
  scale_color_discrete(name="Tipo de evento")+
  # scale_x_continuous(trans = 'log10') +
  # scale_y_continuous(trans = 'log10')+
  theme_bw()
}
tabla_Umbrales

})
EndTime2 <- Sys.time()-StartTime
EndTime2

plan(sequential) # returns to sequential processing

```

```

tabla_Umbrales4 <- list()
for (m in 1:11) {
  tabla_Umbrales4[[m]] <- do.call("rbind",tablas_2var[[m]])
}

tabla_Umbrales4 <- do.call("rbind",tabla_Umbrales4)

names(tabla_Umbrales4) <- c("Region", "Curva", "Cut_Y", "Expon",
  "positive", "negative", "predicted_positive",
  "predicted_negative",
  "tp", "tn", "fp", "fn", "accuracy", "error_rate", "sensitivity",
  "especificity", "precision", "npv", "TSS",
  "positive_test", "negative_test",
  "predicted_positive_test", "predicted_negative_test",
  "tp_test", "tn_test", "fp_test", "fn_test",
  "accuracy_test", "error_rate_test", "sensitivity_test",
  "especificity_test", "precision_test", "npv_test",
  "TSS_test")

View(tabla_Umbrales4)
write.csv(tabla_Umbrales4,paste("Output/",escenario,"Tabla de Curvas
  Umbrales", "-", Sys.Date(), "DB Train.csv", sep = ""),row.names = F)

#===== FIN SECCION 4

# install.packages("rgl")
library(rgl)
# install.packages("magick")

new_clusters <- hydroshed@data[,c("HydroID", "Cluster")]
new_clusters$HydroID=as.numeric(new_clusters$HydroID)
BaseDatos <- inner_join(x=BaseDatos,y=a,by=c("ID_Hydro"="HydroID"))

```

```
consulta <- BaseDatos %>%  
  select(-Cluster) %>%  
  inner_join(.,new_clusters,by=c("ID_Hydro"="HydroID")) %>%  
  dplyr::filter(Cluster==1) %>%  
  dplyr::mutate(Color=case_when(Class==0~"grey55",Class==1~"green"))
```
