A DYNAMIC MODEL FOR LANDSLIDE EARLY WARNINGS FOR THE ROAD NETWORK IN COLOMBIA
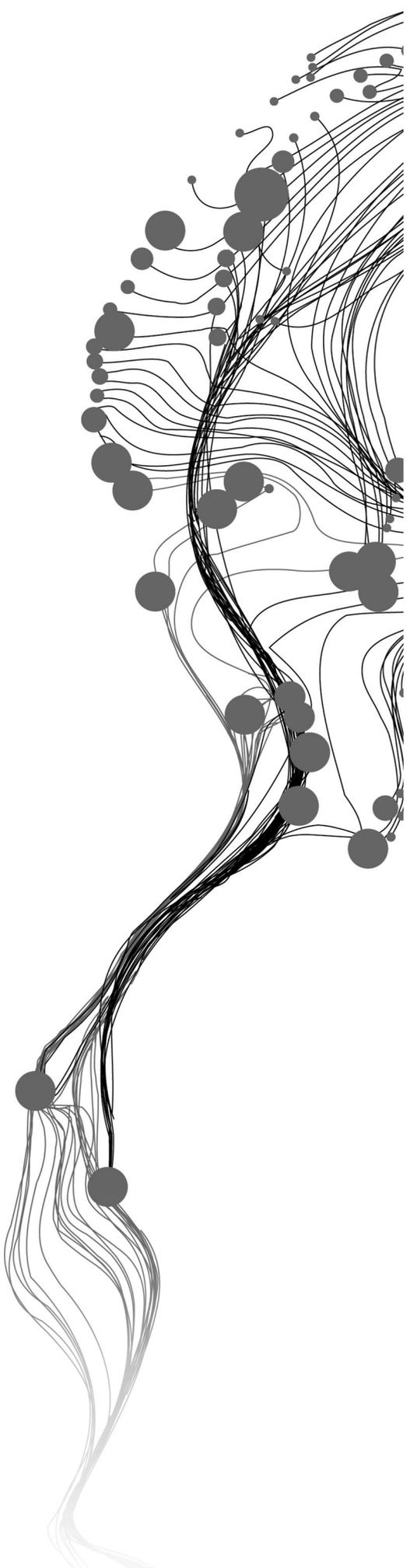
DAVID ALEJANDRO URUEÑA RAMIREZ
June, 2025

SUPERVISORS:
Prof. Dr. C.J. van Westen (First Supervisor)
Dr. L. Lombardo (Second Supervisor)
PhD Candidate M. Moreno (Scientific Advisor)

A DYNAMIC MODEL FOR LANDSLIDE EARLY WARNINGS FOR THE ROAD NETWORK IN COLOMBIA

DAVID ALEJANDRO URUEÑA RAMIREZ
Enschede, The Netherlands, June, 2025

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfillment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialization: Natural Hazards and Disaster Risk Reduction

SUPERVISORS:
Prof.Dr. C.J. van Westen (First Supervisor)
Dr. L. Lombardo (Second Supervisor)
PhD Candidate M. Moreno (Scientific Advisor)

THESIS ASSESSMENT BOARD:
Prof. Dr. V.G. Jetten (Chair)
Dr. A.C. Seijmonsbergen (External Examiner, Institute for Biodiversity and Ecosystem Dynamics, Universiteit van Amsterdam)
B. Virgilio Portela (Procedural Advisor)

ABSTRACT

Background: Landslides are a persistent threat to Colombia's Andean road network, causing frequent closures, damage, and accidents. Traditional landslide early warning approaches in Colombia have been limited by fragmented inventories and static rainfall thresholds that do not account for spatial or temporal variability. This research addresses these gaps by developing a dynamic landslide modeling approach for early warnings on roads. We first evaluated and unified eight different landslide inventories from Colombian agencies (e.g., SIMMA, INVIAS, regional datasets) to overcome data incompleteness and inconsistencies. We then implemented a spatio-temporal predictive model to forecast landslide occurrence, explicitly incorporating rainfall data and environmental factors. The study focuses on enabling operational early warnings; hence, we emphasize model validation, threshold setting for alerts, and practical visualization of results.

Methods: We conducted a comprehensive audit of existing landslide inventories, assessing their attribute quality and completeness using a kernel-density approach for spatial coverage, and a new index to assess temporal coverage, Adjusted Temporal Units Completeness Score (A-TUCS). After harmonizing inventories into a single database (17 824 landslides, 2000-2024), we developed a Generalized Additive Model (GAM) to predict daily landslide probabilities per slope unit. Predictor variables included dynamic rainfall indices (antecedent 15-23 day accumulation and short-term 2-4 day triggers derived from CHIRPS satellite data and CHIRPS-GEFS forecasts) and static terrain, land cover, geology, and road characteristics. We performed extensive model training and validation, utilizing five cross-validation schemes: random, spatially clustered, monthly, yearly sequential, and inventory-based hold-outs, to evaluate model robustness under different dependency assumptions. A Youden's J analysis of the Receiver Operating Characteristic (ROC) curve was used to select an optimal probability threshold for classification of "landslide likely" days. The chosen threshold was then applied in a real-case back-testing scenario (January 2023 storms) to simulate early warning alerts. We designed a traffic-light alert mapping (Green/Yellow/Orange/Red) at the road segment level to convey the model predictions in an operationally meaningful way. All results were interpreted against recent literature on landslide prediction and early warning to contextualize their significance.

Results: The inventory integration yielded a more complete dataset than any individual source, though inventory completeness varies regionally. The spatial completeness analysis showed that certain provinces (Antioquia, Norte de Santander) have landslide densities >1 event/km² (hotspots of reporting), whereas others remain under-reported.

The unified inventory's temporal completeness was moderate (A-TUCS ≈0.86 after year 2000 filter), suggesting some temporal clustering and missing events in quieter periods. The GAM model achieved good predictive performance: average AUROC ≈0.75 across realistic validation scenarios, indicating it can successfully discriminate landslide days from non-landslide days. Calibration was also acceptable (Brier score 0.15 in spatial CV), meaning the predicted probabilities are reasonably in line with observed frequencies. Notably, model performance under spatially and temporally partitioned tests only modestly declined compared to random shuffles, reinforcing confidence in the model's generalizability. Dynamic predictors (rainfall and climate indices) contributed 33% of the model, confirming their critical role; static factors (especially slope steepness and road infrastructure type) also had significant influence, aligning with known hazard determinants. We identified optimal rainfall window lengths of 15days antecedent plus 2 days trigger for CHIRPS, and 23 days + 4 days for CHIRPS–GEFS allow us to forecast - capturing both long-term soil saturation and short intense rainfall bursts as triggers. Importantly, using the CHIRPS-GEFS 5-day rainfall forecasts, the model maintained nearly the same skill as with historical rainfall (forecast AUROC 0.77 vs. 0.78 with observed data), demonstrating the model's suitability for real-time forecasting. The Youden-optimized probability threshold was $T^* = 0.30$ (95% CI: 0.27-0.32), maximizing the sum of sensitivity and specificity. Applying this threshold in the full inventory gave a confusion matrix with True Positives = 3 917, False Positives = 4 016, False Negatives =2 313, True Negatives = 13 483, corresponding to a TSS of 0.533. In a real-case month (Jan 2023), this threshold correctly identified 18 of 21 reported landslides (Recall 85.7%) at slope-unit level, but also issued 62,700 false-positive unit alerts (Precision 0.03%). By aggregating these into road-segment alerts, we reduced actionable alerts to a manageable number. The traffic-light road maps clearly showed authorities when and where the highest risks were concentrated; for Jan 13-15, multiple contiguous segments were Red, coinciding with known road blockages, whereas after Jan 20 the network mostly reverted to Green, matching the all-clear. This suggests that, despite the large false-positive volume at SU level, the system can prioritize effectively and would have provided timely warnings for all major landslide incidents that month. Additionally, a back-analysis of false alarms indicated they were distributed widely rather than repetitively affecting the same site, implying the model's high sensitivity did not overly "cry wolf" at specific locations but rather reflected broad hazard conditions.

Conclusions: We successfully developed a dynamic landslide early warning modeling approach that integrates multi-source data and generates practical alerts for road management in Colombia. The research demonstrates that combining landslide susceptibility factors with real-time rainfall forecasts can significantly improve early warnings compared to static threshold methods. Key contributions include: (1) a unified national landslide inventory with quantified completeness, highlighting data improvements needed; (2) a validated spatio-temporal predictive model (AUROC 0.75, TSS 0.53) that runs on daily updated weather forecasts; and (3) an operational alert framework (traffic-light system) that translates model outputs into clear guidance for decision-makers. The model can reliably identify periods of high landslide likelihood on specific road segments, enabling preventive road closures or other mitigation to be undertaken. The study also underscores the importance of cross-validation strategies that respect

spatial/temporal structure in landslide modeling, providing more realistic estimates of performance for operational use. Recommendations are made for INVIAS and IDEAM to pilot this system, including regular threshold recalibration and data sharing protocols. Limitations such as oversimplified spatial independence assumptions and data gaps are acknowledged, with suggestions for future research to address them. In conclusion, this thesis advances the state-of-the-art in landslide early warning by moving from static regional thresholds to a dynamic, probabilistic model tailored to Colombia's roads. Implementing this approach can enhance disaster preparedness and resilience of critical infrastructure in the face of climatic extremes. It exemplifies how scientific modeling, when coupled with operational considerations, can contribute directly to risk reduction in data-scarce, hazard-prone regions.

ACKNOWLEDGMENTS

Contents

List of Figures

List of Tables

1   INTRODUCTION

Climatological variables are highly dynamic and have the potential to trigger a variety of natural hazards (IPCC, 2012, 2023; WMO, 2024), with floods, droughts, landslides, and wildfires being just some of the most common. Rainfall is a particularly critical variable in multiple Earth systems, shaping the Earth's surface and acting as a primary trigger for landslides. Rainfall-induced landslides are a significant global hazard, causing substantial loss of life (Froude & Petley, 2018; Garcia-Delgado et al., 2022; Gómez et al., 2023b; Pollock & Wartman, 2020), economic losses (Gariano & Guzzetti, 2016; Hallegatte et al., 2017), and disruption of essential infrastructure (Donnini et al., 2017; Winter et al., 2016). The increasing frequency and intensity of extreme weather events are expected to amplify the occurrence of landslides, particularly in tropical regions (Amarasinghe et al., 2024; Gariano & Guzzetti, 2016; IPCC, 2023).

Due to its high susceptibility to rainfall-induced landslides, Colombia serves as an interesting case study (Garcia-Delgado et al., 2022; Gómez et al., 2023a).The Andean region of the country hosts over 75% of the country's road infrastructure and economic activities. This region is also characterized by steep slopes, deep soil profiles, complex geology, and intense rainfall events contributing to frequent landsliding. Landslides have been a significant cause of fatalities in Colombia. According to the Unified Global Landslide Database (UGLD), Colombia recorded 10 393 landslide events between 1903 and 2020, resulting in 35 686 deaths (Gómez et al., 2023b). This has given Colombia the highest number of landslides and fatalities globally during this period (Garcia-Delgado et al., 2022; Gómez et al., 2023b). It is essential to consider that a large portion of these fatalities resulted from a single event, the 1985 Nevado del Ruiz volcanic eruption, which triggered a massive lahar, leading to over 22 000 deaths (Herd, 1986). Even excluding this event, the fatalities are still over 10 000, highlighting landslides as one of the deadliest natural hazards in the country.

Landslides have also affected Colombia's infrastructure, specifically its extensive road network. The road network is the backbone of the country's logistics and transportation system, handling approximately 80% of Colombia's internal cargo and passenger movement (Ministerio de Transporte, 2023). In 2022, this network facilitated the transport of 263 million tons of cargo and over 100 million passengers (Ministerio de Transporte, 2023). The road network is essential for connecting remote communities, facilitating major trade routes, and providing access to vital services. However, landslides frequently disrupt this critical infrastructure, leading to road closures and increased transportation costs. According to the Sistema de Información de Movimientos en Masa (SIMMA; https://simma.sgc.gov.co/)

managed by the Colombian Geological Survey (SGC), around 9 000 landslide events have impacted the roads historically (1950–present), making it the most common reported impact as shown in Figure 1.1 reflecting the high exposure of Colombia's road network. The increased cargo volume and passengers amplify the economic impact of such disruptions, underscoring the urgent need for effective landslide risk management strategies focusing on the road network.
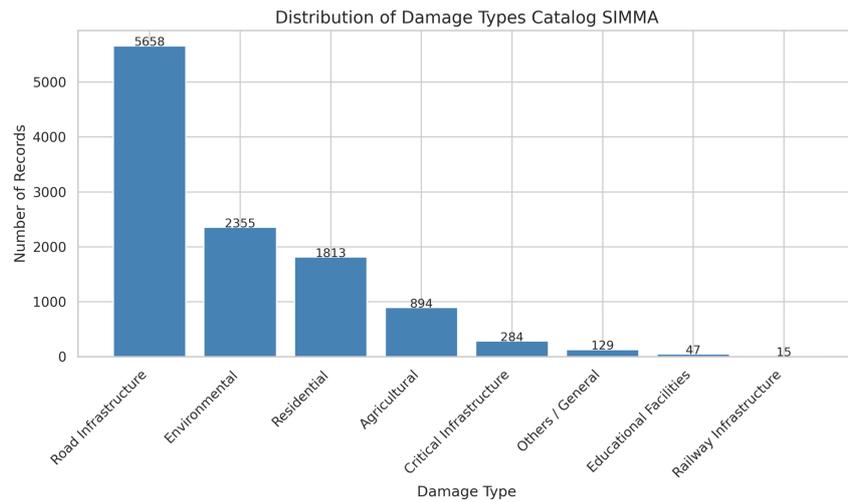
(a) Damage types in the SIMMA landslide catalog.
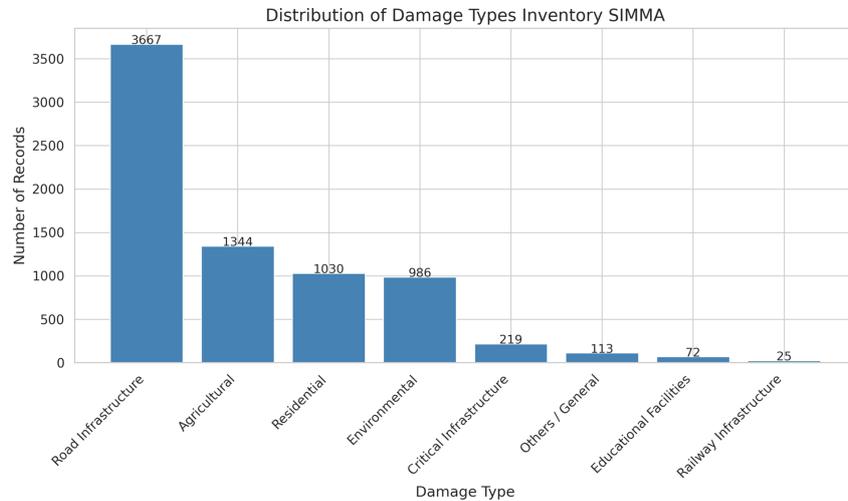
(b) Damage types in the SIMMA landslide inventory.

Figure 1.1: Distribution of landslide-induced damage types recorded by SIMMA in Colombia: (a) the national caTheog; (b) the national inventory. Source: Geological Service of Colombia (SIMMA), https://simma.sgc.gov.co/.

Colombia lacks a practical operational Landslide Early Warning Systems (LEWS) that provide information for potentially exposed segments of the road network. The Institute of Hydrology, Meteorology, and Environmental Studies (Instituto de Hidrología, Meteorología y Estudios Ambientales, IDEAM) forecasts and reports on an administrative level alert levels for landslides only using expected rainfall, but the lack of granularity makes it challenging to make informed decisions with this information. The absence of a detailed LEWS hampers the ability of authorities to anticipate landslide events, allocate resources efficiently, and implement preventive measures. LEWS can provide precise and timely warnings, mitigating the impacts and disruptions caused by landslides (Chae et al., 2017; Guzzetti et al., 2020). The necessity of Early Warning Systems (EWS) is emphasized by the United Nations initiative "Early Warnings for All" (EW4All), looking to ensure that multi-hazard EWS protect everyone on Earth by 2027 (WMO, 2022). A crucial step to advancing toward an operational LEWS is developing reliable spatio-temporal landslide models for predicting when and where landslides will likely occur (Fathani et al., 2016; Guzzetti et al., 2020; Rossi et al., 2019). LEWS saves lives by facilitating timely decision-making and reducing economic impacts by preventing or minimizing infrastructure damage and associated costs, aligning with the EW4All initiative's emphasis on proactive disaster preparedness.

This thesis proposes a comprehensive spatio-temporal landslide predictive framework designed to forecast where and when landslides are most likely to occur within a 24-h horizon. Building on recent advances in dynamic landslide modeling with the generalized additive model (GAM) framework allows us to maintain the high interpretability necessary for a transparent decision-making support (Kruschel et al., 2025). The workflow couples ensemble precipitation forecasts from CHIRPS-GEFS with National landslide inventories, multi-source terrain and lithological data, and road-network attributes. The resulting model prototype for landslide early warning road applications is engineered for real-time ingestion, providing actionable lead time for road management and emergency response agencies. By aligning with the United Nations' EW4All initiative, the modular and open-source architecture offers a transferable pathway for other landslide-prone regions, contributing to the global goal of universal multi-hazard early-warning coverage.

1.1 Landslide Early Warning Systems

A LEWS is a multi-disciplinary and complex architecture that moves gradually—from risk knowledge to real-time action—through mutually dependent phases (Piciullo et al., 2018; Thirugnanam et al., 2020) that we adapt to these four phases, as observed in Figure 1.2:

- **Phase 1 — Setting** defines why, where, and for whom the system is built. Decisions on objectives, spatial coverage, institutional responsibilities, and data-governance rules determine everything that follows; missteps here propagate through the entire system.

- **Phase 2 — Modeling** transforms raw observations into quantitative forecasts. The accuracy, lead time, and

interpretability of these forecasts hinge on the completeness of landslide inventories and on harmonized geo-environmental datasets.

- **Phases 3 & 4 — Warning and Response** operationalize the science and the numbers where thresholds are calibrated, alert levels agreed, and communication protocols tested with the end-users who must act on them (Liu et al., 2023).

This work concentrates on the Setting and Modeling phases, because (i) they are prerequisites for any credible alert service (Rossi et al., 2019; Van Westen et al., 2008), and (ii) they remain the main gap for Colombia's national road network. Strategies for alert dissemination and emergency coordination are also crucial, but those remain for future work with the National agencies IDEAM (Meteorology and climate) and INVIAS (Road management). The correct decisions in the first phase are the foundation for the system's success and deserve deeper discussion.

The development of accurate predictive models (Phase 2) critically depends on the availability of comprehensive and reliable data (Phase 1). Geo-environmental data and landslide inventories are the foundational inputs of any model, and their quality is crucial for any model's success. The landslide inventories provide essential data on the location, timing, type, and triggers of past landslides, forming the empirical basis for understanding landslide processes and modeling future occurrences (Herrera et al., 2018; Tanyaş et al., 2017), making it one of the most critical inputs for any models at the core of the LEWS.

1.1.1 The Data problem: Goals, Completeness & Harmonization

Despite their importance, compiling landslide inventories is challenging. Variations in data sources, mapping methods, scales, and classification systems make it challenging to integrate different datasets (Galli et al., 2008; Garcia-Delgado et al., 2022; Gómez et al., 2023b; Guzzetti et al., 2012; Herrera et al., 2018; Hervás & Bobrowsky, 2009; Tanyaş et al., 2017; Van Den Eeckhaut & Hervás, 2012; Van Westen et al., 2008). Recent advances in remote sensing technologies—such as radar, lidar, and high-resolution satellite imagery have significantly improved landslide detection and mapping capabilities (Casagli et al., 2023). Then again, standardizing and integrating these diverse datasets into unified inventories requires harmonization efforts and the adoption of consistent methodologies (Galli et al., 2008; Tanyaş et al., 2017; Van Den Eeckhaut & Hervás, 2012).

Assessing the completeness and quality of landslide inventories is also an important task. Incomplete or biased datasets can lead to inaccuracies in the model outputs, and EWS outcomes (Guzzetti et al., 2012; Malamud et al., 2004; Tanyas & Lombardo, 2020). This issue is critical in regions like the Andean mountains of Colombia, where landslide activity is high (Amarasinghe et al., 2024; Sepúlveda & Petley, 2015). Colombia does not have the problem of landslide collection, but multiple sources make the harmonization process a challenge, and evaluating the inventory completeness becomes a priority for understanding and making decisions on the different EWS stages. As one cannot expect reliable and unbiased output from LEWS trained on poor landslide data in both quality and completeness.
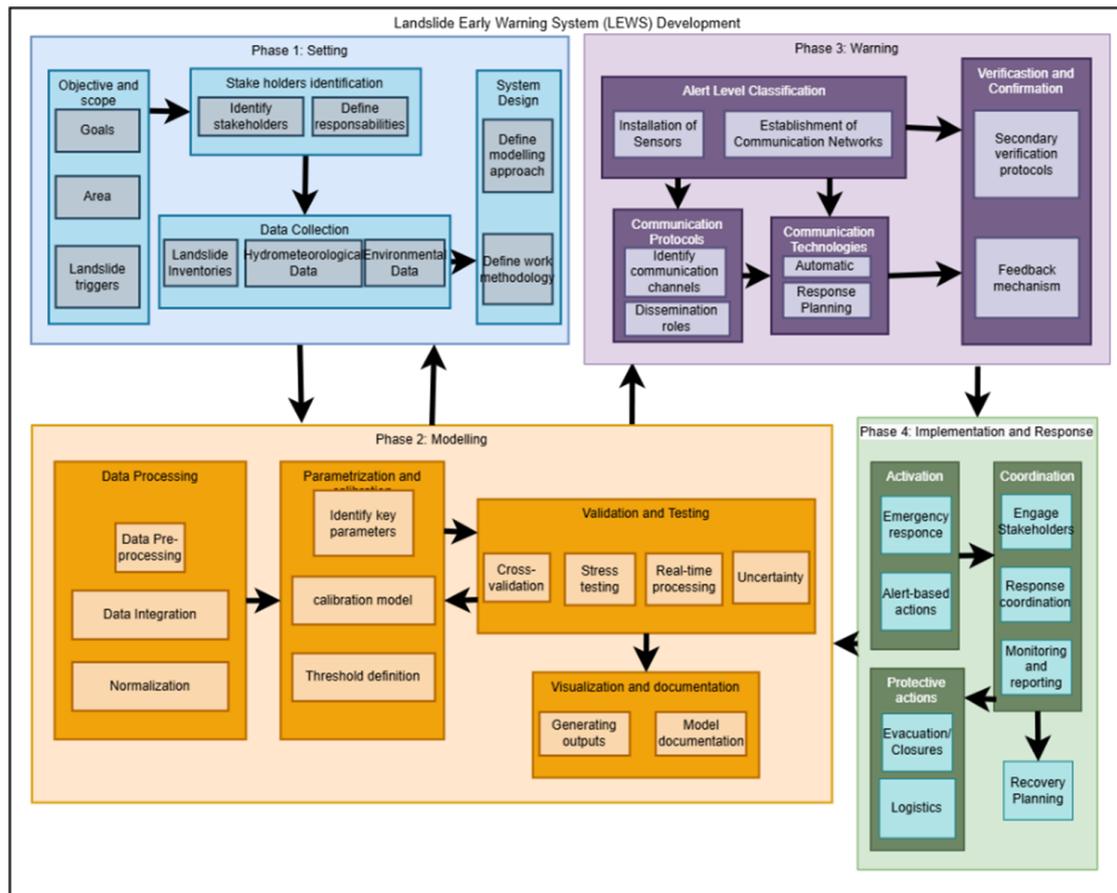
Figure 1.2: Conceptual Framework LEWS Source: Adapted from Piciullo et al. (2018) and Thirugnanam et al. (2020).

1.1.2 Predicting the unstable: Selecting a modeling approach

Determining landslide probability in space and time is necessary for DRM, leading to the development of various modeling approaches broadly categorized into physical, heuristic, and data-driven models. Each approach offers distinct advantages and limitations, influencing their applicability in different contexts.

Physical models simulate the mechanical behavior of slope materials by applying the principles of physics, soil mechanics, and hydrology. They aim to replicate the actual processes leading to landslides, providing detailed insights into the factors causing slope failures (Corominas et al., 2013; Zakaria et al., 2022). Examples include SHALSTAB (Montgomery & Dietrich, 1994), which predicts shallow landslides under saturated conditions; TRIGRS (Baum et al., 2008); SINMAP (Pack et al., 1998), integrating hydrological factors with stability analysis. While valuable for understanding landslide mechanisms, physical models require extensive data on soil properties, hydrology, and precise topography, making them challenging to apply over large regions but helpful for those areas with limited

landslide records.

Heuristic models rely on expert knowledge and qualitative assessments to evaluate landslide susceptibility. These models incorporate local expertise into susceptibility mapping by assigning weights to conditioning factors based on their perceived importance (Anbalagan, 1992), where techniques like the Analytic Hierarchy Process are applied in different contexts. Although straightforward and valuable when data is limited, heuristic models are inherently subjective, potentially leading to inconsistencies and affecting reproducibility and quantitative rigor (Guzzetti et al., 1999).

Data-driven models, or statistical and empirical models, have gained prominence due to their ability to handle large datasets and model complex relationships between landslide occurrences and conditioning factors (Reichenbach et al., 2018). These models utilize statistical techniques to identify patterns in historical landslide data, enabling predictions in other areas. Traditional methods like logistic regression (Ayalew & Yamagishi, 2005) have been widely used, especially because of their straightforward interpretation. Conversely, machine learning techniques such as support vector machine (Huang & Zhao, 2018), random forest (Gómez et al., 2023a), and artificial neural networks (Fang et al., 2023) have traditionally been pursued for the high predictive performance they ensure. Despite their strengths, the effectiveness of data-driven models depends heavily on the quality and completeness of input data, highlighting the importance of comprehensive landslide inventories.

Among data-driven approaches, GAM offers flexibility and interpretability in modeling complex, non-linear relationships between landslide occurrences and predictor factors (Hastie, 2017; Kruschel et al., 2025; Lombardo et al., 2020). These models extend linear models by incorporating functions of predictors, effectively modeling non-linear effects without specifying a particular form. They have been successfully applied in landslide susceptibility mapping, integrating factors like rainfall, topography, and geology (Ahmed et al., 2023; Goetz et al., 2011). Generalized Additive Mixed Models (GAMMs) further enhance this approach by introducing random effects, accounting for spatial and temporal dependencies, and improving predictive accuracy (Moreno et al., 2024; Steger et al., 2024).

Building upon advancements in various modeling approaches, it becomes essential to translate the outputs of landslide models into actionable information for decision-makers. Regardless of whether the models are physical, heuristic, or data-driven, their results—often presented as hazard maps, susceptibility or probabilistic indices that must be effectively communicated to stakeholders involved in DRM (Glade & Nadim, 2014; UNISDR, 2015; WMO, 2022).

Building upon advancements in various modeling approaches, it becomes essential to translate the outputs of landslide models into actionable information for decision-makers. Regardless of whether the models are physical, heuristic, or data-driven, their results—often delivered as hazard maps, susceptibility surfaces, or probabilistic indices—must be communicated in a format that end-users in DRM can immediately interpret and employ (Glade & Nadim, 2014; UNISDR, 2015; WMO, 2022). In practice, this means reducing multi-layered model output to a handful of operational triggers that can slot straight into early-warning workflows and standard operating procedures. Bridg-

ing the gap between complex scientific outputs and practical decision-making tools is crucial for implementing timely interventions and reducing the impact of landslides on vulnerable regions.

Trigger thresholds are a popular tool for closing that gap between scientific outputs and action. Rain-based thresholds satisfy this role because cumulative depth or intensity-duration (ID) values are already monitored by most meteorological agencies and are intuitively understood by road managers and civil-protection staff. A rainfall threshold is simply the precipitation history—usually some combination of event intensity, duration, and antecedent moisture— that marks a sharp increase in landslide probability (Caine, 1980; Nocentini et al., 2023; Wieczorek & Glade, 2005). Thresholds can be derived along a continuum that runs from quick empirical fits to historical landslide-rain pairs (Brunetti et al., 2010; Peruccacci et al., 2012) through deterministic physical simulations to fully probabilistic, region-specific formulations that embed model uncertainty (Steger et al., 2024). Empirical ID curves remain popular because they require little data and can be drawn on a single chart, yet they often transfer poorly between climatic zones, provide no slope-specific detail, and may under-represent small, unreported slides.

1.2 A Local Framework: EWS in Colombia

Colombia has made significant advances in developing operational EWS for DRM, particularly within hydrometeo-rological contexts. IDEAM is the primary institution responsible for issuing technical reports on environmental alerts from hydrometeorological events (Domínguez-Calle et al., 2014). IDEAM employs advanced models such as the Weather Research and Forecasting and Mesoscale Model V5 at various scales. These models utilize initial data from NOAA's Global Forecast System to generate climate and weather predictions for different regions in Colombia (IDEAM, 2008). Predictions encompass short-term, from every 180 seconds up to three days; mid-season from one to two months, and seasonal on a range of three to five months periods, delivering products such as probabilities of macroclimatic phenomena (La Niña/El Niño) occurrences, river level information, precipitation forecasts, landslide risk forecasts, and fire warnings.

Additionally, regional EWS, such as the EWS on The Aburrá Valley (Sistema de Alerta Temprana de Medellín y el Valle de Aburrá, SIATA) and the EWS in Bogotá from the Distrital Institute of Risk Management and Climate Change (Instituto Distrital de Gestión de Riesgos y Cambio Climático, IDIGER), play crucial roles in DRM at a more granular scale. SIATA employs a network of meteorological stations, radar stations, and sensors to monitor weather conditions and issue warnings for floods and landslides in Medellín and its surrounding areas (SIATA, 2024). IDIGER integrates efforts from various public entities, private organizations, and community groups to facilitate risk management and emergency response within Bogotá, utilizing a comprehensive information system to disseminate alerts and coordinate actions (Domínguez-Calle et al., 2014). These regional systems are further complemented by local and community-based EWS, which rely on the participation of local volunteers for monitoring and data collection. No operational EWS focuses on specific sectors or exposed valuables like roads, buildings, or people; they all have a general approach without considering potential impacts.

On the other hand, some recent efforts in Colombia have concentrated on developing comprehensive landslide inventories to understand these events' spatial and temporal distribution. However, there is still a lack of centralized information. Institutional databases like SIMMA, managed by the SGC and the Disaster Inventory System (DesInventar), have been instrumental in compiling landslide data, including location, date, type, and triggering factors (Desinventar, 2022; SGC, 2024). Research initiatives by universities and institutions have augmented these inventories. For instance, the Geohazards research group from the Universidad Nacional in Medellín has made an updated database (https://geohazards.com.co) (Aristizábal et al., 2025), Gómez et al. (2023b) compiled an extensive inventory of 3 536 rainfall-triggered landslides from 1981 to 2019 by integrating data from SIMMA, DesInventar, and Geohazards group. Garcia-Delgado et al. (2022) compiled an inventory from different sources with only fatal landslides and over 2 500 events recorded. Despite these efforts, challenges such as data inconsistencies, spatial and temporal gaps, and difficulties in accurately attributing triggers persist (Garcia-Delgado et al., 2022; Gómez et al., 2023b). There are new inventories that have not been implemented in any other research, such as the one made by INVIAS, that report impacts on the road network by landslides and floods. All of these sources demonstrate the lack of coordination among different entities and the necessity of a unique inventory.

Advanced modeling approaches have been employed to predict landslide occurrences and assess susceptibility in Colombia. Statistical models like logistic regression and weight of evidence have been used to evaluate the relationship between landslide occurrences and conditioning factors such as lithology, slope, land use, and rainfall (Aristizábal et al., 2019). Machine learning techniques, particularly the random forest algorithm, have shown promise in handling complex, non-linear relationships and improving prediction accuracy (Gómez et al., 2023a). These models generate outputs like landslide susceptibility maps and hazard zonation maps, essential for urban planning and disaster risk reduction. However, challenges remain, as spatio-temporal models have been scarcely explored, and their predictive capability is limited to nowcasting since no integration with forecasted rainfall has been done.

Establishing accurate probability thresholds is critical for predicting landslide occurrences in Colombia, and rainfall is the factor that is measured in most cases. Various methods have been employed, including empirical approaches based on statistical analysis of historical data, physically based models simulating slope stability under varying rainfall conditions, and integration of rainfall data into machine learning models (Gómez et al., 2023a; Marin et al., 2021). Specific thresholds have been established in regions like the Central Cordillera, differentiating thresholds based on soil types and antecedent rainfall conditions (Aristizábal et al., 2022). The development of EWS, particularly in areas like Manizales and Medellín, demonstrates significant progress in integrating real-time data, community engagement, and technological innovation (Correa et al., 2020; Marin & Marin-Sanchez, 2024). However, challenges such as the spatial variability of rainfall and the frequency of extreme rainfall events limit the effectiveness of static thresholds. A dynamic probability threshold approach would be more practical for decision-makers, as it can adjust to changing rainfall and environmental conditions, an essential capability currently missing from existing EWS in the country.

While Colombia has made considerable progress in establishing a robust EWS, a spatio-temporal LEWS for specific sectors remains in its nascent stages and has yet to develop fully. Existing systems lack the granularity required to predict landslide events affecting specific road segments and do not fully consider all the dynamic factors in the process.

1.3 Objectives and research questions

1.3.1 General Objective

To develop a spatio-temporal data-driven landslide model for Colombia's Andean road network and enhance early warning capabilities by establishing probability thresholds for individual road segments.

1.3.2 Specific Objectives

First Sub-objective

To compile and integrate existing landslide inventories in the Andean region to create a unified, high-quality database suitable for spatio-temporal predictive modeling.

Research Questions:

- **RQ1.1:** What are the differences in attributes, mapping techniques, data quality, and completeness among the existing landslide inventories in the Andean region of Colombia?
- **RQ1.2:** How can the existing landslide inventories be evaluated and integrated to create a unified landslide inventory suitable for developing a spatio-temporal data-driven model?
- **RQ1.3:** How can the damage information be linked to landslide inventories?

Second Sub-objective

To develop a spatio-temporal predictive landslide model for the Andean road network using suitable dynamic and static factors.

Research Questions:

- **RQ2.1:** Which predictor factors influence landslide occurrences along Colombia's road networks, and how can they be effectively incorporated into a data-driven spatio-temporal susceptibility model?
- **RQ2.2:** What are the optimal time windows for incorporating historical and forecasted rainfall data into the predictive model to enhance its accuracy?
- **RQ2.3:** What is the model performance using the CHIRPS-GEF forecast rainfall dataset?
- **RQ2.4:** What are the differences in landslide prediction between nowcasting and forecasting?

Third Sub-objective

To establish a probability threshold based on the predictive model to translate landslide probabilities into actionable warnings within an Early Warning System for the Andean Road network.

Research Questions:

- **RQ3.1:** How can we demonstrate the effectiveness of the probability threshold in real-case scenarios while reflecting on the most suitable graphical representation?
- **RQ3.2:** How can the established thresholds be applied to provide actionable recommendations and alerts for decision-making by authorities responsible for the Andean Road network?

2 PHYSICAL SETTING AND DATA FOUNDATIONS

Our study area is located at the Northwestern corner of Colombia, which is around 40% of Colombian territory, stretching from the Pacific coast to the eastern slopes to its western slopes, and hosts more than three-quarters of the country's population, GDP, and road infrastructure. Extreme relief, complex litho-tectonic juxtapositions, and one of the most intense equatorial rainfall regimes on Earth turn this corridor into a global hotspot of rainfall-triggered landslides. This chapter equips the reader with: (i) a concise but complete description of the region's physiography, geology and climate (§2.1), and (ii) an exhaustive account of the dynamic and static datasets that underpin our forecasting framework (§2.2).

2.1 Physiographic, Geological and Climatic Context

2.1.1 Relief and geomorphology

The Northwestern part of Colombia, highlighted in the Figure 2.1, forms a triple-range system created where the Nazca oceanic plate converges beneath South America at (Lizarazo et al., 2021). Between the Cordilleras Occidental (Western range), Central (Central range), and Oriental (Eastern range) lie the Cauca and Magdalena grabens—elongated valleys that funnel population, commerce, and, unfortunately, landslide risk. Ten departments (provinces) and the metropolitan areas of Bogotá, Medellín and Cali lie within this mountainous corridor.

2.1.2 Relief and geomorphology

In the study area, we can find elevations range from sea level to $5\,700\,m\,a.s.l.$ at Colon Peak, producing steep hillslopes, deep canyons, and narrow intermontane basins. Active glaciation survives on high volcanoes, while lower belts exhibit fluvio-denudational landscapes, volcanic edifices, and large landslide scars that attest to rapid erosion rates. The Table 2.1 classifies the macro geomorphological domains according to the framework proposed by Carvajal (2012)
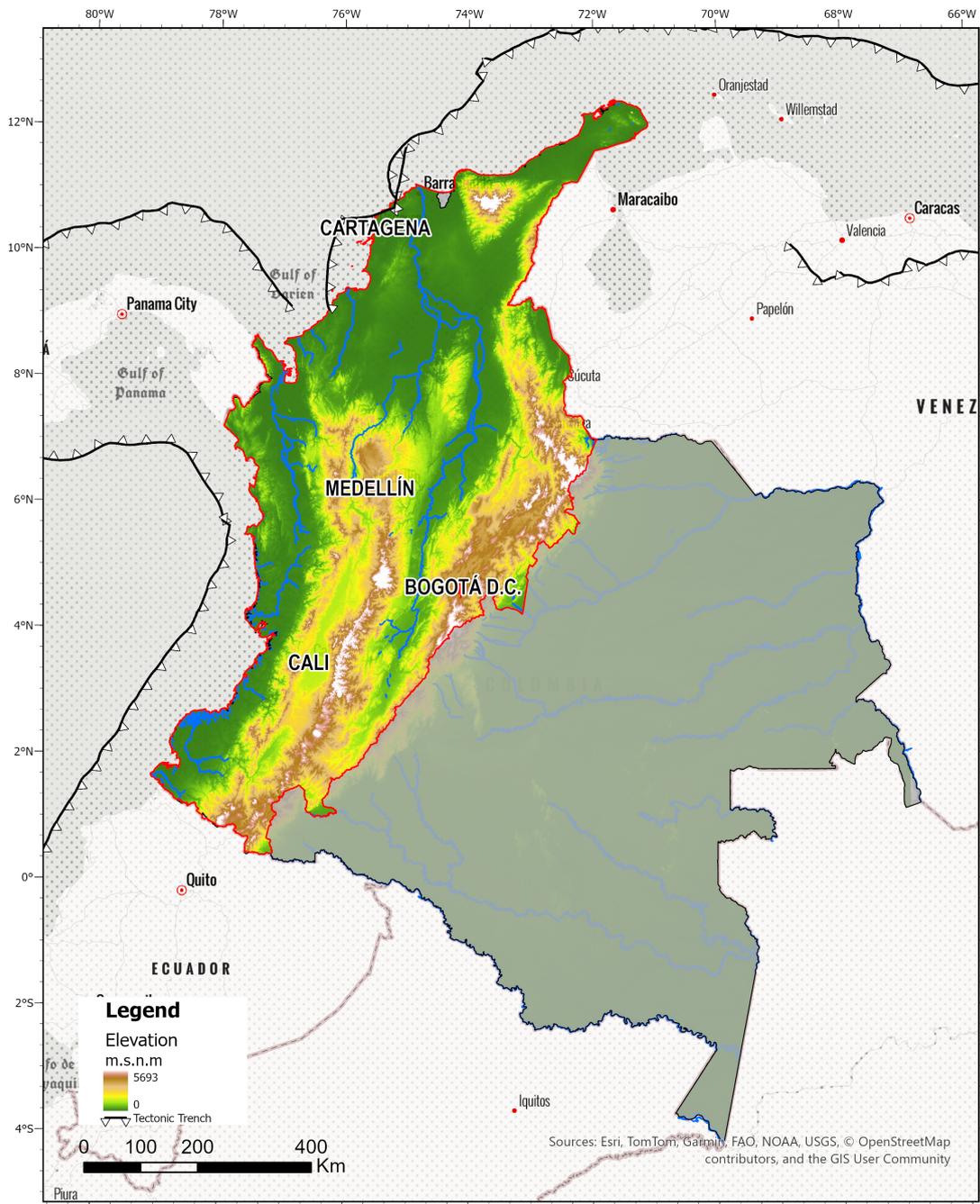
Figure 2.1: Topographic Map of Colombia with Major Rivers, and Coastal Tectonic Trenches. Highlighting the study area in red

Table 2.1: Geomorphological regions in the North Western part of Colombia and their relevance to landslide hazard

Regions	Diagnostic features	Landslide relevance
Orogenic Andean System	Actively uplifting metamorphic & volcanic chains	High seismic shaking, pervasive slope adjustment
Volcanic highlands	Composite cones & pyroclastic plateaux (e.g. Ruiz-Tolima)	Weak, juvenile tephra prone to debris flows
Intermontane valleys	Cauca & Magdalena grabens	Thick weathered colluvium, channel-toe erosion

2.1.3 Lithostratigraphy and fault architecture

The Northwestern Colombian Andes are characterized by the convergence of the oceanic Nazca plate beneath the continental South American plate, which drives arc magmatism and compressional deformation; transpressive coupling with the Caribbean plate further segments the margin into small crustal blocks, as we can observe in 2.1.

The geological composition of the Andean region is diverse and complex. Late Cretaceous island-arc basalts characterize the Western Cordillera, Jurassic-Miocene batholiths intrude the Central range, and a deformed Precambrian-Cretaceous sedimentary pile builds the Eastern Cordillera (Gómez et al., 2025). High fault densities and strong rheological contrasts (Pulido, 2003), especially along the Romeral shear zone, localize deep-seated instability (Herrera-Coy et al., 2023).

2.1.4 Climate and Hydro-meteorological Regime

Colombia's equatorial position places the Andean region under the influence of the Intertropical Convergence Zone (ITCZ), which oscillates between 5° South latitude in January–February and 14° North in July–August (Urrea et al., 2019). This movement results in bimodal rainfall patterns, with two wet seasons occurring from April to May and September to November. Annual rainfall varies significantly across the region, ranging from 1 000 mm to over 4 000 mm per year in some regions.

The Andean region is also influenced by atmospheric circulation patterns over the Atlantic and Pacific Oceans and the Amazon and Orinoco basins (Hoyos et al., 2018). Localized phenomena, such as strong topographic gradients, induce atmospheric circulation that enhances deep convective systems, leading to highly intense and localized storms. These intense rainfall events are the primary trigger for landslides, flash floods, and debris flows in the region.

Interannual climate variability is significantly affected by the El Ninõ-Southern Oscillation (ENSO), which can lead to prolonged dry periods (El Ninõ) or increased rainfall (La Ninã) (Poveda et al., 2011).
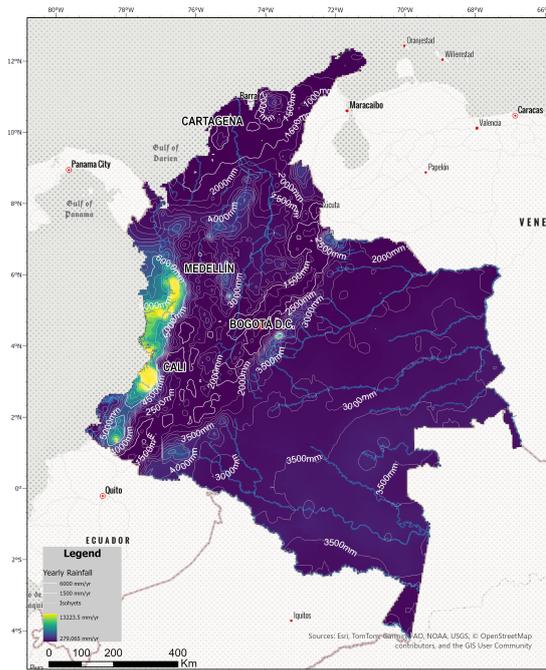
(a) Mean annual precipitation over Colombia (1981-2020, CHIRPS). white isolines at 500 mm intervals. Intervals highlight the Pacific hyper-humid belt (> 6 000 mm), and dry Orinoco < 1 500 mm.
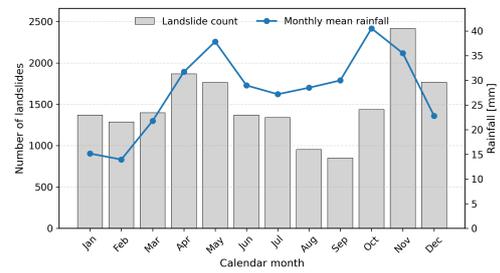
(b) Calendar month landslide counts and mean rainfall.
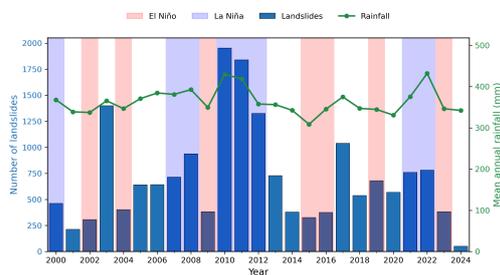
(c) Yearly landslide counts and mean precipitation, shaded by ENSO phases.

Figure 2.2: Spatio-temporal context for rainfall-induced landslides in the Colombian Andes. Data source: CHIRPSV2, CHIRPSV3 (Funk et al., 2015). (a) shows the spatial distribution of mean annual CHIRPS precipitation with key isohyets. (b) displays the seasonal cycle of landslides and rainfall. (c) presents interannual variability with ENSO bands.

2.1.5 Strategic road network and landslide exposure

Colombia's public road system extends for around 206 000km, and it is officially classified according to the *Plan Maestro de Transporte Intermodal* (Ministerio de Transporte, 2019), that hierarchize into:

- **Primary (National) corridors** are managed primarily by *INVIAS* and the National Infrastructure Agency (*ANI*). Some of the new corridors are managed by private companies. These primary roads are 10 500km paved, and carry the bulk (80 %) of the logistic and transportation flows (World Bank, 2016).

- **Secondary (Departmental) connectors** is around 34 000 km on mixed surfacing, it is usually one line, but some of them could also be two lanes. Those roads are usually administered by provincial governments.

- **Tertiary (Rural) feeder roads**: This is the majority of the road network in the country. This category counts approximately 142,000 km, its material could be gravel or earth surface, and it is usually administered by

communal or municipal organizations (Departamento Nacional de Planeación, 2021).

Even though the road network is extensive, most of the transportation lies in a couple of roads that absorb the majority of importance, due to they concentrate most of the population, commerce routes, and of course, landslide risk:

Table 2.2: Key trunk roads crossing the study area. Traffic figures are 2022 average daily traffic (ADT); landslide counts are events recorded in SIMMA for 2000-2023 around each alignment.

Route (Invías ID)	Strategic role	Class[†]	ADT [veh day^{-1}]
Medellín – Bogotá (RN 60/45)	Links the main industrial pole (Aburrá Valley) with the capital. It carries containerized imports from Cartagena	1	12 800
La Virginia – Quibdó (RN 29)	Sole paved outlet from the Pacific Ocean to the Coffee Axis. It presents frequent landslides closures	3 → 4	2 600
Cali – Buenaventura (RN 40)	Gateway for ∼40 % of Colombian trade. It climbs 2000m in <45km around the mountain range of the Andes	1	13 500

The road corridors listed above illustrate the dual challenge faced by the Colombian road infrastructure: high traffic exposure and extreme geomorphic sensitivity. Additionally, INVIAS' incident report logs portal mention that landslides were the single largest cause of unscheduled closures, accounting for 57% of all disruption hours on the national network in 2022 (Instituto Nacional de Vías (INVÍAS), 2023) aligning with our earlier discussion where SIMMA records ∼9 000 landslides that have affected the road infrastructure country–wide since 1950.

2.2 Data and Pre-processing steps

Effective landslide forecasting, particularly within data-driven frameworks, necessitates the integration of diverse and high-quality datasets. These datasets encompass landslide inventories; dynamic variables, such as precipitation patterns; and static predictors, including topographic, geological, land cover, and infrastructural attributes. We present a summary of the static datasets in Table 2.3 that we will be discussed in the following sections. The reliability and resolution of these datasets are paramount, as they directly influence the accuracy and applicability of predictive

models in complex terrains like the Northwestern Colombian Andes.

As a complementary side note, all spatial datasets were reprojected to the *WGS 84 / UTM zone 18N (EPSG: 32618)* to maintain consistency across analyzes. The datasets are categorized into three primary groups:

- **Landslide Inventories:** Compilation of historical landslide occurrences from multiple sources, offering insights into spatial distribution and temporal trends.
- **Precipitation Data:** Inclusion of both historical and forecasted rainfall data, essential for understanding triggering factors and long-term factors.
- **Static Predictors:** Encompassing topographic data, land cover classifications, geological maps, and road network information, these predictors aid in assessing susceptibility and exposure.

Table 2.3: Static covariate layers

Dataset	Native res.	Class	Notes
NASADEM	30 m (1")	Raster	Void-filled SRTM
ESA WorldCover v200	10 m	Raster	11 LC classes, epoch 2021.
South America Geology	1:5M	Vector	Simplified lithologic units.
Colombia Geology	1:1M	Vector	Quaternary units merged into the above map.
ICDE Road Network	1:100K	Vector	Filtered by *Road Type*

2.2.1 Landslide inventories

Colombia is unusually rich in landslide data, yet those data sit in four very different provenance streams that we are going to describe in the following sections. We summarized the information in The Table 2.4 that is the main input of this research. If the reader wants to check a more complete description and characterization, please see Appendix A.

National Geological Service (SGC)

The official source of landslide information is curated by SGC, which has two complementary products:

- **SIMMA Catalogue** is the fastest updated product from SGC, it contains a crowd-enabled interface that can be fed by any third party, including municipal disaster offices, media, and individuals. Its geometry is points located in space and time, with the time of the event. This inventory does not have field validation and *no* trigger or causal information, but it contains the highest volume of information among all the sources (35 244 records).
- **SIMMA Inventory** is the field validated subset; for that reason, it is not updated regularly, contains 9 170 point records and 200 mapped polygons. Its attributes are rich in characterization, and each event stores

three causal fields (trigger, contributory, inherent) and the most complete impact data that includes fatalities and USD losses for ∼30 % of events.

Road-authority feeds — INVIAS

INVIAS is the road management authority in Colombia, a government agency responsible for the construction, maintenance, and regulation of Colombia's national public road network. Inside its repositories, there are two different datasets:

- **Critical Sites** is an internal dataset that needs to be requested; it contains expert-assigned points where recurrent landslides have occurred. It contains 1 700 records with coordinates. However, there is some misalignment between the projected coordinates recorded, and it seems they are not uniform for each entry, making its positional accuracy uncertain and highlighting the necessity to do extra pre-processing steps for using this product.
- **Emergency Logs** is accessed from a public dashboard[1]. Its geometry is unique among the inventories, since it is presented as a line geocoded along specific road segments for every disruption. Its date range is from 2019 to 2023, but it lacks systematic entries for the months of January and December. The dates are precise to the hour and minute when the interruption started and ended, but the corresponding landslide source area is unknown. When this database is downloaded, all its attributes are hard-coded on numerical factors that make its interpretation difficult without any additional documentation and metadata.

University geoportals — GeoHazards

The National University's GeoHazards group in Medellín is built upon DESINVENTAR and SIMMA repositories, after its aggregation started to update daily with reports from social media, municipal logs, and newspapers (Aristizábal et al., 2025). It contains 2 products that contain basically the exact attributes but differ in some details:

- **National** only includes events that has an impact recorded. The impact is recorded on a notes attribute without any specific clear architecture
- **Antioquia** is only localized on the province of Antioquia, but it has a rich repository for the entire province using different local municipalities' sources. This database fills a gap that existed in the other repositories for this specific region.

Research-grade compilations

Researchers have compiled different landslide sources for landslide modeling. Two products highlight:

- **Gómez et al. (2023a)** initially contains 3 536 entries from different sources, including both SIMMA products, DesInventar, and GeoHazard. The efforts are maintained ongoing by the researcher, and on the day there is

a curated database of 16,780 landslide points. The dataset does not contain trigger information for most of the events, and does not discriminate between the sources. It spans from 1981 to 2023.

- **Garcia-Delgado et al. (2022)** produced a 2 400-fatal-only inventory from different sources and not usually used by landslide research: libraries, old newspapers, UNGRD, and municipalities' internal databases, and it goes back to 1828 until 2022. The only limitation of this database is that it is only available in a PDF format, which makes its interoperability difficult without any pre-processing efforts for extracting the data.

Finally, there is one dataset *DESINVENTAR*, we decided not to classify it in any of the categories because it is no longer in operation. Previously, it was the inventory used for UNGRD. This government organization decided to move its disaster inventories into a proprietary one; today, that inventory is not public. The main inconvenience of the DESINVENTAR database is that this inventory does not report hazard events, but impacts, and it does not contain any coordinate attribute, only location descriptions. Many inventories described in this section, like Aristizábal et al. (2025) and Gómez et al. (2023a), have made efforts to spatialize this inventory and avoid losing its rich information. The Table 2.4 summarizes this discussion.

Table 2.4: Core inventories used in this study. "Geo." = geometry type (Pt = point, Poly = polygon, Line = polyline).

Dataset	# events	Span	Geo.	Trigger	Provenance / salient notes
SIMMA Catalogue	35 244	1492-2024	Pt	—	Rapid, unverified crowd + agency feed
SIMMA Inventory	9 170	1900-2024	Pt+Poly	3-field	100 % field validated; richer impact fields
GeoHazards National	2 427	1880-2024	Pt	partial	University compilation, multi-source fatal+economic
GeoHazards Antioquia	5 079	1880-2024	Pt	partial	High-density departmental layer, fills spatial gap
INVIAS Critical Sites [2]	1 700	2018-2023	Pt	—	Road hotspots; CRS uncertain
INVIAS Emergency Logs [3]	706	2019-2023	Line	n/a	Linear disruptions along public road network
Gómez et al. (2023a)	16 780	1981-2023	Pt	yes	Andean merge of SIMMA, DesInventar, Geohazard. No source discriminated
Garcia-Delgado et al. (2022)	2 352	1828-2020	Pt	yes	Fatalities-only inventory, confidence rated, pdf format
DESINVENTAR [4]	10 532	1921-2017	Pt	mixed	Press & government reports; municipality centroids

---

[1] https://hermes2.invias.gov.co/SIV/
[2] Email requested
[3] https://hermes2.invias.gov.co/SIV/
[4] https://db.desinventar.org/

2.2.2 Climatological and rainfall datasets

Understanding when, where, and how much water is coming to a hillslope is the cornerstone of any dynamic landslide-forecasting framework. The Table 2.5 summarizes these datasets. The daily to sub-monthly rainfall grids described below, together with the monthly *Oceanic Niño Index* (ONI), constitute the *dynamic factors* of the landslide probabilistic model. They inform both the short-term trigger pulses and the longer preparatory wetting phases whose combined influence has been shown to improve space-time predictive skill (Moreno et al., 2024; Steger et al., 2023)

CHIRPS v2.0 historical rainfall

The *Climate Hazards InfraRed Precipitation with Stations* (CHIRPS) v2.0 product delivers near-global rainfall fields (50° S - 50° N, ≈ 90 % of the Earth's land area) from 1981-present at 0.05° (∼5 km) resolution, blending IR with 40 000 rain gauges (Funk et al., 2015). This is a daily product released with a one-month lag. We extracted the 2000-2024 daily stacks and resampled to 30 m by `nearest-neighbor` to align with the other raster information spatial resolution.

CHIRPS–GEFS 15 day precipitation forecasts

The second product we are using is the CHIRPS–GEFS, which is a bias-corrected forecast product applied to the NOAA GEFS v12 ensemble and downscaled onto the same 0.05° grid, yielding reliable 1-15-day outlooks and released every day at 0:00 UTC. Two problems were identified during quality control: (i) February 29th is absent on all the leap years, and (ii) a provider outage removed most records for Jan-Sep 2020. Both gaps are retained as `NA` values and handled downstream by the wet–filter logic explained in the Section 3.3.3.

IDEAM gauge-derived rainfall composites

To benchmark the satellite products, we adopt the 1 km daily gauge mosaics for 2022–2023 created by Gómez et al. (2023a) and interpolated using Voronoi polygons (Schumann & Polygon, 1998). These rainfall rasters were tested on a landslide modeling application using random forest on the referenced study with a great performance (AUC=0.88). The rasters are upscaled to 30 meters (nearest-neighbor) for exact overlay with other raster products.

Oceanic Niño Index (ONI)

As discussed before ENSO variability is a strong control for rainfall variability in our study area. We created a database extracting the reported Monthly ONI value anomalies in the Niño 3.4 from the Climate Prediction Center (CPC) archive[5]. Including ONI will allow the model to capture ENSO–conditioned rainfall interannual variability,

---

[5]https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php

and it has proved to be a strong predictor in the country (Gómez et al., 2023a) and aligns with the recommendations from Steger et al. (2023) of capturing seasonal and interannual climate characteristics for landslide modeling.

Table 2.5: Dynamic forcing data layers used in this study.

Dataset	Native res.	Span	Key fields	Ref.
CHIRPS v2.0	0.05° / daily	1981-2024	P_{day}	Funk et al. (2015)
CHIRPS-GEFS	0.05° / daily	2000-pres.	$P_{+1\ldots15}$	—
IDEAM kriged	1 km / daily	2022-2023	P_{gauge}	Gómez et al. (2023a)
ONI v5	1 mon	1950-pres.	SST anomaly	CPC

2.2.3 Topography and static derivatives

Digital Elevation Model

The NASA Digital Elevation Model (NASADEM) offers improved elevation data at a 30–meter resolution. It is a reprocessed version of the Shuttle Radar Topography Mission (SRTM) data, featuring refined corrections and fewer voids (Crippen et al., 2016), particularly beneficial in mountainous regions (Nguyen et al., 2023; Uuemaa et al., 2020) like the Colombian Andes. Because hydrological consistency is critical for hydrological terrain derivatives and a strong control on shallow-landslide initiation, we first conditioned the raw DEM with the open-source *RichDEM v2.2* library (Barnes, 2018). RichDEM fills true depressions while carving spurious "digital dams". This DEM is the reference raster used for all the re-projected and resampling pre–processing steps.

Surface cover and human footprint

Reliable land-cover data is essential for capturing natural and anthropogenic controls on shallow landslides. The ESA WorldCover 2021[6] dataset provides global land cover information at a 10–meter resolution (Zanaga et al., 2022). It classifies land cover into 11 categories, offering detailed insights into vegetation, urban areas, water bodies, and more. Among all other land cover products, we selected this one for its spatial resolution that allows capturing the small but subtle changes near the road network. During exploratory modeling, all 11 legend classes were tested, but only the `Built_up` category displayed a statistically significant positive effect on landslide occurrence. Consequently, the land-cover information was simplified to a binary indicator that flags the presence of built-up terrain within each modeling unit.

---

[6]https://developers.google.com/earth-engine/datasets/catalog/ESA_WorldCover_v200?hl=es-419

Lithology

Regional-scale landslide models need lithological information that is (i) geologically meaningful yet (ii) statistically robust. Having many attributes tends to fragment the study area, leaving many probability units with only a handful of pixels per class and inflating model uncertainty; conversely, over-generalized maps may hide contrasts in rock strength and weathering that dominate slope stability. To balance those needs, we merged two publicly available sources:

- **Geological Map of South America:** It is a continent-wide map compiled at a scale of 1:5M that already aggregates lithology units into a manageable number of rock classes (Tapias et al., 2023).
- **Colombian Geological Map:** Provides detailed geological information specific to Colombia, including Quaternary (Q-Xx) unconsolidated deposits that are absent from the continental map but critically important for shallow failures in the Andes (Gómez et al., 2025).

The two vector layers were clipped to the country outline, rasterized at 30 m with nearest-neighbor resampling, and mosaicked. We then collapsed the 12 source codes into five macro-classes described in the Table 2.6. This "middle-ground" legend follows recommendations that favor mechanical homogeneity over purely genetic criteria.

Table 2.6: Reclassification of source lithological units used in the susceptibility model

Source map	Original description	Merged class
South America	Siliciclastic sedimentary rocks	Siliciclastic
Colombia	Unconsolidated Quaternary deposits	Quaternary deposits[a]
South America	Granitic, gabbroic	Plutonic Igneous
South America	Basaltic, andesitic, volcano-sedimentary	Extrusive Igneous / Volcano-Sedimentary
South America	Low-medium & medium-high grade metamorphics	Metamorphic

[a]No Quaternary class exists in the continental legend; national map supplies it.

2.2.4 Road Network Data

The classification presented for Ministerio de Transporte (2019) in Section 2.1.5 lacks physical interpretation for landslide modeling because it does not define the level of intervention of the road. That is why we are using the classification from the national spatial layer published by the Colombian Spatial Data Infrastructure[7] (Infraestructura Colombiana de Datos Espaciales; ICDE). The dataset includes various attributes, with the Type of Road being partic-

---

[7]https://www.icde.gov.co/node/131

ularly relevant and detailed in Table 2.7, whose eight codes combine (i) pavement condition, (ii) platform width, and (iii) year-round accessibility. We kept that attribute in our modeling approach.

Table 2.7: ICDE national road-type codes and the English labels used in this study. Surface and width thresholds follow the official specification.

Code	Label used	ICDE definition (translated)
01	Paved highway	Concrete or asphalt pavement; > 5.5 m platform; engineered drainage, signage and protective works; open all year except during severe weather.
02	Unpaved dual-lane	Gravel/earth surface, ≥ 5.5 m wide; no major engineering works; trafficable year-round.
03	Paved single-lane	Hard surface ≤ 5.5 m; one effective lane; open all year.
04	Unpaved single-lane	Gravel/earth surface ≤ 5.5 m; one lane; trafficable year-round.
05	Plate-track	Discontinuous concrete "placa-huella" strips for very low-volume tertiary roads (mainly cars & motorcycles).
06	Wheel-track	Earth track without gravel; driveable only in dry weather; used by vehicles, animals or pedestrians.
07	Rural path	Narrow foot/pack-animal path, unsurfaced; no vehicular traffic.
08	Urban alley	Narrow paved or gravel passage in built-up areas, designed for pedestrian flow.

3 RESEARCH METHODS

This section outlines the methodological approach to achieve the research objectives and answer the corresponding research questions. The methodology is structured according to the three sub-objectives: data compilation and harmonization, statistical modeling, and real-scenario testing analysis.

3.1 Compilation and integration of landslide inventories

The first phase details the methods employed to achieve the first sub-objective of this study, which was to compile and integrate existing landslide inventories from the Andean region of Colombia into a unified, high-quality database suitable for spatio-temporal predictive modeling. The methodology addressed three research questions (RQ1.1-RQ1.3) focusing on evaluating differences among inventories, integrating datasets into a unified inventory, and exploring how damage information could be linked to landslide occurrences.

3.1.1 Characterization and Comparison of Landslide Inventories

The first step involved a qualitative characterization and comparison of the available landslide inventories. Initial exploratory analysis was conducted using eight datasets described in Section 2.2.1. This process aimed to systematically document and compare key aspects of each inventory.

A comparative matrix was developed, see Appendix A, with the following categories:

- **Provenance and custodianship:** Information on database ownership, original data sources, update frequency, last revision date, and access conditions.
- **Spatial Definition:** Details regarding geometry type (point, line, polygon), and regional scope of the database.
- **Temporal definition:** Coverage of start and end dates, temporal granularity, and the presence of auxiliary time fields such as event duration or reporting delay.
- **Landslide characteristics:** Inclusion of landslide typology, triggering information, descriptive variables (e.g., material, movement, volume), and impact indicators (fatalities, economic loss, infrastructure affected).
- **Data quality indicators:** Presence of completeness flags, confidence ratings, metadata availability, and documented mapping or validation techniques.

The matrix not only guided the harmonization strategy that follows, but also exposed recurring weaknesses—most inventories lacked polygon geometries, contained patchy time-stamp metadata, or omitted confidence flags—mirroring the heterogeneity reported by earlier cross-inventory audits (Guzzetti et al., 2012; Reichenbach et al., 2018; Van Westen et al., 2008).

3.1.2 Inventory Integration and Standardization

From the eight sources characterized earlier, four were retained for merging—*SIMMA Inventory*, *SIMMA Catalog*, *GeoHazards Antioquia*, and *GeoHazards Colombia*. They were chosen because they (i) provide the greatest spatio-temporal coverage of the Andean region, (ii) add non-overlapping records (minimal duplicate sourcing), and (iii) together contribute the largest volume of events with usable attribute fields, and (iv) it was traceable where its records came. The key attributes maintained are described in the Table 3.1 These criteria is inspired by completeness and representativeness guidelines proposed by Galli et al. (2008) and Hervás and Bobrowsky (2009).

The four datasets were merged through a four-step framework:

S1. **Schema pruning** - retain only attributes present in $\geq$75 % of the four inventories, described in the Table 3.1 (Merge_ID, Source, Landslide_ID, Event_Date, Dept, Municipio, X, Y, Movement_Class, Movement_Subtype, Cause, Reporter_Org, Reliability_Flag, Impact_Binary, Econ_Loss).

S2. **Categorical translation & reclassification** translate all text fields to English and re-code Movement_Class and Movement_Subtype to the Cruden-Varnes lexicon (Cruden & Varnes, 1996). English labels facilitate reproducibility and cross-study comparison.

S3. **Cause hierarchy** - when multiple causes are listed, apply the six-step rule: rainfall trigger $\rightarrow$ rainfall contributory $\rightarrow$ other trigger $\rightarrow$ other contributory $\rightarrow$ inherent $\rightarrow$ unknown. This preserves the distinction between trigger (initiating event) and contributory factors, and prioritizes rainfall consistency with our study focus.

S4. **Reporter re-grouping** - harmonize Reporter_Org into smaller groups of institutional classes. Records flagged with high uncertainties in the Reliability_Flag attribute were discarded owing to uncertain geolocation accuracy.

Table 3.1: Attribute glossary for the merged landslide inventory.

Attribute	Purpose	Type / Format
Merge_ID	Primary key after de-duplication	Integer (auto-increment)
Source	Origin inventory code	String
Landslide_ID	Original event identifier	String
Event_Date	Occurrence or first-report date	Date
Dept	Colombian departamento	String
Municipio	Municipality	String
X, Y	EPSG 32618	spatial
Movement_Class	Cruden-Varnes category	String
Movement_Subtype	Detailed morphology / material	String
Cause	Dominant trigger	String
Reporter_Org	Institution class	String
Reliability_Flag	Confidence reported	string
Impact_Binary	Any reported impact?	Boolean
Econ_Loss	Direct economic loss	USD

3.2 Completeness evaluation and data quality assessment

Evaluating the completeness of integrated datasets involved spatial, temporal, and attribute completeness metrics. For the attribute completeness metrics we assign a simple binary presence flag on each of the database columns and calculate the percentage of present data. On this section we will focus on describing the methodologies used for spatial and temporal completeness.

3.2.1 Spatial completeness

The spatial representativeness was evaluated with a two-dimensional Gaussian Kernel-density estimate (KDE) on a 1×1 km grid. For the kernel bandwidth h, the radius that controls smoothing, we selected it with *Scott's rule* that uses the number and spread of mapped landslides to give a bandwidth wide enough to reveal broad patterns but narrow enough to avoid random speckle, and it does so almost instantly (Terrell & Scott, 1992), making it ideal for large national-scale inventories.

Because a pixel was first binarized to "landslide present/absent", multiple points falling in the same 1 km cell counted as a single presence, thus the density reflects the spatial spread of mapping, not point crowding. We computed a separate KDE for each of the four inventories, then linearly rescaled each surface to the range 0–1 for comparison purposes. The resulting dimensionless surface ranges from ≈1 (highly represented terrain) to ≈0 (little or no reporting) and serves as a first-order indicator of inventory spatial coverage.

To ensure completeness is only evaluated where landslides are geomorphologically feasible, we multiplied each

KDE by a binary mask that removes trivial terrain: DEM cells classified as flat by `r.geomorphon` and whose local slope is $< 5°$ (Jasiewicz & Stepinski, 2013). This slope cut-off follows the "flat lowlands" exclusion used in continental landslide inventories to avoid penalizing geomorphically inactive zones (Steger & Glade, 2017). By constraining the analysis to susceptible slopes, we ensure that low KDE values reflect mapping gaps rather than stable geomorphic settings.

3.2.2   Temporal completeness: the Advanced Temporal–Uncertainty and Completeness Score (A–TUCS)

For decades, landslide inventory development has focused on where landslides occurred: spatial accuracy, developing elaborate rules for mapping scale, positional precision, and geomorphic representation (Guzzetti et al., 2000; Malamud et al., 2004). While such spatially detailed inventories are essential for creating reliable susceptibility models (Guzzetti et al., 2012), their value for dynamic applications like landslide forecasting models and operational EWS is significantly constrained if the temporal component is unreliable (Guzzetti et al., 2020; Kirschbaum et al., 2010).This temporal uncertainty fundamentally compromises key tasks: model calibration, validation of triggering thresholds, and EWS performance assessment. Three persistent temporal-completeness problems exacerbate these limitations:

(i) **Recency bias** — Contemporary risk reflects current slope stressors (e.g., infrastructure development, deforestation). Inventories lacking events from the past 10–15 years systematically underrepresent modern triggers, leading to hazardous "blind spots" in susceptibility assessments (Gariano & Guzzetti, 2016; Glade, 2003; WMO, 2022).

(ii) **Temporal clustering** — Post-disaster response surges create "event bursts" in inventories, distorting rainfall threshold calculations and EWS lead-time estimates. Recording lapses during quite periods compound this bias (Froude & Petley, 2018; Piciullo et al., 2018).

(iii) **Archive depth deficit** — are high-magnitude triggers (e.g., megathrust earthquakes, multi-decadal climate anomalies) require ≥50-year records for robust return-period analysis and threshold calibration (Parker et al., 2017). Most inventories lack this historical coverage.

Because no single scalar captures recency, evenness, and historical depth at once, we introduce the Advanced Temporal-Uncertainty & Completeness Score (A-TUCS). A-TUCS deliberately decomposes the problem into (i) recent coverage, (ii) clustering penalization, and (iii) historical longevity. Their mathematical definition and parameter logic follow below so that readers can reproduce - or adapt - the metric for other hazard inventories.

Component A: Recent coverage

Landslide inventories face a fundamental tension: while historical landslides provide valuable context, their relevance decays as landscapes evolve through deforestation, urbanization, and climate shifts. This component resolves this tension by systematically weighting events based on their age, prioritizing recent occurrences while still crediting

older data. The 20-year threshold emerges as a critical pivot point in this weighting scheme, aligning with reporting changes in the benchmark inventory (SIMMA) from Colombia, and with the inclusion of new technologies for collecting and reporting hazard events. This timespan captures significant anthropogenic changes – new infrastructure, mining operations, and land-use patterns (Gariano & Guzzetti, 2016; Glade, 2003) – while remaining within the memory window of local communities and institutions.

The time-decay function applies distinct exponential decay rates before and after this 20–year threshold, acknowledging that pre–2005 datasets fundamentally differ from post-smartphone–era environments where crowd-sourcing and remote sensing should theoretically eliminate data gaps (Guzzetti et al., 2012). For any landslide occurring in the present year Y_i within an inventory ending at Y_{last}, the weight w_i follows:

$$
w_i = \begin{cases} e^{-\lambda_{\text{recent}}\,(Y_{\text{last}}-Y_i)}, & Y_{\text{last}} - Y_i \leq T_{\text{threshold}}, \\[2ex] e^{-\lambda_{\text{old}}\,(Y_{\text{last}}-Y_i)}, & Y_{\text{last}} - Y_i > T_{\text{threshold}}. \end{cases}
$$

The gentle decay rate ($\lambda_{\text{recent}} = 0.05$) for recent years preserves the diagnostic value of modern slope disturbances – a highway cut in 2020 reveals more about current stability than a 1970s rockslide. Conversely, the rapid decay ($\lambda_{\text{old}} = 0.20$) for older events reflects landscape obsolescence, and changes in data collection, older inventories are expected to have less data and we don't want that the index got affected by this normal behavior of hazard datasets. The half-life metaphor helps visualize this: recent events maintain half their weight for 14 years, while pre–2005 events decay to half-value in just 3.5 years.

A critical insight guides our treatment of monthly counts: a single typhoon month with 100 landslides should not outweigh years of sparse but diagnostically vital activity. To prevent such distortion, we adapt the Michaelis-Menten saturation principle from enzyme kinetics (Michaelis, Menten, et al., 1913). The transformation:

$$
\phi(\text{Counts}_i) = \frac{\text{Counts}_i}{\text{Counts}_i + k}
$$

serves as an "event value compressor." The denominator constant $k = 2$ was calibrated through sensitivity analysis to ensure that:

- A solitary landslide in a month retains significant value ($\phi(1) = 0.33$)
- Two events achieve half-credit ($\phi(2) = 0.50$)
- Major outbreaks (>5 events) approach but never reach full credit ($\phi(10) = 0.83$)

This design acknowledges the value of having a lot of sparse events more than a bunch of events in one month. Seasonal effects and interanual changes can only be capture across the time (Steger et al., 2023), documenting that landslides occurred in a given month matters more than precisely how many. The saturation prevents monsoon seasons from artificially inflating scores while ensuring quiet periods aren't neglected.

The final coverage score C_R synthesizes these principles through weighted normalization:

$$\text{Coverage} = \frac{\sum_{i=1}^{n} w_i\, \phi(\text{Counts}_i)}{\sum_{i=1}^{n} w_i}$$

The numerator represents the "effective coverage" - each month's saturation-adjusted count scaled by its temporal relevance. The denominator, the sum of all possible weights, establishes the theoretical maximum. This ratio inherently penalizes inventories missing recent years: even a historically complete catalog ending in 2010 would score below 0.3 today, as the weights for 2011-2023 (which should constitute >60% of the denominator) would be entirely absent. The parameterization allows changing the thresholds for initiation year, finalizing year, and threshold in case the researcher wants to apply this index into specific landslide inventories, geographic locations or other hazards datasets

The 20–year dual–decay framework forces confrontation with landscape memory and technology changes. The saturation function ensures this temporal sensitivity isn't compromised by media-amplified disaster months. Together, they create a metric where scores below 0.5 indicate dangerously obsolete inventories, while scores above 0.8 require consistent modern documentation –precisely the behavior needed for reliable dynamic modeling. This proposed levels needs to be adjusted in other studies.

While this recency weighting exposes gaps in temporal coverage, it remains blind to another critical distortion: the clustering of reported events during post-disaster response surges. Such 'reporting bursts' create statistical artifacts that mislead rainfall threshold calibration. To penalize this clustering effect, we now introduce Component B: the temporal evenness metric or cluster penalization.

Component B: the temporal evenness metric

Landslide inventories suffer from a hidden distortion we will call the "crisis attention cycle". When major disasters strike, researchers, media coverage, NGOs, and government programs intensifies and emergency funding floods in, creating concentrated bursts of reporting activity appears. This is a trend that can be followed with the number of studies published after a big event; some examples are the Gorkha earthquake in Nepal in 2015 (Kargel et al., 2016; Marc et al., 2019; McAdoo et al., 2018; Roback et al., 2018); The Mocoa disaster in Colombia in 2017 (Cheng et al., 2018; García-Delgado et al., 2019; Peñas, 2017; Vargas-Cuervo et al., 2019); or the Emilia Romagna heavy–rainfall induced landslides in Italy in 2023 (Berti et al., 2025; Dotta et al., 2023; Ferrario & Livio, 2023); and many other examples. Yet during quieter periods when landslides still occur, documentation often lapses due to diminished attention. This problematic clustering pattern distorts rainfall thresholds due the lack of seasonality and variability along the trained databases, and as a Steger and Glade (2017) mentioned thresholds changed with climate patterns.

To quantify and penalize this temporal distortion, we adapt the Gini coefficient—a century-old metric originally developed by Gini (1921) to measure wealth inequality in economics. This robust framework has since been repur-

posed across disciplines: economists use it to identify income disparity, ecologists employ it to diagnose species clustering in fragmented habitats (Bai et al., 2023), and used in the geospatial domain to measure spatial concentration (Rey & Smith, 2013). Our innovation lies in applying this established framework to measure temporal clustering in event reporting, revealing inequality along and inventory.

Before applying the Gini framework, we must first address a fundamental challenge in landslide data: raw event counts exhibit extreme variance following the "crisis attention cycle" we already discussed, and it is a expected behavior, that can mathematically overwhelm the metric. A single month with 100 landslides shouldn't overshadow ten months with 9 landslides each, though both represent valuable information. Following established practice in statistics, we apply a logarithmic transformation to the monthly counts. This transformation, expressed as $x_i = \log(\text{Counts}_i + 1)$, serves three critical purposes. First, the +1 adjustment gracefully handles months with zero landslides by ensuring we never take the logarithm of zero. Second, it compresses extreme values–transforming a 100–landslide month to approximately 4.6 and a 10–landslide month to 2.3, thereby preventing orders of magnitude differences from dominating the analysis. Third, it preserves meaningful distinctions within the typical operational range of 1–20 events per month that matter most for Early Warning Systems.

The core of our temporal evenness metric then applies the Gini framework to these transformed counts, weighted by their recency importance from Component A. The complete equation takes this form:

$$
G = \frac{\sum_i \sum_j w_i w_j \left| \log(\text{Counts}_i + 1) - \log(\text{Counts}_j + 1) \right|}{2\,\mu_{\log,w} \left(\sum_i w_i \right)^2},
$$

where:

$$
\mu_{\log,w} = \frac{\sum_i w_i \log(\text{Counts}_i + 1)}{\sum_i w_i}.
$$

In the Table 3.2 we interpret the possible results of the coefficient and how they can be interpreted.

Table 3.2: Interpretation guide for temporal Gini coefficient values

Gini (G)	Interpretation	Real-World Analogy
0.0	Perfect evenness	Landslides documented steadily across months
0.2	Moderate clustering	Typical inventory with minor reporting spikes
≥0.6	Dangerous clustering	Post-disaster surge dominates record (e.g., 2018 Hiroshima rains)

This formulation deserves careful unpacking. The double summation in the numerator $\left(\sum \sum w_i w_j |x_i - x_j| \right)$ acts as an inequality magnifier—it systematically compares every possible pair of months in the inventory, calculates the absolute difference in their log-transformed event counts, then weights those differences by the temporal relevance of each month. Conceptually, this measures the total "temporal tension" in the inventory caused by un-

even reporting patterns. The denominator serves as a normalization factor, centered around the weighted mean of transformed counts (μ_w), which anchors the inequality measurement to the inventory's overall event density. The denominator's structure ensures the resulting Gini coefficient always falls between 0 and 1, regardless of inventory size or timespan.

The practical interpretation of the resulting G value follows clear thresholds, as summarized in Table 3.2. When G approaches 0, it indicates near-perfect temporal evenness—landslides documented steadily across months without artificial clustering. A moderate score around 0.2 reflects typical inventories with minor reporting spikes but generally acceptable distribution. Values exceeding 0.6 signal dangerous clustering that requires correction before using the inventory for threshold calibration.

To illustrate why this penalty structure works, consider a hypothetical "supercluster" scenario: After a 100-year storm triggers 1,000 landslides, international organizations and government agencies thoroughly document these events. Five years later, a researcher requests the landslide inventory and discovers that 92% of events remain concentrated in just three post-disaster months, with sparse coverage elsewhere despite available satellite data. The logarithmic transformation prevents these three massive months from dominating the metric, while the pairwise comparisons in the numerator amplify penalties for the stark contrasts between disaster months and quiet periods. Critically, the recency weighting ensures a 2020 cluster impacts the score more severely than a 1990 cluster would, maintaining alignment with Component A's philosophy.

While this Gini-based penalty effectively diagnoses and penalizes clustering within an inventory's observation window, it cannot assess whether that window spans sufficient time to capture rare megastorms or multi-decadal climate patterns. For this essential dimension of historical depth, we now introduce Component C: the historical longevity bonus.

Component c: the historical longevity bonus

While Components A and B address contemporary temporal quality, truly robust landslide hazard assessment requires historical perspective. Long observation windows serve as irreplaceable foundations for understanding landscape behavior—only inventories spanning decades can capture the full spectrum of destabilizing triggers. These include multi–decadal climate oscillations like ENSO that alter regional rainfall patterns over 20–30 year cycles, and rare high-magnitude earthquakes (Mw > 7) whose seismic legacies may reshape slope stability for generations. As documented in the Colombian Andes, such low-frequency events fundamentally modulate landslide regimes yet remain invisible in short-term records (Garcia-Delgado et al., 2022; Sepúlveda & Petley, 2015).

The longevity bonus explicitly rewards such temporal depth through a carefully balanced equation:

$$\text{Longevity} = \underbrace{\alpha}_{\text{Balancing Parameter}} \cdot \underbrace{\frac{\ln(\text{duration} + 1)}{\ln(101)}}_{\text{Depth Reward}} \cdot \underbrace{\exp\left(-\frac{\text{median gap}}{10}\right)}_{\text{Gap Penalty}} \qquad (3.1)$$

Depth Reward mechanism   Human-modified landscapes evolve nonlinearly, making additional years of data more valuable during early inventory development than in mature archives. Consider two examples:

- Extending from 5 to 15 years captures transformative changes (new highways, deforestation, urban expansion)
- Extending from 85 to 95 years adds marginal insights when earlier periods already document long-term patterns

The natural logarithm ($\ln$) reflects this diminishing returns principle—it strongly rewards initial timespan extensions while preventing century-old data from dominating scores. The +1 adjustment ensures even single-year inventories receive minimal credit rather than undefined values. Normalization by $\ln(101)$ scales the output to a 0-1 range relative to a 100-year reference baseline.

Gap Penalty Mechanism   Duration alone cannot guarantee quality—an inventory with a 40-year span but 30 years of silence remains scientifically inadequate. The exponential gap penalty addresses this by penalizing extended reporting voids while tolerating short pauses:

- A single 10-year silence reduces the bonus by 63% ($e^{-1} \approx 0.37$)
- Ten separate 1-year gaps reduce it by only 9.5% ($e^{-0.1} \approx 0.905$ per gap)

This structure aligns with empirical evidence that short voids can be filled through targeted archive searches (Garcia-Delgado et al., 2022; Herrera et al., 2018), while extended silences indicate systemic documentation failures. Using the median gap rather than the mean ensures robustness against outlier pauses.

The Balancing Parameter   The α coefficient caps longevity's maximum contribution at 10% of the total score. This critical safeguard prevents historical depth from masking deficiencies in contemporary documentation, a century old inventory with poor recent coverage cannot score above 0.10 without improvement.

The Integrated A-TUCS Framework

The complete Advanced Temporal-Uncertainty & Completeness Score synthesizes all three dimensions into a unified metric:

$$\text{A--TUCS} = \underbrace{\text{Coverage}}_{\text{Recency}} \times \underbrace{(1 - G)}_{\text{Clustering Penalty}} + \underbrace{\text{Longevity}}_{\text{Historical Depth}} \tag{3.2}$$

This structure creates a symbiotic relationship between components while preventing any single aspect from dominating:

1. The $(1 - G)$ term acts as a clustering corrector—even inventories with excellent coverage (Component A) are penalized if events cluster in post-disaster bursts (Component B). This ensures rainfall thresholds derived from the inventory won't be artificially inflated by media-amplified events.

2. Longevity (Component C) contributes additively rather than multiplicatively. It provides standalone credit for historical depth without requiring perfect modern documentation. This acknowledges that century-old landslides still offer value for understanding long-term patterns, even if contemporary records are sparse.

3. The $\alpha = 0.10$ cap ensures longevity remains a "bonus" rather than a primary driver-no inventory can exceed 1.0 without strong performance in both recency and evenness.

Innovation and contextual limitations   As the first metric to decompose temporal completeness into recency, evenness, and depth, A-TUCS addresses a critical gap in geohazard inventory assessment. By adaptively weighting components-prioritizing recent documentation while crediting historical context—it provides a more nuanced evaluation than simple "oldest inventory wins" approaches. Its modular approach allows flexibility an adaptation into different contexts. Nevertheless, three limitations warrant consideration:

1. The 20–year decay threshold reflects Colombian landscape dynamics but may need adjustment in rapidly changing urban or glacial environments.

2. The Gini component assumes landslides are small-scale events; inventories dominated by other physical processes may require modified clustering metrics.

3. While validated for regional Colombian inventories, global application requires testing in diverse settings including crowd sourced databases.

Through this framework, we enable quantitative comparison of the landslide inventories used in this research, revealing how temporal weaknesses invisible to conventional can be assessed. Future research should link this index into two different directions, (i) how the temporal weaknesses can propagate into dynamic probabilistic model performance and (ii) balancing datasets across the time for a more evenness training data.

3.3   Development of a spatio-temporal predictive landslide model

This section details the methodology for developing a spatio-temporal predictive landslide model for the Andean road network. The model integrates dynamic rainfall data and static environmental factors, datasets that were already introduced in the Section 2.2 within a GAM framework.

3.3.1   Spatial Framework and Mapping units

The spatial units of analysis are slope units (SU), which are geomorphologically homogeneous regions bounded by drainage and divide lines (Alvioli et al., 2016). Slope units were automatically delineated using the `r.slopeunits`

algorithm (Alvioli et al., 2016) applied to a 30-meter resolution DEM without trivial terrain areas (Steger & Glade, 2017), where landslides are not expected. For that purpose, we used the slope mask developed before in Section 3.2, and we removed the permanent water bodies.

The key parameters and post-processing operations for SU delineation are in the Table 3.3 defined after iteration and texting of different set of parameters, we invite the reader to check Alvioli et al. (2016) for more information about the parameters. The Slope Unit approach balances the need for spatial detail with computational feasibility and aligns with best practices in landslide susceptibility modeling (Carrara et al., 1991; Eeckhaut et al., 2009).

Table 3.3: Key *conceptual* parameters adopted for the automatic slope-unit delineation workflow and post-processing steps (Alvioli et al., 2016)

Parameter (conceptual name)	Value used
Drainage-area threshold	$500\,000$ m^2
Minimum slope-unit area	$12\,500$ m^2
Minimum coefficient-of-variation (relief)	0.30
Reduction factor applied	5
Maximum number of refinement iterations	5
Minimum polygon area kept after cleaning	$10\,000$ m^2
post-processing steps	
Conversion of cleaned raster to polygons	`r.to.vect` (type=area)
Generalization of polygons limits	`v.generalize` (type=area)

A 100-meter buffer was applied to Colombia's national road network. Only slope units whose overlap this buffer were retained, resulting in approximately $152\,310$ slope units with an average area of 0.43 km^2. This focus on road-proximal units ensures the model's relevance to road risk management.

3.3.2 Co-factors aggregation

All predictors—terrain, hydrographic, lithologic, and rainfall—were therefore aggregated from their native raster or vector representation to the SU polygons prior to modeling. The candidate factors aggregated are described in Table 3.4.

Table 3.4: Static candidate covariates screened in the univariate–GAM step.

Variable	Description	Type	Source (res.)	SU aggregation
area_su	Polygon area of slope unit (m^2)	cont.	SU polygons	`v.to.db area` (sum)
slope_average	Mean slope angle (°)	cont.	30 m DEM	`v.rast.stats` `method=average`
slope_stddev	Std. dev. of slope angle (°)	cont.	30 m DEM	`method=stddev`
LR_range	Local relief (max-min elev., m)	cont.	30 m DEM	`method=range`
VRM_average	Vector Ruggedness Measure (mean)	cont.	derived VRM (30 m)	`method=average`

Continued on next page

Table 3.4 (continued)

Variable	Description	Type	Source (res.)	SU aggregation
VRM_stddev	Vector Ruggedness Measure (sd)	cont.	derived VRM (30 m)	`method=stddev`
TWI_average	Topographic Wetness Index (mean)	cont.	GRASS `r.topidx`	`method=average`
TWI_stddev	Topographic Wetness Index (sd)	cont.	GRASS `r.topidx`	`method=stddev`
profCurv_average	Mean profile curvature ($\times 10^{-3}$)	cont.	30 m DEM (`r.param.scale`)	`method=average`
profCurv_stddev	Profile curvature (sd)	cont.	idem	`method=stddev`
planCurv_average	Mean plan curvature ($\times 10^{-3}$)	cont.	30 m DEM (`r.param.scale`)	`method=average`
planCurv_stddev	Plan curvature (sd)	cont.	idem	`method=stddev`
rainfall_average	25-yr mean annual rainfall (mm yr^{-1})	cont.	CHIRPS 2.0 (5 km) $\rightarrow$ res. 30 m	`method=average`
river_sum	Total rasterised river pixels	cont.	1:500 k network	`v.rast.stats sum`
river_length	Stream length within SU (m)	cont.	idem	pixel-count $\times$ 30 m
river_density	Stream length / area (m m^{-2})	cont.	idem	derived
Categorical factors				
Road_maj	Road type (7 classes)	cat.	1:100 k road lines	mode per SU $\rightarrow$ **random effect**
Aspect_maj	Dominant 8-point aspect	cat.	derived aspect	circular mode
litho_group	Lithological group (5 classes)	cat.	1:500 k geology	mode per SU
river_presence	Binary: river intersects SU	cat.	river network	presence/absence
Landcover_maj	ESA land-cover main class (11)	cat.	ESA WorldCover 10 m	mode per SU
LC_TreeCover	Binary: tree cover present	cat.	as above	indicator
LC_Grassland	Binary: grassland present	cat.	as above	indicator
LC_BuiltUp	Binary: built-up present	cat.	as above	indicator

Elevation-derived rasters were computed from the 30 m NASADEM with standard GRASS modules (`r.slope.aspect`, `r.param.scale`, `r.vector.ruggedness`, `r.topidx`). For each raster, cell values intersecting a given SU were summarized by the statistic best suited to the geomorphological meaning of the index: *mean* and *standard deviation*, and *range* for local relief. Distance-to-stream, stream length, and density were obtained from the 1:500k hydrographic IGAC network using `r.stream.distance` and zonal summaries. Categorical layers (road type, lithology, 8-point aspect, ESA land-cover) were rasterized at 30 m and the modal class within each SU was recorded.

Daily rainfall grids from (i) CHIRPS 2.0 and (ii) CHIRPS-GEFS ensemble forecasts (both 0.05°) were *nearest-neighbor* resampled to 30m—preserving original values while ensuring at least one rainfall pixel falls inside every SU. For each product, we built a time-ordered stack $P_0 \ldots P_{30}$ (where P_0 is the day of the landslide and P_{30} is thirty days before) and derived cumulative antecedent and trigger totals.

All aggregated variables were finally merged into a single attribute table and exported to *gpkg* format used for the GAM modeling workflow. Landslides where aggregated on a binary form (landslide presence–absence) duplicating SU attributes and creating one new entry for each landslide date and ignoring.

3.3.3 Sampling Design and Wet-Filter

Of the 152 310 SU intersecting the national road buffer, only 7 501 contained dated landslide events (presences). To address the extreme class imbalance while maintaining spatiotemporal representativeness in our absence sampling, we implemented a stratified random sampling procedure following the methodology of Steger et al. (2023) and refined by Moreno et al. (2024). This three-tiered balancing approach ensured that absences were distributed:

- **Spatially balanced:** Across different geographic locations

- **Annually balanced:** Proportional representation across years

- **Seasonally balanced:** Proportional representation across months

Rather than sampling from the prohibitively large daily spatiotemporal domain (which would create computational inefficiency and potential temporal bias), we first assigned random multiple replicates per location across the study period. We then iteratively refined this selection to enforce balanced annual and monthly distributions. This rigorous stratification process yielded 24 900 carefully curated absence cases that accurately represent stable conditions throughout the study area. The final merged dataset (2000–2024) comprises 32 401 SU: 7 501 presences and 24 900 temporally–stratified absences.

Model skill depends critically on excluding spatial units (SU) during dry periods where landslides are extremely unlikely, retaining only SU-days under 'potentially rainy' conditions (Moreno et al., 2024). To implement this systematically, we developed a precipitation uncertainty filter that accounts for known systematic biases in satellite rainfall products. As documented in South America and globally (Bai et al., 2018; Shrestha et al., 2017; Wu et al., 2019), satellite estimates consistently deviate from ground observations. We quantified these biases using the high-resolution IDEAM gauge raster developed by Gómez et al. (2023a) as ground truth, comparing daily rainfall values across our study area during 2022-2023.

For each satellite product, we calculated the daily pixel bias as $Bias = Rainfal_{Estimated} - IDEAM_{Observation}$ This produced a distribution of daily differences from which we derived two key statistical parameters:

- μ as the mean systematic bias
- σ as the standard deviation of saily errors

Then The wet filter threshold (T_{wet}) was then calculated as:

$$T_{\text{wet}} = |\mu| + \sigma \tag{3.3}$$

SU–days were retained only when satellite estimates exceeded T_{wet}, ensuring rainfall signals were sufficiently strong to overcome product-specific errors. Complete bias statistics are provided in Table 3.5.

The dry bias of CHIRPS V2 relative to IDEAM gauges ($\mu = -2.18$ mm; Table 3.5) mirrors earlier evaluations for the Tropical Andes (López-Bermeo et al., 2022) where the CHIRPS estimations are underestimated. By contrast, the CHIRPS–GEFS ensemble shows a wet bias ($\mu = +1.30$ mm) but a smaller spread, in line with regional validation studies of GEFS precipitation (Lien et al., 2016; Yue et al., 2022).

Table 3.5: Daily bias statistics used to set the *wet filter* from Equation (3.3).

Product	Mean (mm)	Std. dev. (mm)	$T_{\mathbf{wet}}$ (mm)	Gauge days (n)
CHIRPS V2	-2.18	4.84	7.02	730
CHIRPS-GEFS	1.30	3.08	4.38	730

With *wet filter* in place, we applied a complete filter to reconcile the five-hour time-zone mismatch between IDEAM gauges (UTC-05) and the gridded rainfall products (UTC). We aggregated a 48-h cumulative rainfall $\text{precip}_{48h} = P_0 + P_1$ on each SU and applied the following filters:

1. Removed records with missing dates;
2. Excluded leap-day entries (29 Feb);
3. Excluded the January-September 2020 gap where CHIRPS-GEFS fields are unavailable;
4. Discarded rows with missing rainfall in 48 hours;
5. Retained observations with $\text{Precipitation}_{48h} > Wet_{filter}$.

The results after applying the filter are summarized in the Table 3.6 and it can be observed on how our final datasets are distributed across months in Figure 3.2 that grantees seasonal precipitation changes is well captured in our model; annually in figure 3.1 for interanual variability.

Table 3.6: Samples retained after successive filters.

	Before filter	CHIRPS V2	CHIRPS-GEFS
Presences	7 501	4 537	6 354
Absences	24 900	11 365	18 180
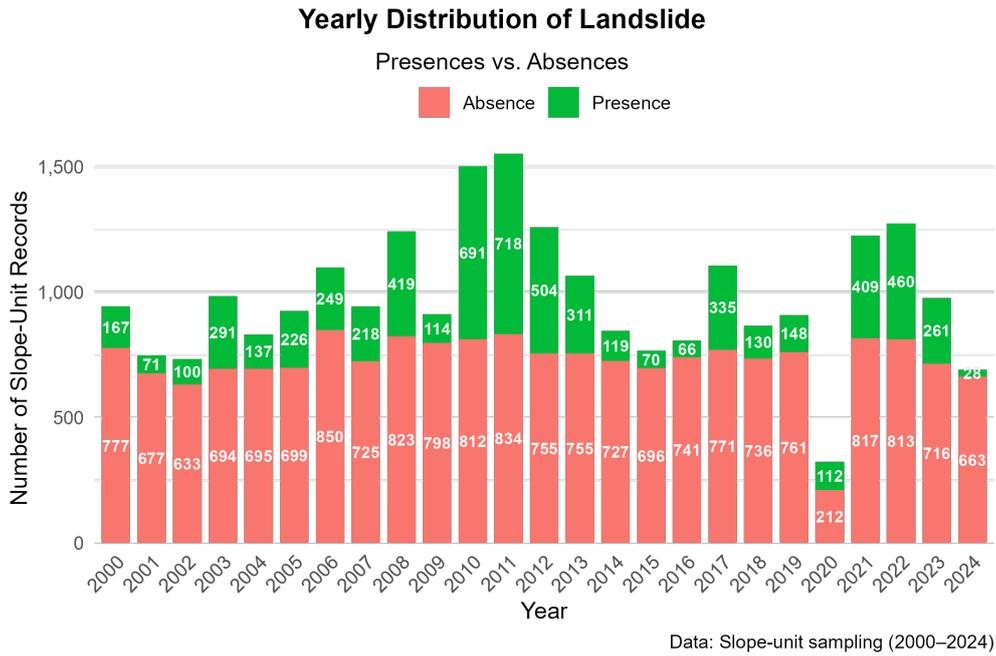
(a) CHIRPS-GEFS presences (green) and absences (red) per year
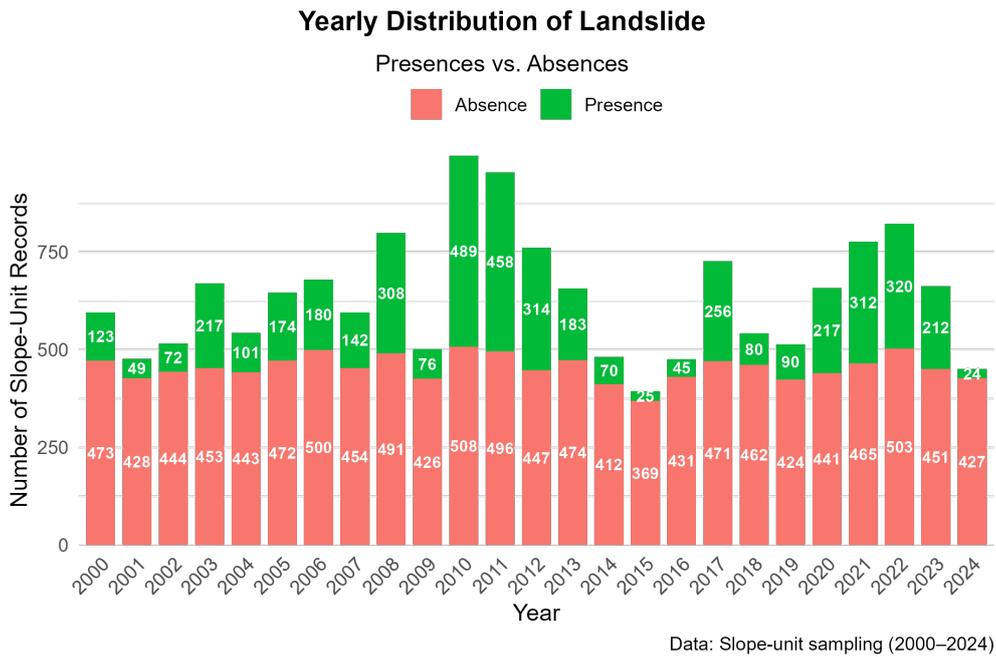
(b) CHIRPS V2 presences and absences per year

Figure 3.1: Year-by-year distribution of presence/absence samples after applying the wet filter.

Monthly Distribution of Landslide

Presences vs. Absences
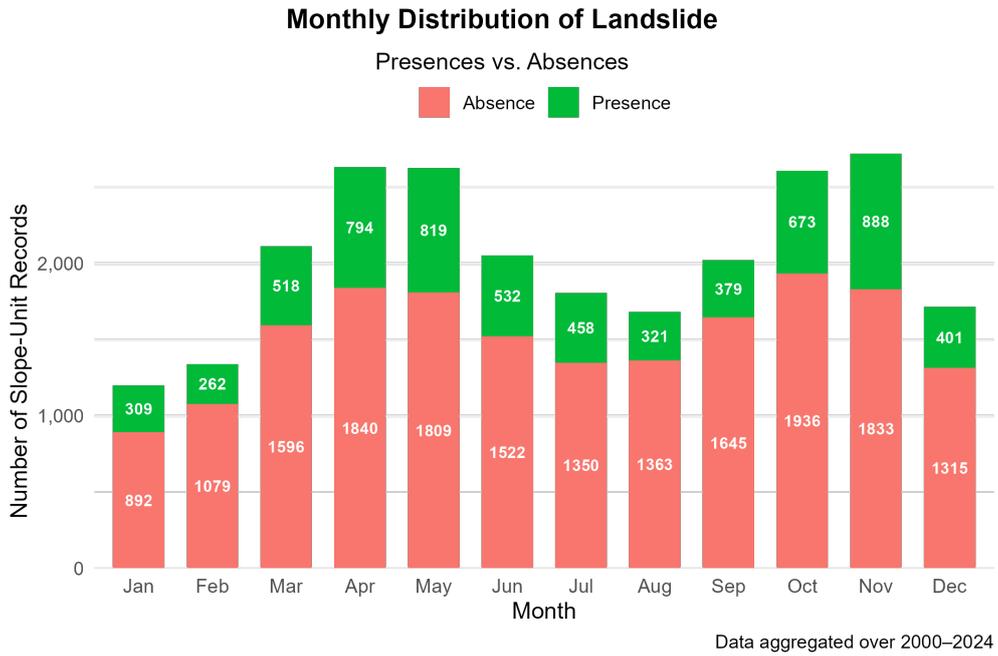
(a) CHIRPS-GEFS monthly distribution

Monthly Distribution of Landslide

Presences vs. Absences
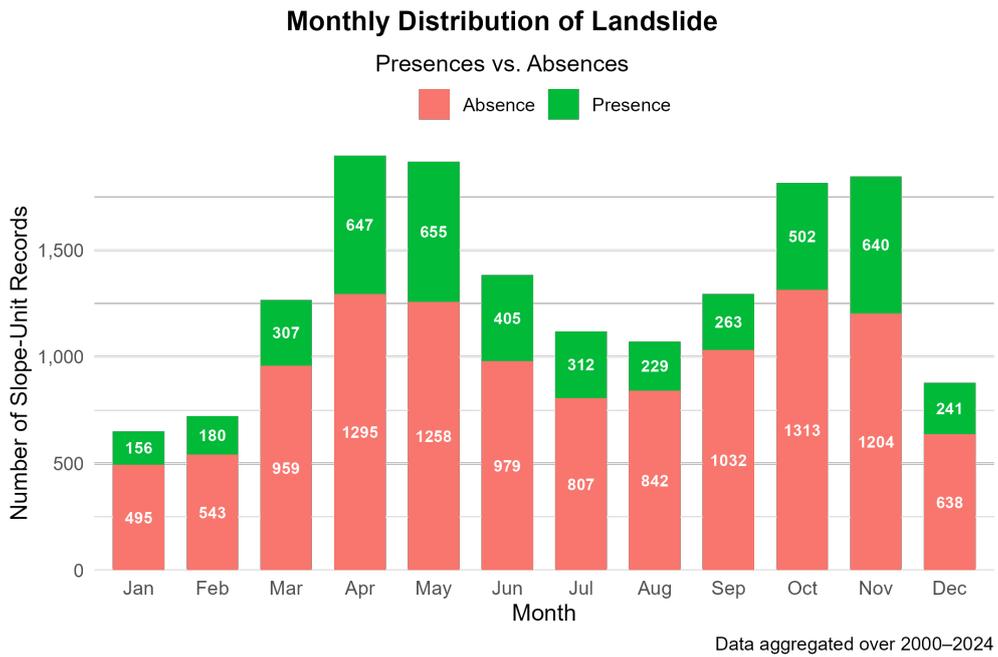
(b) CHIRPS V2 monthly distribution

Figure 3.2: Monthly distribution of presences and absences retained for modeling.

3.3.4 Predictive Framework

The spatio-temporal landslide model was developed using a binomial GAMs (Hastie, 2017; Wood, 2017). We adopted GAMs because they offer the rare combination of (i) proven predictive skill for landslide models(Moreno et al., 2024; Wang et al., 2024) and with environment variables (Pedersen et al., 2019), (ii) an additive structure that lets each co-factor be visualized and interpreted in isolation(Ramampiandra et al., 2023), and (iii) mature, peer-reviewed software with its library in R, `mgv` (Wood, 2017), that supports spatial smooths, tensor interactions and mixed-effect terms. GAMs fitted the formula

$$
\underbrace{\mathrm{logit}\{p_i\}}_{\text{linear predictor}} = \underbrace{\beta_0}_{\text{intercept}} + \underbrace{\sum_{j=1}^{J} f_j(x_{ij})}_{\text{smooth functions of continuous drivers}} + \underbrace{\sum_{k=1}^{K} \beta_k\, z_{ik}}_{\text{categorical (fixed or random) effects}} + \underbrace{s(x_i, y_i)}_{\text{Gaussian–process spatial smooth}} .
$$

(3.4)

Equation (3.4) is the linear predictor of the GAM used in this study. Starting from the left, the `logit` link transforms the predicted probability (p_i) into the log-odds scale so that the model can be fitted with standard maximum-likelihood tools. *The intercept* β_0 provides a global reference level; every other term describes how the log-odds deviate from that baseline. *The smooth terms*, each continuous environmental driver–terrain, cumulative antecedent rainfall, daily trigger rainfall, day-of-year, among others are allowed their own flexible spline f_j. The thin-plate basis means we never need to pick break-points by hand; the amount of wiggle is chosen automatically by REML, and we control how much freedom we want to give to each one with parameters inside the modeling, and the resulting curves can be plotted with 95%. *The Categorical effects* are binary factors, such as river presence, that are included with ordinary dummy coefficients. Road hierarchy is different: the "trail" and "footpath" classes contain <100 slope-units, whereas "primary highway" occurs in thousands. Treating this variable as a random-effect smooth (`bs="re"`) shrinks those tiny classes towards the global mean and prevents them from monopolizing the model. *Spatial surface* explains spatial similarities that nearby SU could have or the spatial bias associated with the uneven report across the country. A Gaussian-process surface $s(x, y)$ soaks up that residual spatial autocorrelation, keeping the standard errors of the other terms valid.

Incremental modeling workflow

Step 1 – Univariate screening of static terrain factors Seventeen candidate predictors, see Table 3.4, were first fitted one-by-one using thin-plate regression splines with a maximum basis dimension of 10. Each partial-dependence curve was inspected to understand the expected respnse to landslide presence. Artifacts like flat profiles that stayed after the aggregation step were judged non-informative and those predictors where discarded..

Step 2 – Static spatial GAM   All surviving static factors were then combined on an iterative process prioritizing those that have a stronger physical meaning and could be more interpretable. Categorical attributes where tested on their confidense and importance for the model and some of them where merged. The selected co factors were merged with a two-dimensional Gaussian-process smooth s(x,y) that captures residual spatial dependence.

Step 3 – Dynamic rainfall engineering   Rain-induced slides depend on both short "trigger" bursts and multi-week antecedent wetting (Gómez et al., 2023a; Steger et al., 2023; Vega et al., 2024). We therefore created a 30–day rolling cube of daily CHIRPS values and tested every legal pair of windows where (i) trigger (0-5 days) and (ii) anticipatory (1 to 30 days) excluding the trigger rainfall (Moreno et al., 2024). Additional dynamic cofactor where included in this model like the day of the year for seasonal characteristics and ONI for long term precipitation changes (Vega et al., 2024).

Each pair of trigger and anticipatory rainfall was evaluated in an otherwise identical GAM including the complete set of dynamic factors; discrimination was scored with the median AUROC across ten 10-fold random splits. The same grid-search was repeated with CHIRPS–GEFS forecasts. To prevent the very wet Pacific coast, where few landslides are reported, from dominating the training process of the model, daily rain values above the 98th percentile were capped at 98th percentile: they were set to that percentile before accumulation. This "tail-capping" stops the model from concluding that 150 mm day^{-1} implies stability simply because those totals coincide with sparsely inhabited rainforest.

Step 4 - Building the full spatio-temporal GAM   The final model pools:

- the optimal trigger and antecedent windows (separately for analysis and forecast modes);

- the static terrain, land-cover, and lithology terms retained after Step 2;

- a cyclic response for day-of-year to absorb seasonality;

- Atensor interaction $ti(antecedentrain, ONI)$ that lets the accumulative rainfall behave differently in different ENSO stages without forcing the partial effects to run "parallel"; and

- a random-effect response for road hierarchy

3.3.5   Validation Strategy

A predictive model is only as trustworthy as the evidence that it generalizes beyond the data on which it was trained. `Out-of-sample validation` is therefore a cornerstone of any landslide model because it reveals whether apparent skill is genuine or merely the result of over-fitting to one period, one region, or one particular inventory (Brenning, 2012; Moreno et al., 2024). Two complementary, *threshold-independent*, statistics are used:

- *Area under the Receiver Operating Characteristic curve* (**AUROC**).AUROC measures how well the model *ranks* slope-units: an AUROC of 0.80 means that in 80% of all randomly drawn landslide-absence pairs, the landslide receives the higher probability (Hanley & McNeil, 1982). Because it is unaffected by the rarity of landslide events, AUROC is ideal for extremely imbalanced data sets such as ours.

- *Brier score*. Defined as the mean squared difference between predicted probability and the observed binary outcome, it captures *calibration*: a perfectly calibrated model achieves the minimum possible Brier score of 0 (Brier, 1950; Schlögl et al., 2025). Good calibration is essential for risk communication, because emergency managers need to know whether "40%" really means that two out of five similar slopes will fail.

To probe every dimension in which the model might fail, five cross-validation (CV) schemes were adopted (Table 3.7). In every split, we preserved all 24 900 absence records to maintain the true landslide / non-landslide ratio and recomputed *both* metrics; the median across repetitions is reported. All resampling was carried out in `R 4.3` with `rsample` and `blockCV`. Seeds were varied automatically ('`set.seed(NULL)`') so that each repetition uses an independent random draw.

Table 3.7: Cross-validation designs used in this study. n_{rep} indicates how many *independent* repetitions were run; the model is refitted from scratch in every fold.

Scheme	Motivation	How folds were defined	What stays constant	n_{rep}
Random 10×10	Baseline sampling variability	Uniformly assigns the 32 401 SU-day cases to ten folds; repeat ten times	All absences retained in every fold	10
Spatial k-means	Transfer to unseen terrain	10 spatial clusters derived from SU-centroid k-means ($k=10$) (Brenning, 2012); one left out each time	Cluster geometry; full absence set	10
Leave-one-month-out	Seasonal generalization	For each calendar month, *all* Januaries, all Februaries, … are withheld in turn (Moreno et al., 2024)	Year-to-year order; absences	10
Rolling-origin	Inter-annual forecasting	Training period always starts in 2000 and expands by one year; test set is year $t+1$ ($t = 2009\ldots2023$)	Earlier years remain in train	15
Inventory hold-out	New data–source robustness	One entire source inventory kept for testing while the other six feed training; cycle through all seven combinations	Presence / absence balance	7

===

3.4 Operational Implementation for Landslide Early Warning Systems

The third sub-objective quantifies how effective the predictive model, coupled with the chosen threshold, is in real case scenarios (RQ 3.1) and evaluates the actional recommendations in a context-based (RQ 3.2).

3.4.1 Threshold Selection

The binary decision cut-point is the Youden-optimal value

$$T_Y = \arg\max_{p_t} \Big[\text{TPR}(p_t) + \text{TNR}(p_t) - 1 \Big],$$

which maximizes the True-Skill Statistic $\text{TSS} = \text{TPR} - \text{FPR}$, a bias-free measure widely used in landslide early-warningRossi et al., 2019; Wang et al., 2024. On the training folds; this same τ is applied in the back-test below.

Diagnostic metrics reported at $p = \tau$:

- **TSS (= Youden's J)**: robustness to prevalenceYouden, 1950.
- **F_1**: harmonic mean of precision and recall.
- **Brier score & AUROC**: model-wide calibration and ranking.

The decision threshold was fixed at the Youden-optimal cut-point obtained from Section 4.2.3. An *alert* (actionable warning) is therefore issued for any SU-day where $P \geq \tau$.

A *hit* is recorded when an alerted SU contains at least one landslide presence on the same calendar day; otherwise, the alert is a *false positive*. Conversely, an inventory point falling in a non-alerted SU counts as a *miss*. Remaining SU-days are *correct negatives*. Daily contingency tables yield hits (TP), misses (FN), false alarms (FP) and correct negatives (TN)

3.4.2 Predicting road interruption

The back-test focuses on the shallow landslide cluster that paralyzed the Quibdó-Pereira trunk road and adjacent coffee-region corridors between **3–20 January 2023**. Official *INVIAS* traffic bulletins first reported partial closures between kilometer-posts 2-18 on 3 January, escalating to a full blockade after a 20 000 m^3 failure at km 16 on 10 January (Infobae, 2023; INVIAS, 2023). Regional newspapers and radio confirmed additional blockages near Manizales, Filandia, Apía, and Guática through mid-January (Carocol Radio, 2023; El Pereirano, 2023). Daily situation reports from the National Disaster Risk-Management Unit (UNGRD) kept the municipalities of Pueblo Rico and Santa Cecilia on *red* alert throughout 14-18 January (UNGRD, 2023a, 2023b). Taking into account the last description, we decided on a temporal scope of January 2023 around the affected area. A bounding box covering the five affected municipalities ($\sim 6{,}946$ km^2) was intersected with the SU yielding 6 277 unique SUs.

Following the rainfall-window optimization in Chapter 4.2.1, each SU was paired with (i) accumulated precipitation over the preceding 23 days and (ii) four daily trigger-rain lags (P_0-P_4) derived from CHIRPS-GEFS (Funk et al., 2023). Static cofactors and the remaining dynamic ones remained identical to those used in model training. To avoid 31 separate model calls, a long database was built with one record per *SU-day* (6 277 SUs $\times$ 31 days $=$ 194 587 rows).

The fitted GAM model was trained on the full inventory up to 31 December 2022, thus excluding all test-period events, using the framework described in 3.3.4. A single inference run produced daily landslide-probability rasters for 1-31 January 2023. For operational deployment, this routine should be scheduled once per calendar day, writing Cloud-Optimized GeoTIFFs for all Colombian corridors; this section's back-test analysis only covers the sub-extent described above.

3.4.3 From Slope Unit Probabilities to Segment-Level Risk

For each road segment and day, we derive an "expected landslide count" by treating each neighboring slope unit (SU) with probability p_i as an independent Bernoulli trial $X_i \sim \mathrm{Bernoulli}(p_i)$. The total count $N = \sum_{i=1}^{n} X_i$ follows:

$$E[N] = \lambda = \sum_{i=1}^{n} p_i, \quad \mathrm{Var}[N] = \sum_{i=1}^{n} p_i(1 - p_i).$$

For small λ, this is well approximated by $\mathrm{Poisson}(\lambda)$ We estimate λ per segment via length-weighted SU probabilities.

Estimating SU-segment intersection via buffer area

Since a center-line approach sometimes misses marginal SU intersections, we instead intersect each segment with a 20 m buffer polygon. Then each overlapping area A yields an approximate clipped line length

$$\hat{L} \approx \frac{A}{2w}, \qquad w = 20\,\mathrm{m},$$

because a buffer of width w has area $\approx 2wL$ This gives the fragment length, which we divide by the full segment length to obtain weights $w_{i,j}$. Then

$$\hat{\lambda}_{\mathrm{seg}} = \sum_i \Big(p_i \times w_{i,j} \Big).$$

Hybrid density metrics for short vs long segments

To compare segments of different lengths, we define density metrics:

$$d_{\mathrm{watch}} = \frac{m_{\mathrm{watch}}}{L_{\mathrm{km}}}, \quad d_\lambda = \frac{\lambda}{L_{\mathrm{km}}},$$

with m_{watch} = count of SUs with $p_i \geq 0.30$ and L_{km} = segment length in km. For very short segments ($L_{\mathrm{km}} < 0.5$ km), densities become unstable, so we use absolute values instead

Risk levels and thresholds

We categorize segments into four operational alert levels—Green, Yellow, Orange, Red—based on the following rules:

1. **Red:** Extreme threat. Either

$$(p_{\max} \geq 0.50 \text{ and } d_\lambda \geq 2.5) \quad \text{or} \quad (p_{\max} \geq 0.50 \text{ and } d_{\text{watch}} \geq 3).$$

   This captures either very high single-SU risk ($p \geq 0.50$) with moderate expected slide counts, or clustered moderate probabilities.

2. **Orange:** Elevated risk. Either

$$(p_{\max} \geq 0.45 \text{ and } d_{\text{watch}} \geq 0.30) \quad \text{or} \quad (p_{\max} \geq 0.45 \text{ and } d_\lambda \geq 0.20).$$

   These are calibrated to produce a manageable Orange rate while recognizing early escalation.

3. **Yellow:** Warning stage. At least one watch-SU *or* $\lambda \geq 1$.

4. **Green:** No SU above $p_i \geq 0.30$ and $\lambda < 1$. Indicates normal operations.

Threshold selection is based on a combination of Youden-optimal AUROC performance, distributional inspection, and operational meaning.

Geometry and road-type re-attachment

After calculating segment-level metrics and alert colours, all summary geometries are dropped. The original centre-line geometry and road-type attribute `TIPO_VIA` from the source `rd` layer are re-attached via a spatial join using the `seg_id` key, ensuring accurate mapping and format consistency for final outputs.

The workflow produces (i) a daily vector `.gpkg` layer with P, alert flag and observation status for every SU; (ii) color-scaled probability maps for visual briefing; (iii) color-threshold map with alert levels and (iv) a CSV summary of TP, FP, FN, TN and H-K per day. Although alert dissemination is beyond the thesis scope, these artifacts satisfy the interoperable data format outputs that enable direct ingestion into decision-making dashboards.

4 RESULTS

4.1 Inventory completeness results

This section will include a complete characterization of the 8 inventories and a description of each one of the characterization attributes, which will be presented in the Section 4.1.1. Later, the Sections 4.1.2 – 4.1.4 will evaluate the four selected inventories in the spatio-temporal domain.

4.1.1 Qualitative contrasts among inventories

Figure 4.1 distills the information in Annex Tables A.1-A.5 into a ten-column heat-map. The color grid makes two facts immediately evident: (i) attributes are decidedly uneven across the eight candidate inventories, and (ii) only four: SIMMA Inventory, SIMMA Catalog, GeoHazards Colombia and GeoHazards Antioquia score "present" (green) or "partial" (amber) in most of the methodological categories. This section unpacks those patterns, moving through the five categories described in our methods in Section 3.1.1.

Provenance & custodianship

Ownership ranges from a single agency (INVIAS) to multi-institutional sources like the Gómez et al. (2023a). Regular updating, recognized as a key quality criterion in inventory guidelines (Van Den Eeckhaut & Hervás, 2012), it is interesting to observe that irregular but post-2023 uploads are confirmed only for Gómez et al. (2023a) (annual), the two GeoHazards inventories, and the SIMMA catalog. By contrast, Garcia-Delgado et al. (2022) has not been revised since 2020, and the INVIAS lists are not being updated since 2023 (Table A.1). Access conditions diverge as well, SIMMA exports require a user request for formats different than PDF, whereas GeoHazards, Garcia-Delgado et al. (2022), and the Emergency Reports from INVIAS provide open downloads, aligning with open data recommendations.

Spatial definition

Seven inventories are point-only, while SIMMA Inventory supplements points with some mapped polygons |(200), allowing footprint-scale analyzes. GeoHazards Antioquia restricts itself to a single department, whereas the others claim nationwide coverage, as we can see in Table A.2 in the Appendix A. To highlight in this category, INVIAS
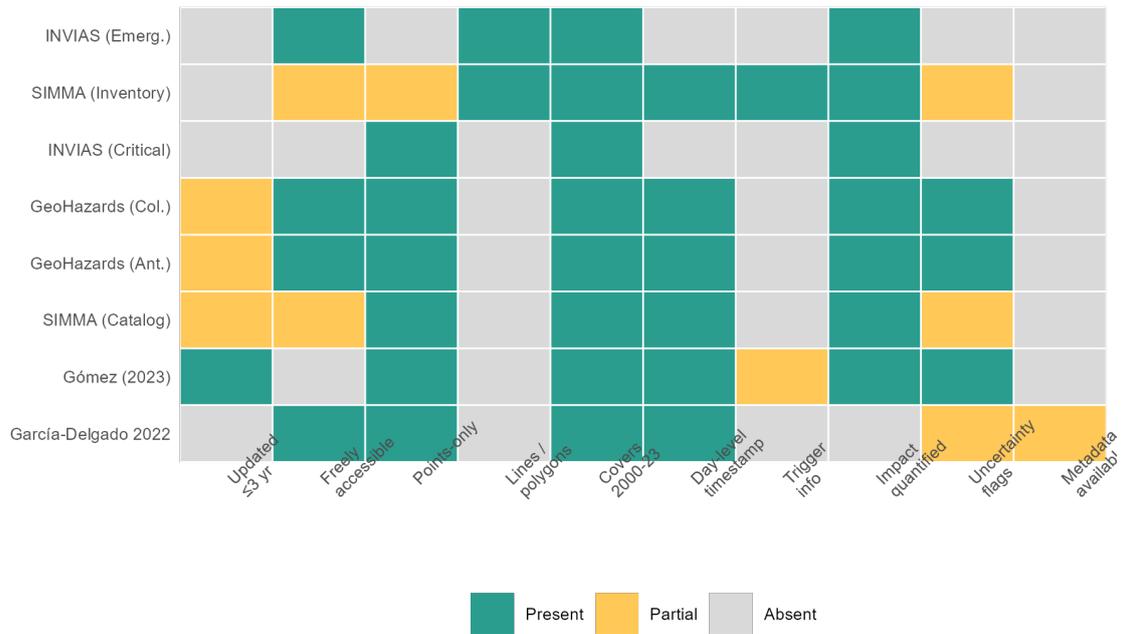
Figure 4.1: Qualitative cross-comparison of eight Colombian landslide inventories. Each coloured square indicates whether a key attribute is *Absent* (grey), *Partial* (amber) or *Present* (green). Attributes are grouped according to the methodological categories defined in Section 3.1.1, enabling a rapid visual assessment of coverage, content richness and data quality. Results highlight why subsequent modelling focuses on four inventories (SIMMA Inventory, SIMMA Catalog, GeoHazards Colombia, and GeoHazards Antioquia) that meet the minimum information and quality thresholds.

Emergency Reports uniquely uses line geometry tied to specific impacted road corridors, which is valuable for infrastructure studies. There is no possibility of linking it to a landslide event, and classifying this inventory as an impact database more than a landslide occurrence inventory.

Temporal definition

All datasets are inside the time range between 2000 and 2024, satisfying the CHIRPS-GEFS overlap, yet their historical depth varies by three orders of magnitude, as we can see in Table A.3. SIMMA Catalog extends to 1 492, echoing the long-tail archives recommended for hazard trend analysis (Gaidzik & Ramírez-Herrera, 2021), whereas INVIAS begins only in 2018. Day-level time-stamps are standard except in INVIAS Emergency (hourly). Garcia-Delgado et al. (2022) and SIMMA report multiple events on one date entry.

Content attributes

The Trigger information is critical for landslide susceptibility modeling (Sinčić et al., 2024), it is fully recorded only in SIMMA Inventory, recording up to three different causes: Triggering, Inherent, and Contributory, there is always at least one reported for each Landslide event. GeoHazards marks most events as "unknown", and the INVIAS lists and SIMMA Catalog omit triggers entirely, as we can see in Table A.4. Impact indicators range from boolean "fatalities" (Garcia-Delgado et al., 2022) to monetized damage categories in SIMMA Inventory, reported in USD. SIMMA Catalog includes five centuries of events yet lacks any trigger field, underscoring the trade-off between volume and thematic richness and highlighting the necessity to use a rainfall filter for maintaining rainfall-related landslides without losing entries as we described in Section 3.3.3.

Data-quality indicators

Explicit uncertainty flags appear in SIMMA and both GeoHazards inventories. Metadata completeness is patchy; only SIMMA provides a public schema, whereas others lack explicit documentation. The complete report is in Table A.5.

Probabilistic modeling emphasizes the necessity of having updated inventories, polygon/line capability, or high-accuracy points, day-level dating, and trigger attribution. Figure 4.1 shows that SIMMA Inventory, SIMMA Catalog, GeoHazards Colombia, and GeoHazards Antioquia are the only datasets that are close to meeting those benchmarks across all categories. Consequently, the subsequent evaluation and integration described in Section 3.2 and modeling workflows concentrate on these four datasets. Both GeoHazards inventories were merged into the following analysis for matching the evaluated spatial extension.

4.1.2   Spatial completeness

The masked 1×1 km Gaussian-KDE surfaces for the three source inventories are shown in Figure 4.2. Continuous hot-spots occur in the central Andean provinces, especially Antioquia, Norte de Santander and Cauca, where event density exceeds ~ 1 landslides km^{-2}. The western part of the country and some parts of the Eastern Cordillera, by contrast, exhibit extensive voids. This heterogeneity underscores the need for merging multiple inventories to fill geographic gaps. Spatial completeness scores from Table A.6 confirm the visual impression: **SIMMA Inventory** attains the highest coverage with 31.8 %, **SIMMA Catalog** follows with 29.4 %, whereas **GeoHazards** covers barely 10.8 %. Note that record count alone does not guarantee spatial coverage, SIMMA Inventory (9 175 records) outperforms SIMMA Catalog (35 658 records) and GeoHazards (7 519 records) in spatial representativeness.
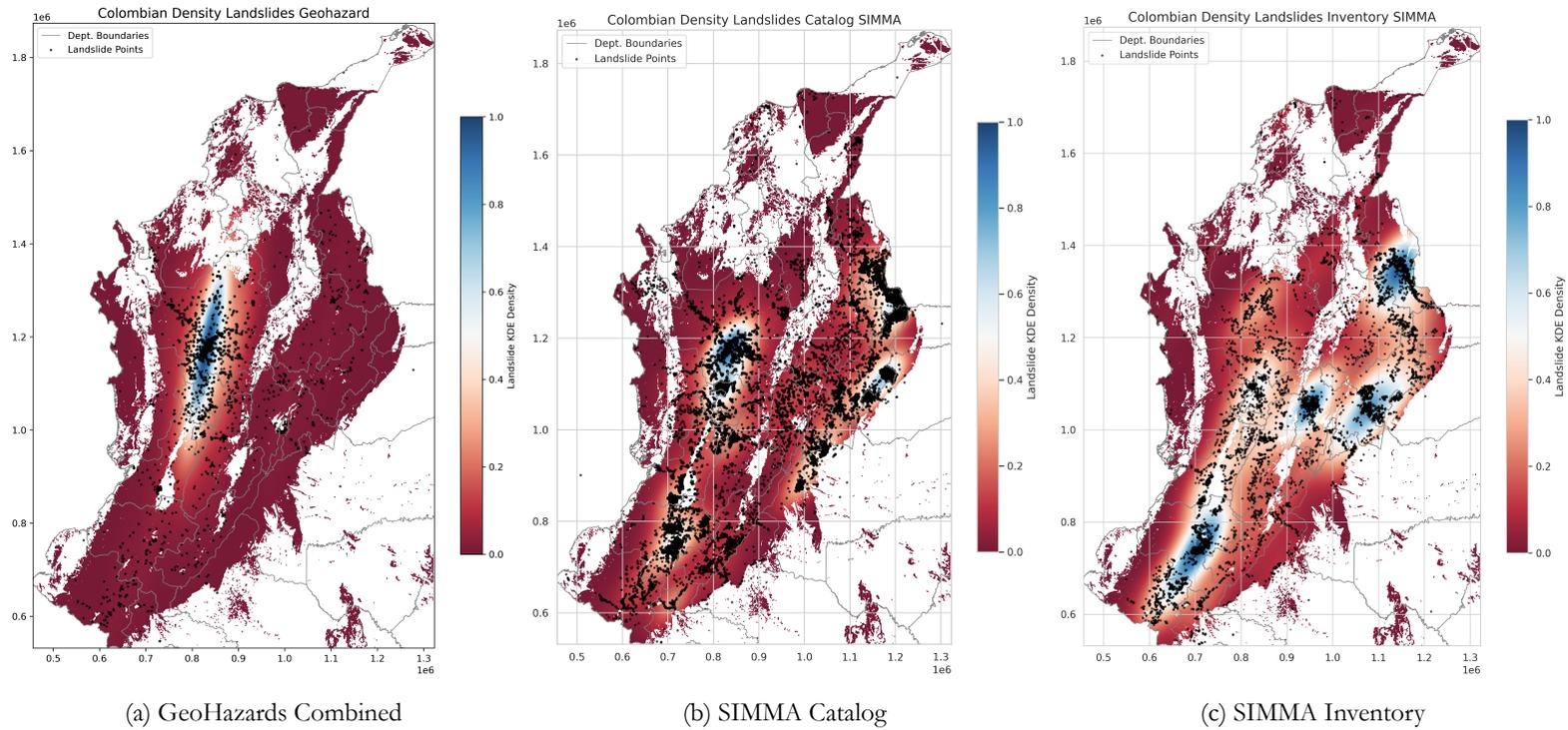
Figure 4.2: Kernel-density maps (normalized scale) for the three source inventories. Colder colors denote higher mapped-event density; white areas mark terrain masked as geomorphologically improbable for landslides. **(a)** GeoHazards Combined exhibits clustered coverage in Antioquia; **(b)** SIMMA Catalog provides broad but uneven density; **(c)** SIMMA Inventory, despite fewer records, achieves the most even coverage, particularly along the Eastern Cordillera and the south-western Andes.

4.1.3 Temporal completeness (A-TUCS)

Temporal robustness of the inventories was quantified with the Advanced Temporal-Uncertainty and Completeness Score (*A-TUCS*) see Methodology, Sect. 3.2.2. The three source inventories and the unfiltered merge differ markedly in the contribution of each component, as we can see in Table 4.1.

An overview of the presented results:

- **SIMMA Catalog** reaches the highest overall score (A-TUCS $= 0.887$), driven chiefly by its long temporal coverage (1724-2024) and low temporal clustering (Gini $= 0.143$). The archive, therefore, provides a strong historical baseline, albeit at the expense of missing trigger information, as we can see in Figure 4.3.
- **SIMMA Inventory** attains a balanced A-TUCS $= 0.734$; the shorter span (1900-2023) is compensated by better recent coverage and moderate clustering, making it well suited for calibration of daily rainfall thresholds.
- **GeoHazards Combined** shows the lowest score (0.515), reflecting episodic reporting and long silent periods. Although there is an overall completeness in other attributes, its temporal unevenness limits its direct value alone for spatio–temporal applications.
- **Merged (no filter)** improves coverage and longevity but inherits high clustering from duplicate events, still yielding a high A-TUCS $= 0.863$ which proves the capping mechanisms' effectiveness included in its formula.

Table 4.1: A-TUCS components and overall temporal completeness.

Inventory	Coverage	GINI	Longevity	A–TUCS	Temporal Scope
GeoHazards (Combined)	0.692	0.288	0.022	0.515	1880–2025
SIMMA (Inventory)	0.834	0.217	0.081	0.734	1900–2023
SIMMA (Catalog)	0.948	0.143	0.074	0.887	1724–2024
Combined Inventory - No filter	0.950	0.177	0.082	0.863	1724–2024

Figure 4.3 situates the temporal findings in a broader six-dimensional completeness context. From the Figure, we can see that SIMMA Inventory balances spatial, temporal, and trigger data; SIMMA Catalog excels in temporal completeness but omits triggers; GeoHazards contributes trigger accuracy yet remains spatially sparse. These metrics give us a more complete overview of each inventory.
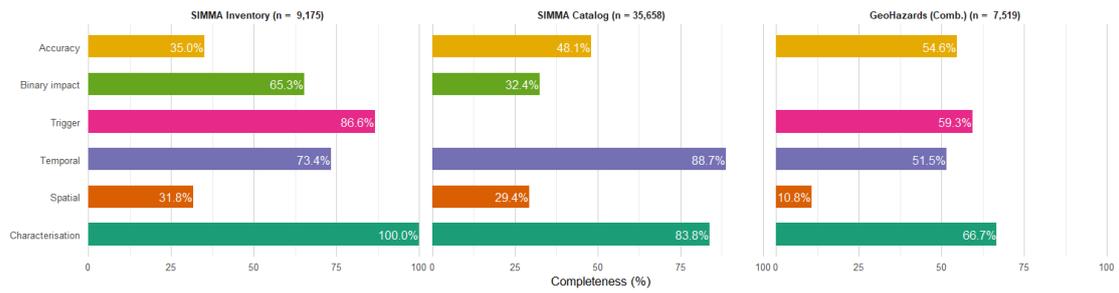
Figure 4.3: Each panel shows percentage fulfillment for six attributes; record counts appear in strip labels.

4.1.4 Consequences for integration

The contrasts observed between the datasets and the subsequent post-2000 filter are possible to see in Figure 4.4, which sacrifices trigger information in exchange for a bigger database compatible with Data-driven models. The resulting 17 824 point–event database minimizes time-dependent bias in further predictive GAM modeling, maintains the overall values on spatial completeness compared with the other inventories, and contains strong characterization attributes.

Damage variables were excluded from modeling because only the SIMMA Inventory supplies consistent quantitative fields. GeoHazards lists contain qualitative notes, and SIMMA Catalog uses a different unit scheme, as we can see in Annex Table A.7. Harmonizing these would have discarded over 70% of records, outweighing potential benefits for predictability modeling on a national scale. Future research could use the damage information to model potential impacts in smaller scales in the most dense places.
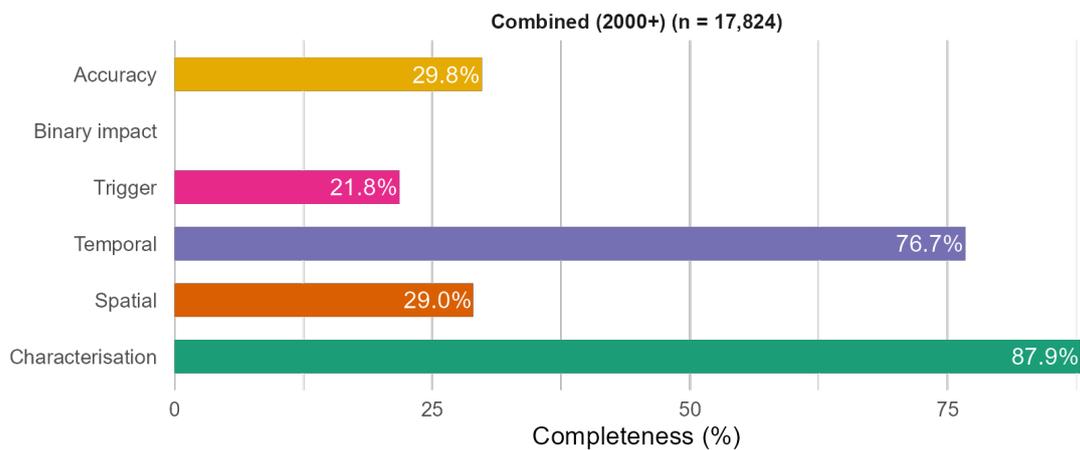
Figure 4.4: completeness of the unified inventory after the temporal 2000-2024 filter. Gains in characterization and volume are offset by reduced trigger and accuracy information.

51

4.2   Spatio-temporal landslide modeling

In this section, we will discuss (i) the rainfall windows that maximize predictive skill, (ii) the relative importance and physical interpretation of each static and dynamic co-factor in the final GAM model, and (iii) the model performance when the model is driven by the CHIRPS-GEFS rainfall dataset. And a fuller validation, including spatial and temporal transferability, and the applicability for forecasting, and threshold diagnostics.

4.2.1   Optimal antecedent-trigger rainfall windows

Figure 4.5 shows median AUROC obtained for every combination of antecedent (AR) and trigger (TR) windows. For the historical CHIRPS product the optimum is a compact $AR = 15$ d + $TR = 2$ d tile (gold square), whereas the forecast CHIRPS-GEFS product peaks at a longer $AR = 23$ d + $TR = 4$ d (Fig. 4.5b). In both cases, the two runner-up tiles differ by ≤ 0.002 AUROC, confirming a relatively flat optimum.

When static predictors and the spatial generalize are added, the difference narrows further: AUROC$= 0.78$ for the full CHIRPS model versus AUROC$= 0.77$ for the full CHIRPS-GEFS model, as we can observe in Figure 4.6.
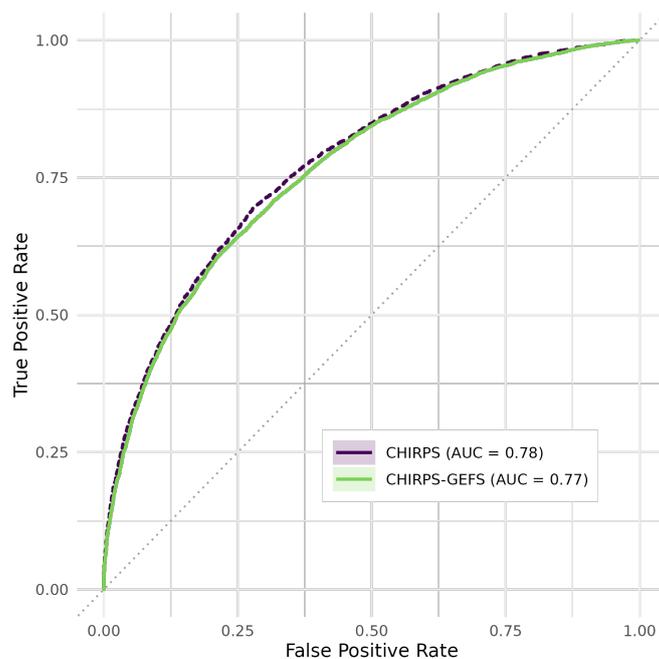
Figure 4.6: Receiver-operating characteristic for the two *full* GAMs (dynamic + static + spatial terms). Dashed diagonal = no-skill; AUC values are shown in the legend.

Precision-recall curves tell the same story, AUPRC $= 0.61$ for both, confirming that the CHIRPS-GEFS driven model preserves the same discriminatory power in the landslide presence domain as the CHIRPS driven model.
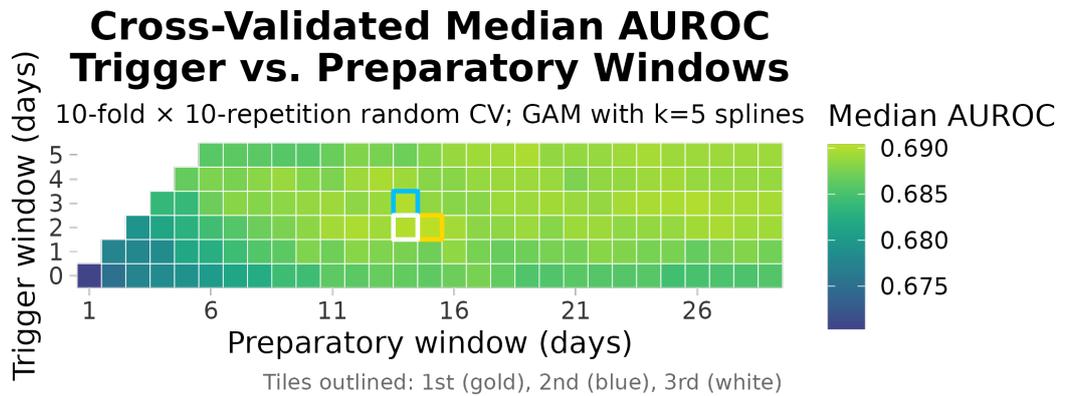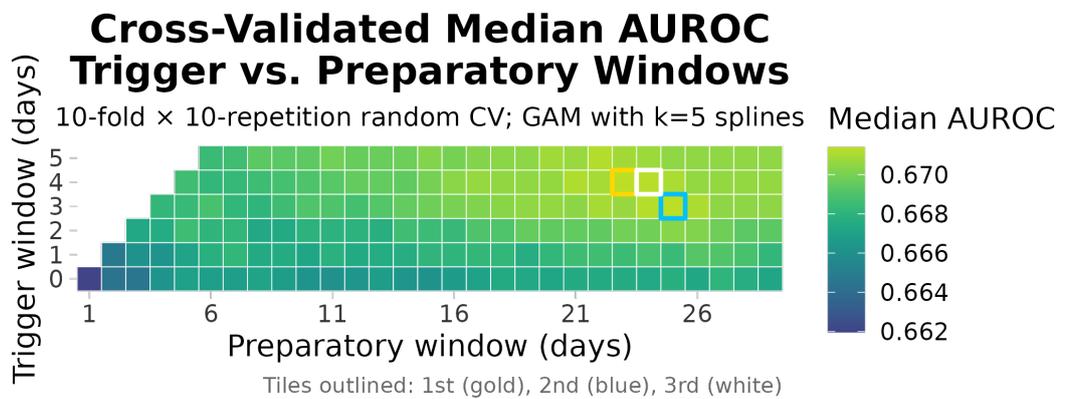
Figure 4.5: (a) CHIRPS: best tile at **AR = 15 d, TR = 2 d** (gold); the next two tiles (blue, white) differ by ≤ 0.002 AUROC. (b) Median AUROC as a function of antecedent-rainfall length (x-axis) and trigger-rainfall length (y-axis). Squares mark the three best windows for each product; warmer colors imply higher skill.

Because CHIRPS grids are released with a $\geq$1-month latency, whereas CHIRPS-GEFS fields are available in near real time, the **AR = 23d + TR = 4d** CHIRPS-GEFS configuration is adopted for all subsequent validation and forecasting experiments.

4.2.2 Which predictors matter?

Final predictor set   The nine terms retained on the final model are summarized in Table 4.2. Their contribution contributions are visualized in Figure 4.7, where we can observe that Dynamic terms together account for $\approx 33\,\%$

of the model contribution, highlighting the importance of explicitly modeling rainfall and climate variability in addition to terrain and geology.

Table 4.2: Predictors included in the final GAM.

Domain	Variable	GAM term	k/levels
Precipitati	Trigger rainfall (Ptrig)	$s(\cdot)$ (tp)	5
	Antecedent rainfall (Pacum)	$s(\cdot)$ (tp)	5
Climate	Oceanic Niño Index (ONI)	linear	-
	Pacum × ONI interaction	$ti(\cdot, \cdot)$	5×3
Seasonal	Day-of-year (doy)	$s(\cdot\,\text{cc})$	5
Terrain	Mean slope (*slope_avg*)	$s(\cdot)$ (tp)	3
	Slope-unit area	$s(\cdot)$ (tp)	3
Hydro	River presence (binary)	factor	2
Land use	Built-up land-cover (binary)	factor	2
Lithology	Litho group	factor	5
Infrastructure	Road hierarchy (*Road_maj*)	$s(\cdot)$ (re)	7
Space	Coordinates (x, y)	$s(\cdot, \cdot\,\text{gp})$	8

Road type dominated contribution (>57%) when fitted as a fixed factor. This was due to the larger number of reported landslides along the major road network, as we can see in 4.3. To mitigate this reporting bias associated with more important roads reports, we re-specified the co-factor as a random effect; this reduced the contribution share from 57% to 33%, see 4.7, while still capturing the engineering and maintenance differences across Colombia's network.
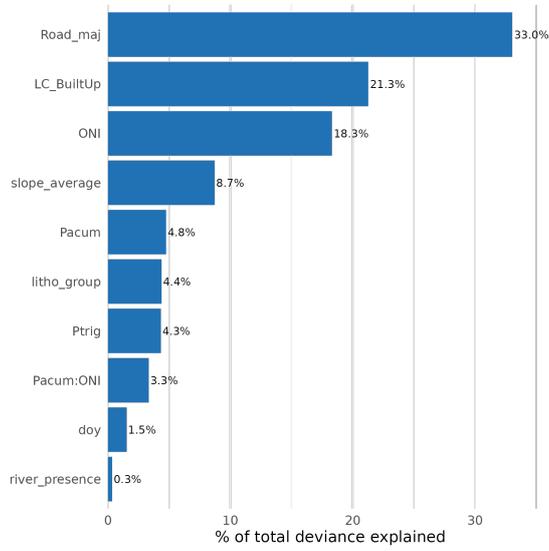
Figure 4.7: Relative contribution explained by each covariate in the final spatio-temporal GAM. Bars sum to 100 %

Table 4.3: Road-hierarchy frequencies within the 152 310 slope-units analyzed.

Hierarchy	Description	# SU with Landslide presence
Type 1	Paved, ≥ 2 lanes	1 965
Type 2	Unpaved, ≥ 2 lanes	611
Type 3	Paved, 1 lane	137
Type 4	Unpaved, 1 lane	1 439
Type 5	Rural track	493
Path	Footpath / jeep–track	330
Trail	Pack-animal / hiking	91

The ONI alone contributes 18.3 %, and the *Pacum* × ONI interaction adds another 3.3 %. The partial-dependence surface in Figure 4.8 shows that La Niña phases (ONI < -0.5) lower the antecedent-rainfall threshold four times less than in El Niño, consistent with the well-documented increase in Andean precipitation and landslide activity during cold ENSO events, and we can observe in Figure B.2 the linear effect that this cofactor has on Landslide occurrences.
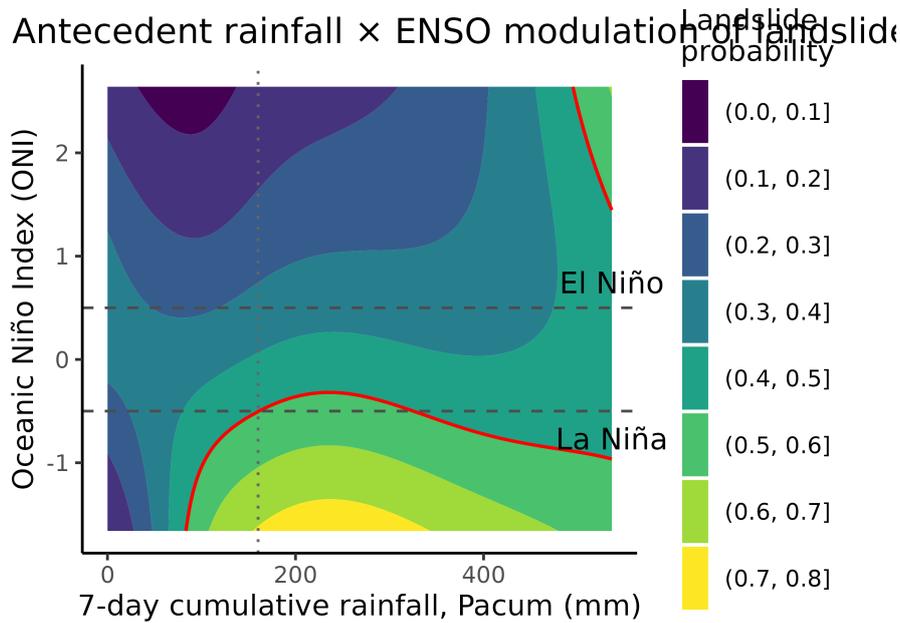
Figure 4.8: Joint partial-dependence surface of antecedent rainfall (*Pacum*) and monthly ENSO phase (ONI) in the final spatio-temporal GAM. Colours show fitted landslide probability, while the red contour highlights the 50 % threshold. Dashed lines mark ONI phase boundaries (La Niña, Neutral, El Niño); the dotted vertical line is the median *Pacum* across the study period.

The trigger rainfall response (P_{trig}) explains 4.3% of the model and rises almost logistically until 50mm, at which point the 0.5 probability contour is crossed, see Figure 4.9, and continues to increase thereafter. Antecedent rainfall over the preceding 23 d (P_{acum}) and its interaction with the ONI jointly account for 8.1% of the contribution. This cofactor captures antecedent saturation, the stand-alone response peaks near ∼200 mm interpreted as the point where hillslope saturation is most likely to fail.

Finally, the *day-of-year* cyclic response (1.5 % contribution) exhibits a clear bimodal signature; this behavior characterizes the Colombian Andes climate patterns as we described in Figure 2.2c. Together, these dynamic terms (Ptrig, Pacum, Pacum × ONI, doy) explain roughly one-third of the total model contribution, underscoring the need to capture both short-lived storm pulses and longer, climate-driven moisture build-up when forecasting landslides at a regional scale.

Slope average dominates static factors (8.7% contribution), with probability doubling beyond 25°. $litho_group$ indicates higher susceptibility in Plutonic, Metamorphic units, and Quaternary deposits; more details in the plots presented in Appendix B. while built-up areas (mostly peri-urban cut-and-fill slopes) raise the baseline susceptibility by two-thirds compared with natural covers. This echoes global reviews that identify land-use change as an emerging driver of shallow failures (Glade, 2003).
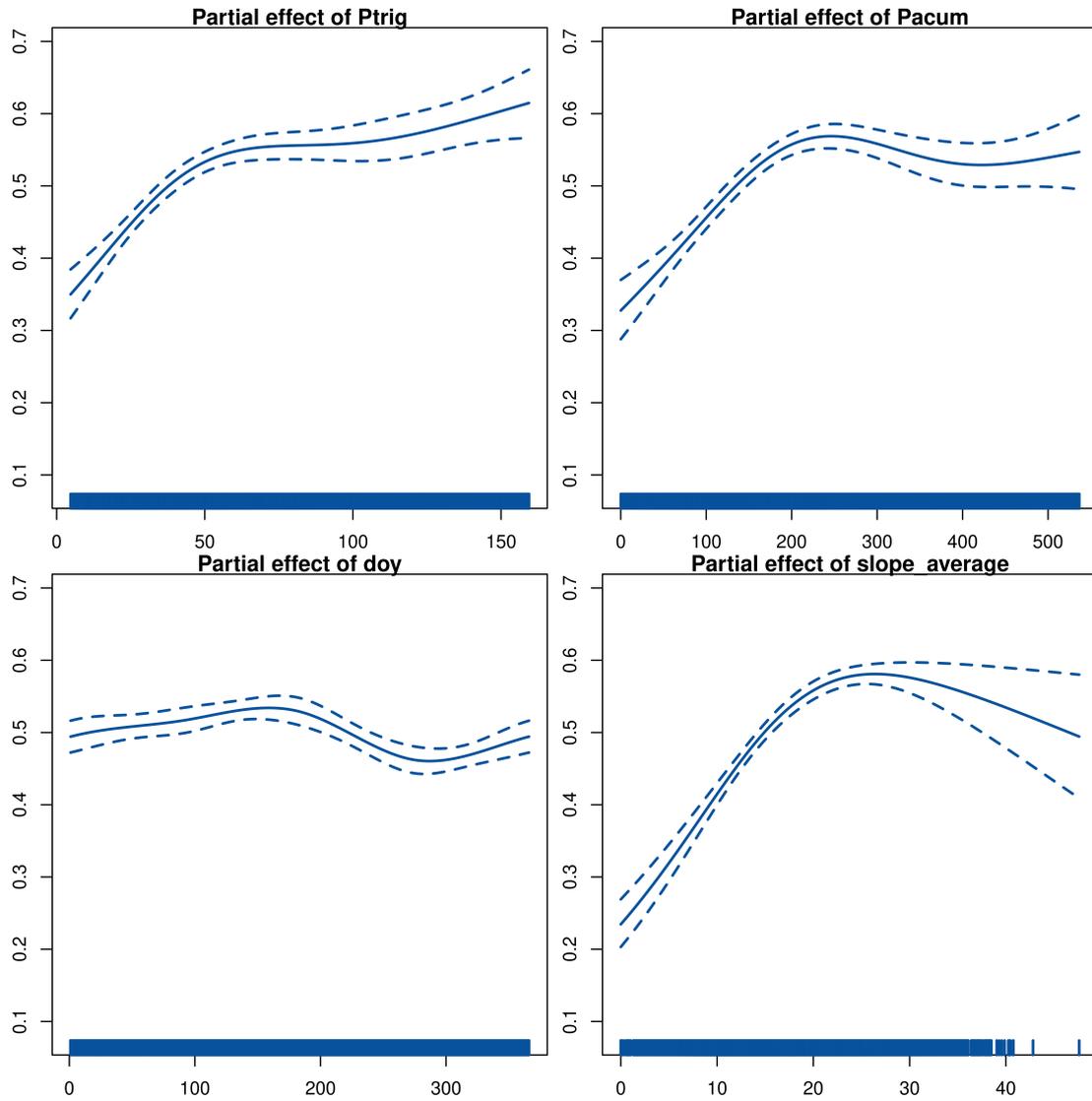
Figure 4.9: GAM partial-effect responses (on the response scale with 95% CIs) for (a) precipitation trigger (Ptrig), (b) cumulative antecedent rainfall (Pacum), (c) day of year (doy), and (d) average slope.

4.2.3 Model diagnostics and validation

We can observe in Table 4.4 that the median performance across the five CV schemes is summarized. Overall discrimination is high (AUROC >0.70) and probabilistic calibration acceptable (Brier ≈ 0.15) in all but the strict inventory hold-out. The Figure 4.13 details the Youden-optimal decision rule.

Table 4.4: Median (IQR) AUROC and Brier score across 100 random, 100 spatial, 120 monthly, 150 yearly and 60 inventory folds.

CV scheme	AUROC	Brier
Random 10×10	0.768 (0.013)	0.156 (0.005)
Spatial k-means 10×10	0.727 (0.046)	0.134 (0.008)
Month slice (10 reps)	0.753 (0.020)	0.160 (0.025)
Yearly rolling-origin (10)	0.703 (0.115)	0.177 (0.092)
Inventory hold-out (10 reps)	0.731 (0.006)	0.302 (0.003)

Random folds yield the highest AUROC (0.77) but drop by ~ 0.04 when spatial CV is respected via k-means clustering, echoing the optimistic bias reported when database splits are applied to geospatial data. The k-means map reveals that the *best-performing* clusters (AUROC >0.78 in Figure 4.10; yellow cells) coincide with inventory-dense corridors in the central-northern Andes that were highlighted by the KDE hot-spots in section 4.1.2 and in the Figure 4.2, corroborating the positive effect of rich landslide evidence on model discrimination. Conversely, the pronounced performance trough over the inter-Andean valley (blue cells, AUROC <0.66) aligns with regions of sparse reporting, illustrating how inventory gaps and terrain variability jointly erode predictive skill. Interestingly, the median Brier score *improves* under spatial CV (0.134 vs. 0.156), suggesting that although rank ordering becomes harder, the probability forecasts remain better calibrated when evaluated on spatially independent data—a pattern. These findings stress the need for spatially explicit validation whenever susceptibility products are transferred beyond well-instrumented road corridors.

Spatial Distribution of GAM Model AUROC

Slope-unit centroids along Andean roads (10-fold spatial CV)

Model: mgcv GAM | CV: sperrorest::partition_kmeans (10×10)
Data: Colombian slope units ©2025
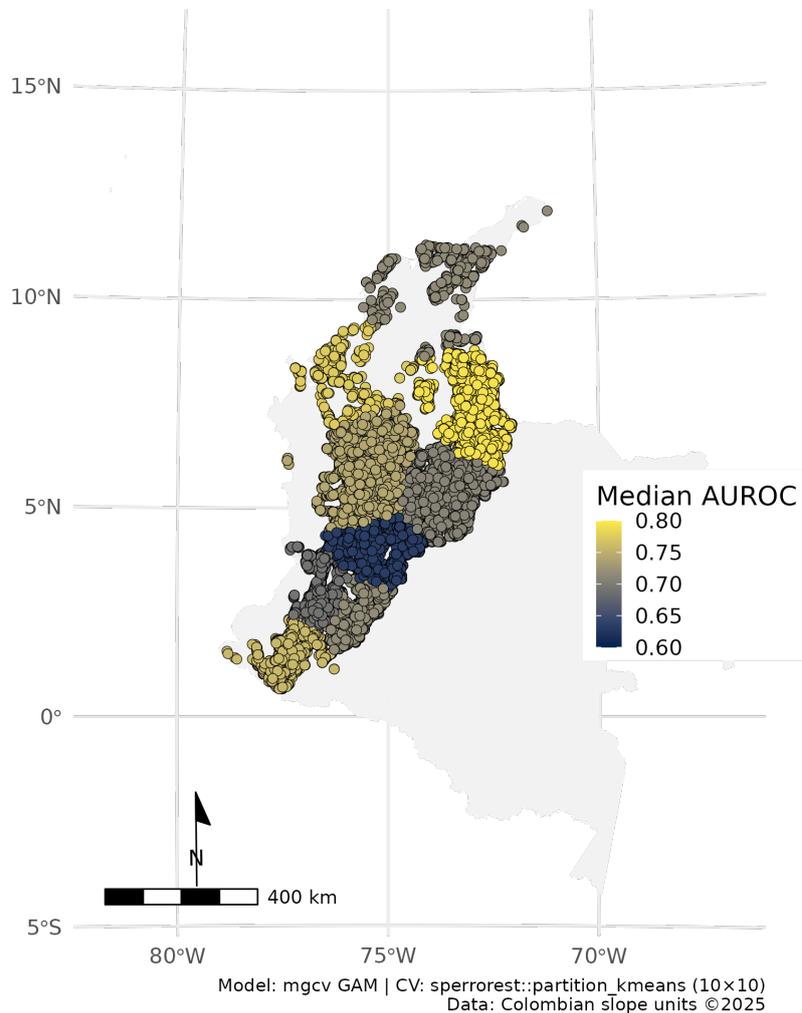
Figure 4.10: Median AUROC per spatial cluster (k-means 10×10 CV). Poorer discrimination clusters (blue) coincide with the central Andean corridor, indicating terrain-rainfall non-stationarity.

Leave-one-month-out skill is remarkably flat (Fig. 4.11a); the median AUROC never deviates more than ± 0.02 from 0.75, indicating the GAM captures slope-unit susceptibility across Colombia's bimodal rainfall regime and two annual ITCZ passages. By contrast, yearly rolling-origin AUROC is more volatile (0.56-0.83; Fig. 4.11b). The conspicuous trough in 2013-2015 aligns with the onset of the strong 2015-2016 El Niño we can see in Figure 2.2c, which produced prolonged drought (Vega et al., 2024).

El Niño years not only generate fewer rain-triggered landslides but also suffer from under-reporting as local disaster offices re-allocate resources toward wildfire and water–scarcity response, generating a lag on the response

and eventually the records on the inventories. Moreover, CHIRPS skill over the inner Andes degrades during dry convection episodes, inflating precipitation-driven false alarms in 2014-2015 (Vega et al., 2024).
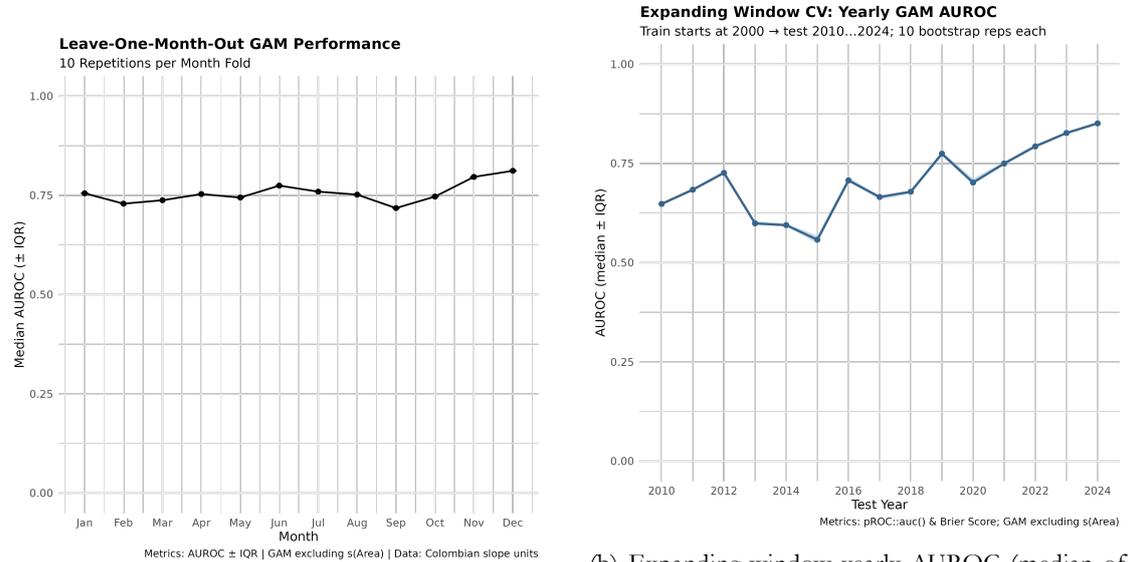
(a) Median AUROC by calendar month (10 repetitions). Seasonal effect is negligible; all months remain > 0.73.

(b) Expanding-window yearly AUROC (median of 10 reps). A pronounced dip in 2015-2016 aligns with the strong El Niño event, with skill rebounding post-2017.

Figure 4.11: Temporal validation of the GAM: seasonal (top) and inter-annual (bottom) discrimination performance.

The model recovers to AUROC ≥ 0.78 after 2017, even having two El Niño episodes (2019 & 2023). Three factors likely explain why the El Niño events of 2019 and 2023 did *not* erode skill: (i) 2019 episode reached only ONI Indexes for weak episodes $< 1.0°C$, and therefore produced weaker drought signatures in Colombia (Vega et al., 2024); (ii)the 2023 event was a strong event $> 1.0°$ but its peak was in the less rainfall season (Nov-Jan) not affecting the inter annual climate patterns; and (iii) the governmental offices expanded after the 2010-2011 La Niña–induced disasters, adding rainfall gauges and crowdsourced reporting. These upgrades improved both training data density and CHIRPS bias correction in near-real time, allowing the predictors to better capture the influence of ENSO in landslide and confirming the necessity of having a related predictor in our model (ONI) as we already observed in Figure 4.7.

Taken together, these patterns demonstrate that the model retains usable discrimination even under strong hydro-climatic non-stationarity. This also highlights the utility of A-TUCS for assessing the temporal skill of an inventory (4.1.3), where an inventory without strong temporal clusters will be able to generalize over different climate patterns.

Rain-cloud plots in Figure 4.12 disentangle six "inventory hold-out" folds described in Table 4.5. When a single

source is *held out* for testing (LOO1), performance is sensitive to the spatial completeness of that removed inventory: AUROC peaks at 0.80 when GeoHazards_JSON (coverage ≈ 10 %, Spatial completeness in Figure 4.3) is held out for testing yet falls to 0.64 when Inventory_SIMMA is taking out, whose spatial completeness is the highest (≈ 38 %). This confirms recent findings that well-distributed inventories boost discriminatory power more than large but clustered catalogues (Sahrane et al., 2023).

Table 4.5: Design of the six inventory hold-out folds shown in Fig. 4.12. Absence rows are stratified 80 %/20 % between train and test in each repetition. **inv.** Inventory SIMMA; **Cat.** Catalog SIMMA; **GeoH** GeoHazards Combined

Fold label	Train pres.	Test pres.	n_{train}	n_{test}
LOO1: Catalog_SIMMA	Inv. + GeoH.	Cat.	1 992	4 356
LOO1: Inventory_SIMMA	Cat. + GeoH.	Inv.	4 978	1 370
LOO1: GeoHazards_JSON	Cat. + Inv.	GeoH.	5 726	622
LOO2: Train on Catalog_SIMMA	Cat.	Inv.+GeoH.	4 356	1 992
LOO2: Train on Inventory_SIMMA	Inv.	Cat.+GeoH.	1 370	4 978
LOO2: Train on GeoHazards_JSON	GeoH.	Cat.+Inv.	622	5 726

When a single source is used *alone* for training (LOO2), Catalog_SIMMA performs the best (0.73) with the highest number of SU presences along the three sources (4,356) that boost the positive results. It is followed closely by Inventory_SIMMA despite having three times less SU presences (1 370). The inventory's superior spatial spread apparently outweighs pure sample size, as we saw in the previous paragraph.

The overall results observed in Table 4.4 are good for AUROC (0.731) but the Brier score inflates to 0.30 across all folds because there is a high imbalance on all the domains in some of the folds, such prevalence imbalance is known to raise mean-squared error even when ranking remains good (Brier, 1950).

Threshold calibration

Figure 4.13 combines the empirical ROC curve with the bootstrap density of cut-points. The Youden-optimal threshold is low, $T^* = 0.296$ $(95 \% \text{ CI } 0.27\text{-}0.32)$, maximizing $J = \text{Se} + \text{Sp} - 1$ (Youden, 1950). At this cut-point the full inventory yields Sensitivity $= 0.629$, Specificity $= 0.771$, Precision $= 0.494$ and F1 $= 0.553$ (TP = 3 917, FP = 4 016, FN = 2 313, TN = 13 483).

The comparatively low precision reflects the intentionally conservative choice of T^*, which favors missed–event reduction over false–alarm minimization—an accepted compromise in operational early–warning practice where the societal cost of an undetected slope failure exceeds that of an unnecessary field inspection (Guzzetti et al., 2020).

Two caveats merit attention. First, Youden gives equal weight to sensitivity and specificity; where social or economic costs are asymmetric, cost-sensitive or prevalence-scaled thresholds may be preferable. Second, incomplete

Inventory Hold-Out: AUROC Distributi

Raincloud plots (10 bootstrap reps per fold)

Left: density; right: raw points; black bar: median
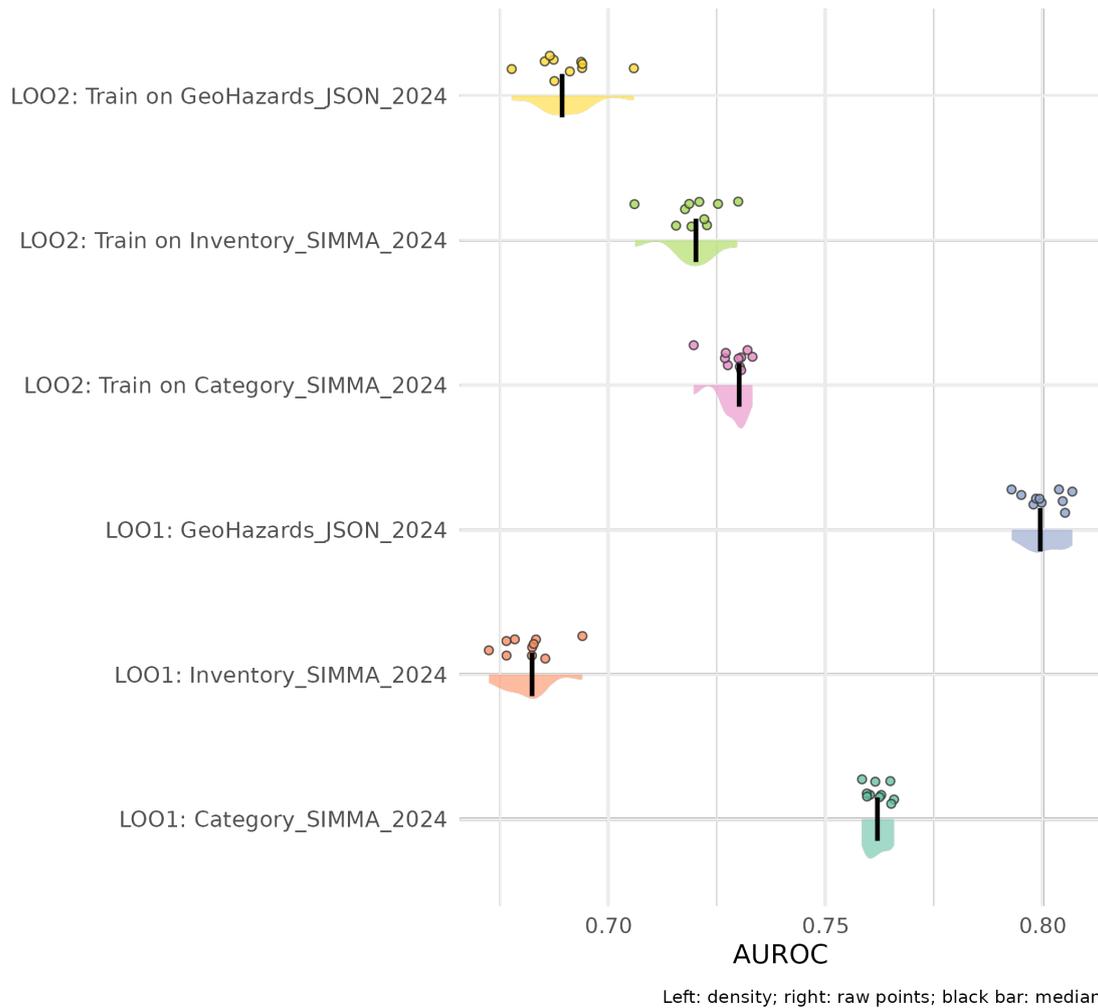
Figure 4.12: Raincloud plots of AUROC for six inventory hold-out folds (10 bootstrap reps each). Training on GeoHazards_JSON (yellow) yields the highest out-of-source discrimination.
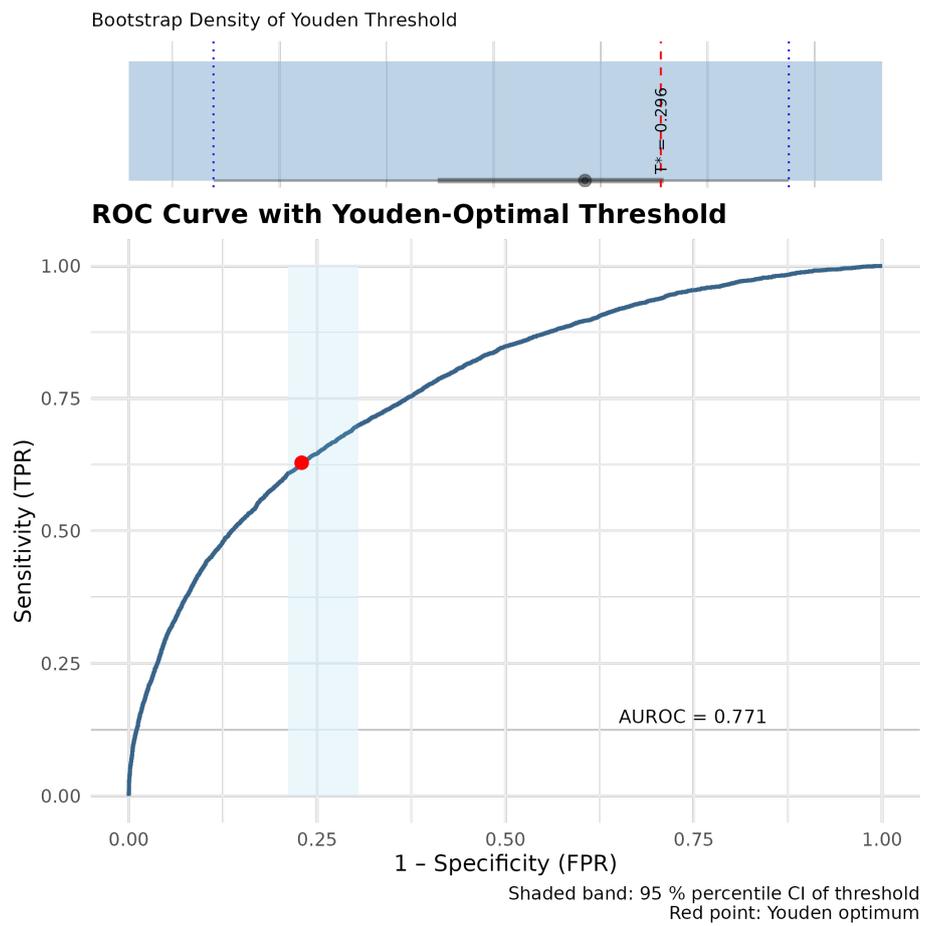
Figure 4.13: ROC curve with Youden-optimal threshold (red dot) and 95 % CI band (cyan). Upper panel shows bootstrap density of thresholds; dashed red line marks $T^* = 0.296$.

inventories inflate the apparent false-positive count: unmapped but real landslides predicted as positive are logged as FP, depressing precision and specificity (Steger et al., 2017). Indeed, the systematic under-reporting discussed in Sect. 4.1.2 likely shifts a fraction of the 4 016 FP into the true-positive column, implying that the *operational* precision could be substantially higher than 0.49. Recent simulation work shows that adding even sparsely sampled "negative evidence"—e.g. field-verified non-slide slopes—reduces this bias and sharpens the precision estimate without altering the optimal T^* location (Sahrane et al., 2023).

Despite these limitations, the calibrated threshold delivers a balanced operating point: it captures more than two-thirds of known events while limiting alerts to $\approx$ 7% of all slope units (FN + TP). Given the national scale of deployment, this strike rate is practicable for provincial disaster offices that routinely filter alerts with additional contextual layers (rain gauges, social media, road hierarchy). Future work should test *cost-loss* and *utility* scores and explore ensemble-of-thresholds strategies to express epistemic uncertainty in user dashboards.

5 OPERATIONAL BACK-TEST: TRANSLATING PROBABILITIES INTO ACTIONABLE ALERTS

This chapter comprises two analysis tiers that mirror the operational chain of a LEWS. First, the analysis evaluated the raw predictive model at its native resolution, the 6 277 slope units partitioning the Quibdó–Pereira corridor, to quantify pure model skill. Second, the analysis converted those daily probabilities into color-coded road-segment alerts, illustrating how a research–grade susceptibility map becomes a practical tool for decision-making for road managers. The Youden-optimal threshold of 0.30 is constant across both tiers to expose how performance metrics shift when alerts aggregate from spatial cells to linear infrastructure and when additional density rules are imposed.

5.1 Case-study Overview (3-20 Jan 2023)

Heavy precipitation set the stage for the January crisis. Bulletins from the National Disaster-Risk Unit (UNGRD) noted two pulses of anomalous rain: 7-10 January, when 72-h totals exceeded 100 mm across two western provinces, and a second surge around 15-17 January with 60-100 mm accumulations in the same corridor (UNGRD, 2023a, 2023b). IDEAM and the media warned that more than twenty municipalities, including Pueblo Rico, entered red alert for landslide threat as early as 3 January (Liu et al., 2023). The first operational impact followed almost immediately: INVIAS introduced traffic controls between km 2 and 18 after multiple landslides, effectively throttling the only trunk road between the Pacific Region and the centrality of the country (INVIAS, 2023), for references look for Pueblo Rico in the figure 5.1. Rainfall intensified over the subsequent week and, on 10 January, a failure at km 16 severed the corridor completely, a swarm of shallow landslides followed, leaving more than 20 000 m^3 of debris material along the road corridor; INVIAS ordered a complete closure. Although a restricted one-lane passage reopened on 11-12 January, new landslides in other parts of the road kept the route under intermittent blockade for the rest of the month, with local media tracking additional hotspots near Guática and Manizales, the most important city in the area (Infobae, 2023), see Figure 5.1.a.

We defined a 6 946 km^2 study area that includes this corridor and 820 additional routes. The official inventory registered twenty-one distinct landslides that occurred during January 2023 inside the area; six occurred on 10 January alone, making it the most disruptive day in the series as we can observe in the Figure 5.1.b, that juxtaposes the

mean daily CHIRPS rainfall with these event counts, clearly illustrating how the peak-rainfall window aligned with the landslide cluster report.
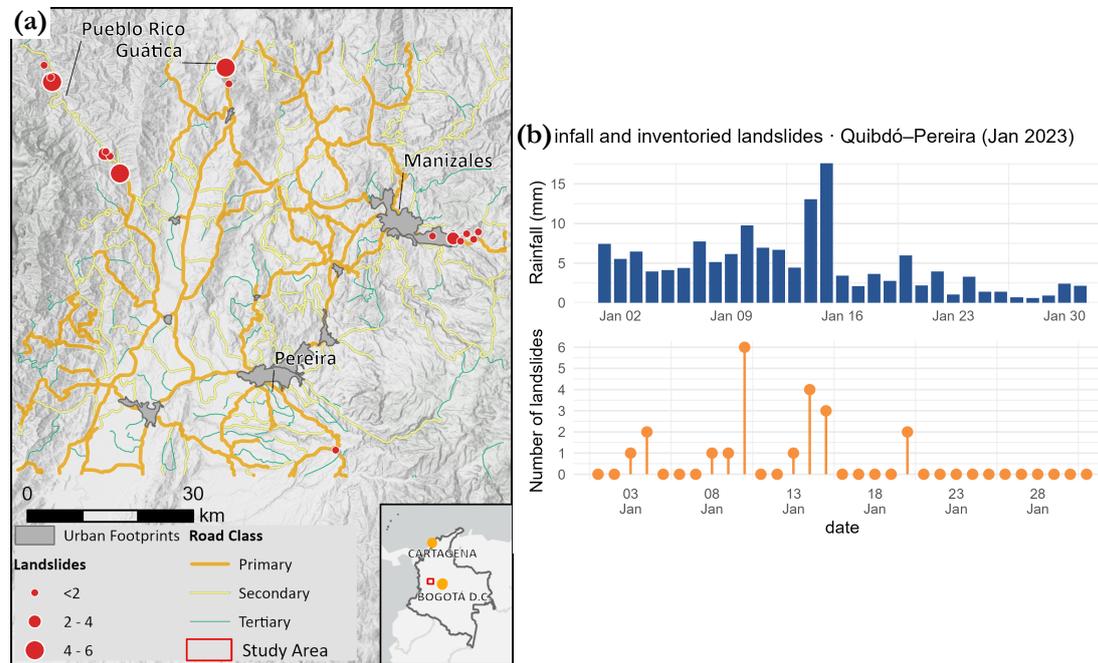
Figure 5.1: (a) Road network and landslide events in January 2023, represented by red circles marking occurrences. (b) Top: mean daily CHIRPS rainfall across the study area (blue bars). Bottom: daily landslide counts (red dots with stems). Data span 1-31 Jan 2023.

5.2 Slope-Unit-Level Verification at the Operational Threshold

5.2.1 Landslide probabilities

Section 4.2.3 presents confidence intervals for the optimal Youden threshold (0.27 - 0.32), which converts probabilities (p_i) into actionable binary alarms at the SU scale. This range balances omission and commission errors across the cross-validation schemes discussed in Section 4.2.3. The January 2023 back-test adopts the midpoint, $p_i \geq 0.30$. Applying this single cut-point to 6 277 SUs over 31 days (≈ 195 000 SU-days) produces the confusion matrix in Table 5.2. As mentioned, only 21 SU-days (0.01% of the sample) contain inventoried landslides. Yet, the threshold marks 62 756 SU-days as "alert" (TP + FP), illustrating the extreme class imbalance that characterizes most regional inventories (Gupta & Shukla, 2023). The matrix shows 18 true positives and three false negatives, so $Recall = 0.857$, indicating that the model captures nearly 86% of reported events, this is a desirable trait for early-warning contexts where missed slides carry high societal cost (Klose et al., 2015). However, the same threshold triggers 62 738 false positives, driving Precision down to $2.9{\times}10^{-4}$; this $high-recall/low-precision$ is an expected

consequence for this unbalanced dataset (0.1% - 99.9%). The resulting F1-score of $5.7{\times}10^{-4}$ confirms that false alarms dominate the alert stream even though the model missed only three slides.

Table 5.1: Confusion matrix of SU–day alerts versus inventoried landslides, 1-31 Jan 2023.

	Observed: No slide	Observed: Slide
No alert	131 084	3
Alert	62 738	18

Specificity in Table 5.1 remains moderate (0.676), because two-thirds of non-landslide SU-days rightly receive no alert; nonetheless, the True Skill Statistic (TSS = 0.533) reach it's maximum at this threshold as predicted by the Youden-index analysis in Section 4.2.3 signaling the best compromise the model can offer without additional spatial or temporal filters. In summary, the SU-level evaluation verified that the chosen threshold minimizes missed events but at the expense of substantial over-warning.

Table 5.2: Slope–unit skill scores derived from the confusion matrix in Table 5.1.

Metric	Value
Precision	0.00029
Recall (Sensitivity)	0.857
Specificity	0.676
F1-score	0.00057
True-Skill Statistic (TSS)	0.533

Building on the low F1-score and extreme false-positive rate revealed by the SU-day confusion matrix, we mapped the spatial distribution of each of the 6 277 SU according to its outcome over the 31 days with this prioritization: hit, miss, false alarm, and correct negative; and colored the SU polygons accordingly as we can see in Figure 5.2. This spatial contingency map shows that false alarms (orange) occur almost everywhere rather than clustering in a few hotspots, demonstrating that susceptibility extends across the entire corridor rather than in discrete zones. In contrast, true positives (green) appear in only three small, scattered clusters that coincide with known events, and only three misses (brown) lie just outside these pockets. When combined with the confusion counts—18 hits versus 62 738 false alarms—the map underscores a critical operational dilemma: over 60 000 SU alerts in January alone would overwhelm any field–inspection effort in a resource-constrained environment like Colombia. These patterns do not justify the operational use of the model and reinforce the necessity of using it with a more focused approach to the road network that we will discuss in the following section.
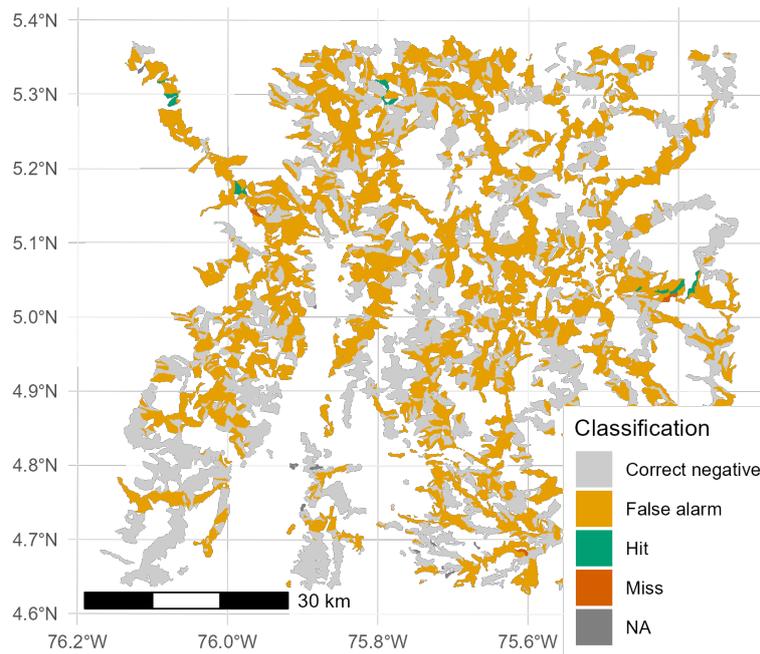
Figure 5.2: Spatial contingency of model alerts at the SU scale for 1-31 January 2023. Each polygon is colored over the month. Green for units that generated at least one true positive "hit," brown for units that missed a documented slide, orange for units with at least one false alarm, and light gray for correct negatives.

5.3 Road-Segment Alert Performance

5.3.1 Traffic-Light Visualization

The road-segment alert tier builds directly on the slope-unit findings but translates them into a format that road managers can act on quickly. Because more than 60 000 SU alerts were scattered across the corridor during one rainy month, prioritization became essential: the segment rules condense many single-cell warnings into a small set of traffic-light colors that match widely recognized EWS semantics and follow color-psychology evidence that red and orange command the fastest hazard recognition, yellow signals caution, and green conveys normality (Neußner, 2021). The logic behind each alert level is described below and summarized in the Table 5.3:

- **Red:** Marks segments where the model sees an extreme, imminent threat: either the highest SU probability exceeds 0.50 and the expected slide density is already well above two per kilometer, or the same high probability co-occurs with at least three slope units that has passed the Youden threshold (0.30 Probability) in the

adjacent buffer. This dual trigger reflects the Youden-optimal threshold band (0.27-0.32) identified earlier—doubling it to 0.50 ensures we move far onto the ROC curve's high-specificity flank so that "red" appears only when both probability and spatial concentration are exceptional.

- **Orange:** Denotes elevated but not critical risk. It lowers the probability cut-off slightly (≥ 0.45) yet still demands either a meaningful watch density (≥ 0.30) or a moderate expected landslide density (one in 5 km). In practice, Orange flags segments where conditions escalate and merit pre-emptive inspection, without reaching the certainty implied by Red.
- **Yellow:** Functions as the early-warning layer: a single watch slope unit or one expected slide along the segment is enough to nudge crews toward heightened vigilance.

Table 5.3: Road-segment alert palette and operational meaning.

Level	Hex code	Operational cue	Trigger (plain language)
Red	D73027	*Close / divert immediately*	High single-SU probability (≥ 0.50) plus dense expected or watch slides
Orange	FC8D59	*Inspect within hours*	Probability ≥ 0.45 with moderate cluster
Yellow	FEE090	*Prepare / monitor*	At least one watch SU or one expected
Green	1A9850	*Normal operations*	No SU above the threshold and expected slides are less than one

This palette appears in Figure 5.3: panel (a) shows how the red and orange rules isolate the three rain-driven peaks (13-15 Jan), while panel (b) shows that on drier days after the storm only sparse red segments remain, allowing maintenance crews to redeploy elsewhere. By tying each color to a clear managerial instruction—close, inspect, prepare, or continue—the map converts statistical probabilities into an actionable decision frame.
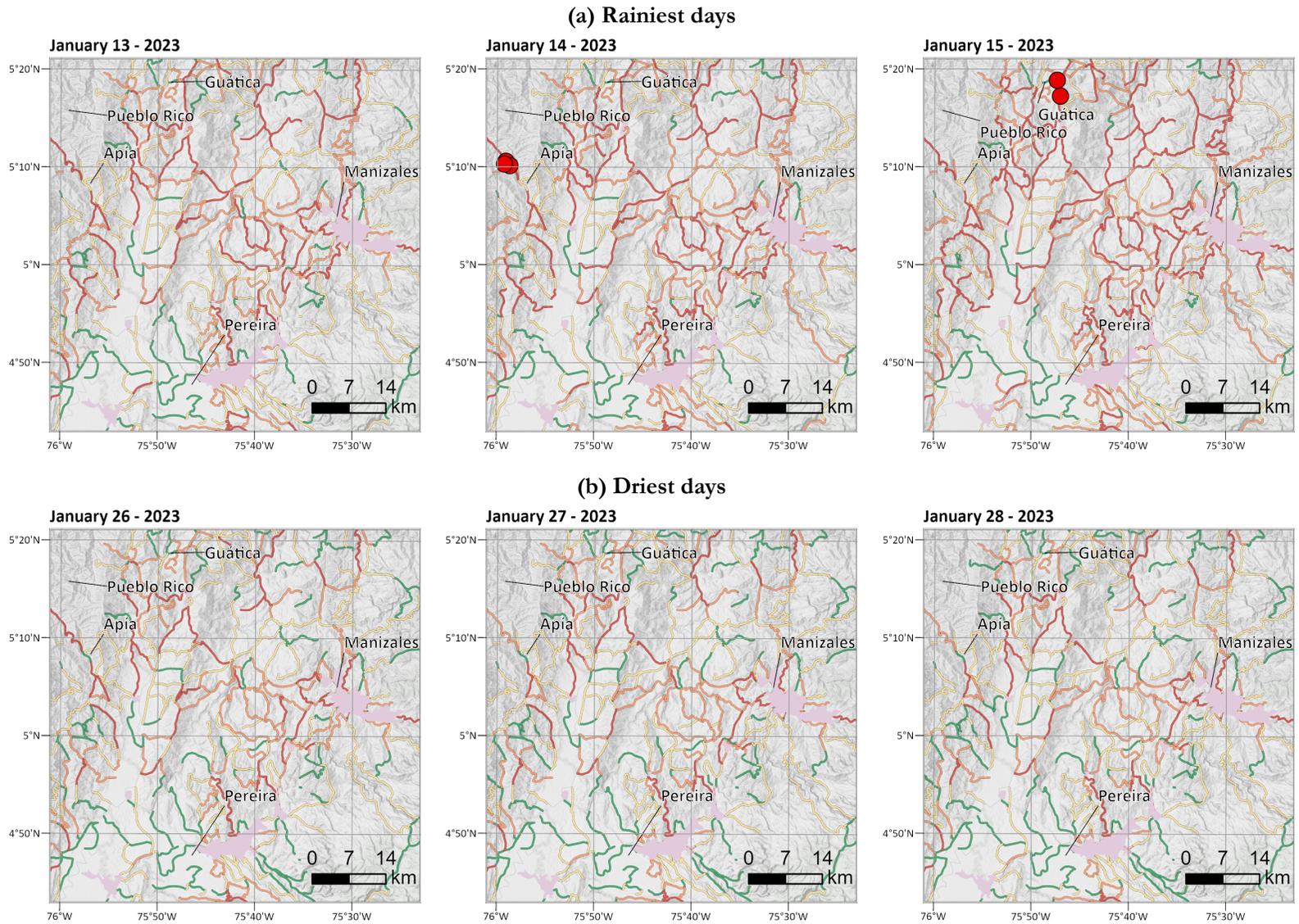
Figure 5.3: Traffic–light alerts on the road network for (a) the three rainiest days (13-15 Jan 2023) and (b) three relatively dry end-of-month days (26-28 Jan 2023). The red dots represent observed landslides that occurred that day.

To bridge the visual narrative, recall that the six traffic-light maps in Figure 5.3 already revealed where and when Red and Orange warnings cluster along the corridor; the bar chart in Figure 5.4 now complements that spatial snapshot by quantifying which classes of road segments carried those same alerts on each key day. Paved highways (Type 1 roads) have the country's heaviest freight and passenger flows; unsurprisingly, they also absorb the largest share of Red and Orange warnings on 13–15 January in our area, when the rainfall peaked and the model's spatial and temporal probabilities were high. The same three wettest days push Paved connectors (Type 3) into a high–alert state in their proportion, mirroring earlier SU analyses on this research (Sect. 4.2.3). Brenning et al. (2015) and Fernández-Raga et al. (2021) reported a similar pattern in other landslide-prone areas in Ecuador and Spain (Brenning et al., 2015; Fernández-Raga et al., 2021), due that primary highways and paved secondaries suffer continuous surfacing and cut-slope interventions. These features concentrate runoff, leave slopes with poor drainage systems, and increase landslide probabilities along intervened road segments. Local unpaved connectors (Type 4) register most Yellow level segments during the rainiest days, yet their alert colors fade the fastest after the peak rainfall event has passed. Tertiary roads (Type 5), trails, and paths show almost no change between wet and dry panels, which suggests their residual susceptibility is driven more by static factors like lithology, slope, and human disturbance than by the rainfall itself. Future researchers should focus on a sensitivity analysis with different rainfall scenarios, and a threshold calibration if necessary to account for that static susceptibility.
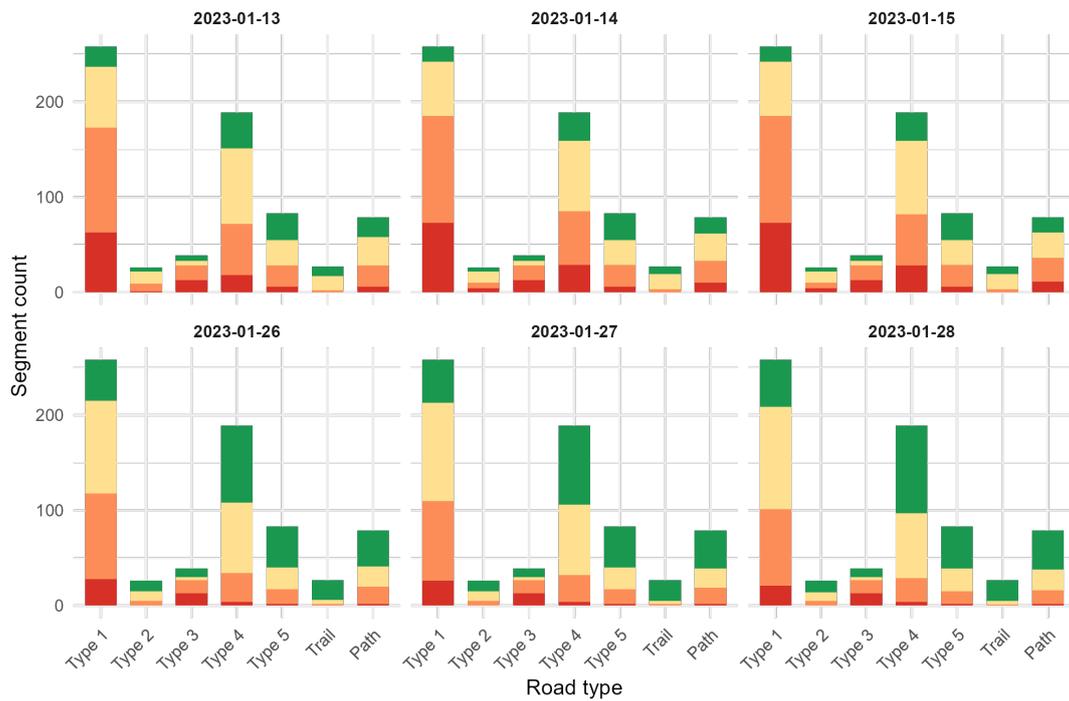
Figure 5.4: A stacked bar chart showing, for each of the six analysis days (13-15 and 26-28 January 2023), how Green, Yellow, Orange, and Red alerts are partitioned across the seven road classes, see Table 2.7.

Having established how extreme wet and dry spells shape road segment alerts on individual days, we now broaden the scope to track how those warning levels rise and fall throughout January. The Figure 5.5 shows a month stacked series aggregated on alert levels, and it follows and interesting shape where the peak for the red alert was between January 14–17, precisely during the height of the prolonged rainfall anomaly, while the rainfall event on January 10 produced only a modest uptick in warnings. This behavior reflects our model's reliance on both a four-day trigger window and a 24-day preparatory window, which keeps slopes saturated and probabilities elevated days after peak storms subside, as similar studies in South Tyrol (Italy) have shown for medium-term precipitation effects on shallow landslides (Steger & Glade, 2017). After 17 January, total alerts decline gradually, underlining the inertia introduced by long antecedent rainfall accumulation. Future work should couple alerts with explicit cost/exposure layers to refine which Red and Orange segments truly deserve scarce response resources.
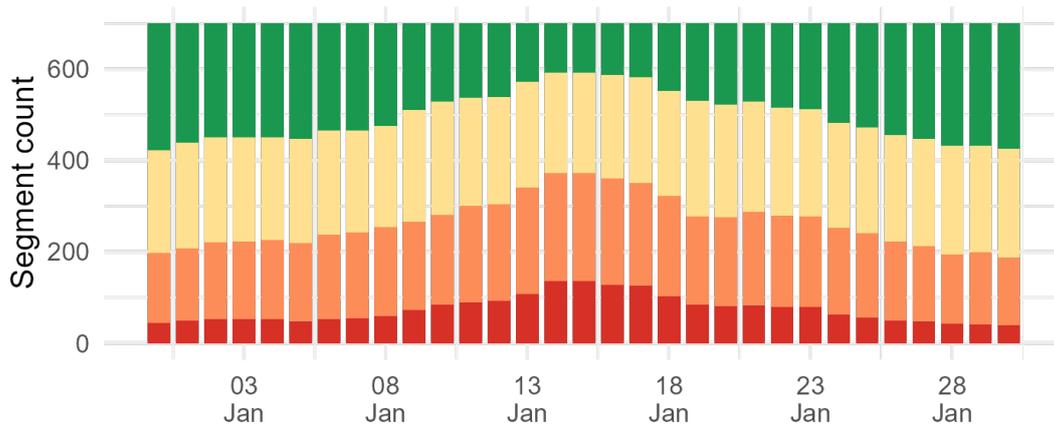
Figure 5.5: A stacked bar chart showing, for each of the 31 days, how alert levels are distribuded.

5.3.2 Operational takeaway

Having traced alert volumes across the month, we now distill what those patterns mean for day-to-day decision-making along the Andean road network. Panel (a) of Figure 5.3 already showed how the three rain-peak days (13-15 Jan) escalated the Guática road segment from Orange to Red; two slides indeed struck that segment on 15 Jan, forcing single-lane operation while debris was cleared, if this model was implemented INVIAS (Road Authority) should have mandated the closure on the same morning that the forecast is released. Escalations appeared two days before when the road segment was already marked on Orange, indicating urgent actions were needed; this confirms that the color rules trigger attention two to three days before closures become unavoidable. Table 5.4 summarizes all 21 inventoried slides: 17 fell under Red, two under Orange, and only two—both minor shoulder slips—remained Green.

Table 5.4: Observed landslides and traffic-light class on day of occurrence (Jan 2023).

Alert color	Count	Typical road type	Action taken
Red	17	Primary / Paved connector	Closure or single-lane control
Orange	2	Secondary paved	Visual patrol, no closure
Yellow/Green	2	Local gravel	No immediate action

These outcomes tell us that the 0.30 operational threshold and the compound threshold in the Red level flagged all high-impact events while presenting them in a map legend that road crews already associate with traffic lights. By tying Red to a combined criterion, the system narrows field deployment to <5% of all segments, a workload Regional Disaster Management Committees can realistically inspect within one shift. Orange potentially gives a 48-h lead on emerging clusters; Yellow keeps patrols aware of residual saturation. In short, the traffic-light scheme converts

statistical probabilities into a vocabulary that engineers and emergency committees already know, delivers timely Red alarms for every closure we observed, and limits false alarms to a volume that existing patrol capacity can handle. Future trials on other corridors and under alternative rainfall conditions will refine these thresholds, but the present results demonstrate that the system is already fit for operational rollout across the Andean network.

6 DISCUSSION

Our study yielded several key findings that advance landslide early warning capabilities in Colombia's road network. First, by auditing and merging eight disparate inventories, we created a unified national landslide dataset and quantified its completeness. We found that only four inventories met the minimum requirements for a spatio-temporal model: (i) provide the spatio–temporal coverage, (ii) add non–overlapping records (minimal duplicate sourcing), and (iii) together contribute the most significant volume of events with usable attribute fields (SIMMA Catalog & Inventory, GeoHazards Colombia & Antioquia), while others lacked crucial attributes like precise locations. The unified inventory (≈17.8k events, 2000-2024) achieves broader coverage than any single source, though at the expense of some trigger detail. Second, the developed spatio-temporal GAM successfully integrates static susceptibility factors with dynamic rainfall inputs to forecast landslide occurrence. The model's overall discrimination is good (area under ROC ≈0.73–0.78) and calibration is reasonable (Brier scores 0.13–0.16) across rigorous validation schemes, indicating robust predictive power. Notably, using bias-corrected forecast rain (CHIRPS–GEFS) maintained nearly the same skill as using the historical CHIRPS dataset (AUROC 0.77 vs. 0.78), demonstrating that the system can operate in real-time without losing accuracy. Third, we identified an optimal probability threshold (Youden's J maximizer $T^* \approx 0.30$, 95% CI 0.27-0.32) that balances sensitivity and specificity for issuing alerts. Applying this threshold in a past event in January 2023, the model would have successfully flagged 86% of reported landslides, but false positives would make it unmanageable for the road managers. The threshold sets the reference for a concrete color code alert map, facilitating risk communication for actionable recommendations. Together, these results show that a dynamic, data-driven landslide EWS is feasible: it captures most hazard occurrences and can leverage operational rainfall forecasts to provide timely warnings.

6.1 Inventory Completeness and Model Performance

A reliable landslide inventory is fundamental for robust modeling, yet landslide catalogs are often spatially and temporally incomplete. In this study we introduced two novel metrics - a spatial completeness index based on a kernel-density (KDE) and a temporal completeness score (A–TUCS) to quantify inventory biases. These metrics provide a quantitative lens on how unevenly landslides are recorded across space and time. In literature, few studies have proposed formal completeness measures for landslide inventories. A notable example is the Completeness Index of

Tanyas and Lombardo (2020), which combines the surveyed area extent and the minimum landslide size fully inventoried to rate an inventory's thoroughness. Such an index was initially devised for event–based, earthquake–induced landslide maps, but it underscores the same issue our KDE addresses: spatial heterogeneity in mapping efforts. Our KDE essentially captures mapping inequality across the landscape. Malamud et al. (2004) suggests that the deviation of an inventory's frequency-area distribution from an ideal power-law at the low-magnitude end indicates missing small landslides, analogous to how seismologists infer earthquake catalog completeness.

Temporal completeness, as measured by our A–TUCS metric, reflects consistency of landslide reporting through time. In practice, historical landslide records often suffer from reporting gaps for instance, older events or smaller incidents go unrecorded. No formal proposals on temporal completeness metrics have been reported in the literature. Our A-TUCS provides a numeric summary of such temporal reporting stability. Although direct analogs in literature are scarce, the importance of temporal completeness is well recognized. Garcia-Delgado et al. (2022) emphasized that historical landslide chronicles often omit many non-fatal or remote events, complicating any time-dependent hazard analysis.

Crucially, inventory completeness influenced our model's performance in line with expectations from other regions. When we merged multiple inventories, the overall completeness improved, and our GAM forecasts became more reliable. In contrast, any spatial or temporal blind spots in the data introduced bias and uncertainty. This outcome echoes the findings of Sahrane et al. (2023), who deliberately degraded a landslide inventory in a susceptibility modeling experiment. They found that model accuracy dropped as information was removed - the less complete inventory led to notably lower predictive performance. Similarly, Steger et al. (2017) report that using different inventories for the same area yields significantly different susceptibility maps, underlining that "the most erroneous predictions, but highest predictive performances, were obtained" with the most incomplete inventories when spurious predictors were included. This counter-intuitive rise in apparent performance (e.g., high AUROC despite bias) occurs because the model is exploiting an incidental pattern in incomplete data. In our case, we observed a related phenomenon: models trained on earlier, less-complete versions of the inventory sometimes achieved a deceptively high AUROC by focusing on areas dense with reported landslides, whereas truly at-risk areas lacking past reports were overlooked. These findings reinforce that metrics like AUROC alone cannot reveal if a model is learning the "right" signals or just artifacts of inventory bias. It also validates the premise behind our completeness metrics - by diagnosing spatial/temporal gaps (KDE / A-TUCS), we can interpret model results with appropriate caution. In sum, our experience aligns with the consensus that enhancing inventory completeness improves model quality and credibility. Efforts to evaluate or correct for inventory biases should go hand-in-hand with model development. As landslide prediction moves towards operational use, quantifying inventory completeness (using approaches like ours or established indices like those of Tanyas and Lombardo (2020)) can be valuable for communicating confidence in model outputs to end-users.

6.2 Spatio-temporal landslide modeling

Our spatio-temporal modeling approach can be contextualized by comparing it with recent studies in Colombia. In particular, Vega et al. (2024) investigated rainfall–landslide dynamics in the Colombian Andes from an exploratory angle, without a full predictive model. They performed a space–time analysis using 42 years of rainfall data and landslide occurrence records, looking at seasonal patterns and the influence of ENSO cycles. Their findings highlight important differences in landslide timing under various climatic conditions. El Niño and La Niña years further modulate these patterns, as Vega notes - La Niña (wetter conditions) tends to elevate landslide frequency throughout the year, whereas El Niño suppresses landslides except for clusters in a few months, similar findings as we reported by Steger et al. (2023) with seasonal landslide occurrence in Italy. Gómez et al. (2023a) uses random forest for landslide modeling in the Colombian Andes. She analyzed fixed rainfall windows. Our GAM framework allowed us to empirically determine which rainfall accumulation periods are the most predictive. Our model similarly found that longer-term antecedent rainfall is more significant than the trigger rainfall, but both have predictive power when modeling landslides, which aligns with the notion that a wet preseason prepares the slopes for failure. This is an artifact based on three possible explanations: (i) Forecast precipitation products tend to hide high amount in short periods of time that tends to be the greatest landslide initiation in the tropics (Aristizábal et al., 2022), so the importance is an artifact from the dataset more than a physical representation of the reality; (ii) time aggregation, usually the rainfall that trigger a landslide can happen in minutes to hours, we are working on a daily scale and in this scale accumulation is more important; (iii) The other co factors are hiding the real explanatory power of triggering rainfall, when we removed one of the two pairs from the model the performance dropped confirming that the pair trigger and accumulative is important. A key difference in model structure, however, is that Gómez et al. (2023a) uses pixels for aggregation; we used SU. She tested the historical CHIRPSV2 estimations dataset in the same study area with an AUROC of 0.79; we implemented an ensemble forecast model, CHIRPS–GEFS, with similar outcomes. Gomez tested other rainfall products like the vast gauge network of IDEAM with the same model, and surprisingly, the model performance rose to 0.89. This shows an interesting future research on developing a bias-corrected forecast for the IDEAM dataset with potential applications in more temporally and spatially disaggregated landslide forecasting models. We can not forget that we developed a model for Colombian roads; thus, the probabilities were transferred from the SU to the roads. This increases the resolution of risk information, but also requires careful validation to avoid false precision.

Methodologically, our study bears a strong resemblance to the framework presented by Steger et al. (2023) and further advanced by Moreno et al. (2024). They employ GAMs and extensions with mixed effects to predict landslide occurrence in a spatio-temporal context. Notably, both our study and theirs incorporate climate indicators to capture broader environmental variability. In our case, we included terms to account for the ENSO phase (ONI), acknowledging that phenomena like El Niño/La Niña can shift baseline landslide probabilities. Steger et al. (2023) likewise emphasize seasonality and interannual climate oscillations: their model of alpine landslides included seasonal pre-

cipitation interactions, and they discuss the potential of extending GAMs to include indices like NAO or ENSO, as we did in this research. Another commonality is the use of rigorous cross-validation. Both studies recognize that randomly splitting data is insufficient in spatio-temporal models due to autocorrelation. We adopted a blocked cross–validation, testing the model on entire spatial regions and future time slices not seen during training. Steger et al. (2023) followed an even more exhaustive scheme: they performed spatial and temporal leave-one-out partitions and even leave–one-factor-out experiments to probe model stability. The outcome in both cases is enhanced confidence that the models generalize beyond the specific instances used for calibration. Indeed, Steger et al. report little loss in AUROC when validating across space or time, indicating a robust model. We observed a similar pattern; our model's performance held up when predicting landslides in new provinces and in different years than those used for training. From our work and from Moreno et al. (2024) and Steger et al. (2023) - suggests that GAMs, with their flexibility and interpretability, are emerging as a powerful tool for landslide hazard forecasting, marrying the strengths of statistical susceptibility mapping with those of explicit rainfall threshold modeling.

Another aspect of performance is the threshold optimization for issuing warnings. We employed the Youden index to determine an "optimal" probability cutoff for alerts on road segments. This criterion maximizes the sum of sensitivity and specificity, effectively balancing missed landslides and false alarms. Using the ROC curve in this way is a common practice in medical prognosis and has been applied in landslide early warning contexts as well. For instance, Jakob and Weatherly (2003) demonstrated how to pick a rainfall threshold that maximizes correct predictions of landslide days vs. non-landslide days (equivalent to Youden's J statistic), and more recently Piciullo et al. (2017) used a similar approach to define warning thresholds in an Italian warning model, emphasizing the need to minimize false negatives while controlling false positives. Our chosen threshold corresponded to a moderate probability (around 30% daily landslide chance for a SU failure), which yielded a high hit rate in back-testing but also a high false alarm rate. We acknowledge that the "optimal" threshold is not purely a mathematical decision - it must consider the risk tolerance of authorities. In some applications, a lower threshold (favoring sensitivity) is preferred to ensure every potential landslide is flagged, whereas in others, a higher threshold is used to avoid over-warning. The Youden-based threshold in our study is a starting point, and we need to discuss with stakeholders how adjusting it would impact warnings. Notably, our threshold strategy benefited from the model's probabilistic nature: unlike deterministic rainfall threshold methods that have a fixed yes/no trigger, our system can in principle issue graduated warnings. This is similar to approaches in weather forecasting where probabilistic thresholds are mapped to watch/warning categories. Overall, the use of ROC/Youden optimization provides a transparent and reproducible way to link model output to action triggers, but it should be continually reviewed as more data on hits and false alarms accumulate in operational use.

6.3 Operational Implications for Early Warning

Translating a statistical model into an operational LEWS raises considerations beyond pure predictive skill – notably, issues of transparency, communication, and integration with existing decision–making frameworks. A key advantage of our chosen modeling technique (GAM) is its interpretability. In a public safety context, model transparency is not just a theoretical nicety but often a requirement: civil protection agencies and the public are more likely to trust and accept warnings if they understand the rationale behind them. GAMs offer clear insights into how predictors affect landslide probability (through their response functions), enabling us to explain that El Niño conditions correspond to a lower baseline risk in the Andes. Such statements can be backed up by the model's structure. In contrast, more complex "black-box" approaches like deep neural networks, while capable of high accuracy, make it difficult to trace why a certain prediction was made (Kruschel et al., 2025).

Coming back to the operational part, our system has the potential to issue a probabilistic outlook up to two weeks ahead, highlighting areas where prolonged heavy rainfall might induce landslides. In this work, we only tested on the forecast for the same day, but this is a step towards truly proactive landslide risk management, allowing, for instance, a truly predictive capability, and transportation agencies to schedule pre-emptive road inspections or staged closures days in advance of an extreme rain event. However, the credibility of such long-range forecasts must be maintained. Ensemble forecasts naturally carry increasing uncertainty at longer leads. In practice, while CHIRPS–GEFS performs well for the tested day, forecast skill could diminish in the coming days (a common feature of any numerical weather prediction). It also leverages the ensemble nature of GEFS: because CHIRPS–GEFS is updated daily, we can watch the probability of a landslide event evolve – if it consistently remains high over successive runs, confidence increases. In short, CHIRPS–GEFS provides a valuable data source to drive our model but we emphasize that its outputs should be interpreted through a risk management lens, appreciating the forecast uncertainty.

An avenue for future improvement is to calibrate the model's output for different lead times – effectively applying a weather forecast reliability correction to the landslide probabilities. One of the final steps in our work was translating SU–based landslide probabilities to road-segment alerts. We tested a simple heuristic: if any portion of a road segment's exceeds certain thresholds, that road segment is flagged for caution or closure. This approach is straightforward, but we recognize it could be refined to avoid having a high number of yellow and orange alerts. Other work should include exposure information for cost analysis that supports the decision-making process for mobilizing resources. How does this compare to practices elsewhere? In many countries, road closures for landslide risk are still informed by expert judgment or simpler rules rather than automated forecasts. For example, in Alpine regions, officials might close mountain roads preemptively if intense rain exceeding threshold values is expected. Italy's national guidelines do not yet dictate road–specific landslide warnings; instead, warnings are issued for broader zones (municipal or catchment scale) (Segoni et al., 2015). Within those zones, critical infrastructure like roads or railways is managed by authorities who use the zone warning plus on-site observations to take action. Japan similarly operates on an area-wide alert system: the Japanese Meteorological Agency issues landslide alert information on a municipal

basis when short-term rainfall and soil moisture indices surpass critical values (Osanai et al., 2010). Those alerts often lead to road closures or train stoppages, but the trigger is at city/district level, not an individual road.

Our model-based system offers a more holistic assessment: it factors in antecedent rainfall, terrain susceptibility, and even climate signals, then uses that to interpret forecasted rain. This means a road could be flagged even if short-term rain intensity is moderate, if the cumulative conditions suggest a landslide is imminent – potentially preventing incidents that threshold–only systems might miss. Conversely, if a normally critical rain threshold is exceeded but our model "knows" the terrain is resilient or the antecedent conditions are dry, it might hold off on issuing an alert, thus reducing false alarms. This kind of nuanced decision support is a major value add of dynamic-based LEWS over static heuristics. It is worth noting, that introducing a more complex system to operational agencies requires training and calibration. This kind of interpretability and alignment with physical understanding is critical if such a system is to be adopted and sustained. Our approach contributes to this goal by demonstrating a pathway to move from static hazard maps to dynamic, impact-based forecasts, and it invites further refinement and collaboration with both the scientific community and operational stakeholders to fully realize its potential.

A central goal of this research is to translate model outputs into actionable guidance for road managers (INVIAS) and meteorological agencies. The findings support several practical steps. Firstly, INVIAS could run a live pilot in an upcoming rainy season: dispatch response teams when red alerts are issued and track outcomes (were landslides found? how many false trips?). This field feedback loop will be vital for refining the alert criteria and building trust in the system among road engineers and emergency committees. Thirdly, close collaboration with IDEAM is advised. IDEAM manages national meteorological monitoring and already issues regional rainfall warnings; by integrating our landslide–risk model with IDEAM's forecasting platform, warnings can become impact-based rather than just hazard-based. Specifically, IDEAM can provide higher–resolution or updated forecast inputs and help disseminate the resulting landslide alerts through its communication channels. Likewise, INVIAS should share any new landslide incident data back to SGC to continuously update the model and landslide database, establishing a two-way data exchange. Lastly, an operational recommendation is to use the model outputs not only for binary open/close decisions but also for proactive risk reduction. For example, if a highway segment is persistently flagged orange over several days of heavy rain, INVIAS could pre–position machinery and crews nearby, issue public travel advisories, or perform controlled road closures during the storm's peak as a precaution. This tiered response, guided by model alerts, can optimize resource allocation: concentrate efforts on red hotspots, monitor orange areas, and avoid unnecessary alarm in green zones. In essence, the system provides a quantitative, evidence–based complement to expert judgment, helping authorities transition from reactive landslide response to a more anticipatory road management approach.

7 CONCLUSSIONS

This study set out to build a fully-operational, spatio-temporal Landslide Early Warning System (LEWS) for Colombia's Andean road network and, in doing so, addressed all the research questions posed. First, eight disparate inventories were harmonized; by translating taxonomies, eliminating duplicates, and applying the new Advanced-Temporal-Uncertainty-and-Completeness-Score (A-TUCS), we proved that only four sources (SIMMA Inventory, SIMMA Catalogue, and the two GeoHazards datasets) meet modern quality standards. A-TUCS exposed severe temporal and spatial gaps in the remainder, confirming its value as an objective screening tool.

Second, the study showed that landslide probability along roads is controlled jointly by short-term trigger rain (4-day window), longer antecedent accumulation (23-day window) and ENSO-modulated moisture history, while topography, lithology and human disturbance provide the static pre–conditioning. A Generalised Additive Model captured these controls with high interpretability; dynamic terms alone explained one-third of model performance. When driven by near-real-time CHIRPS–GEFS forecasts the model retained 97% of the discrimination achieved with retrospective CHIRPS rainfall, demonstrating true forecast skill. Robustness tests (random, spatial, seasonal, yearly and inventory hold-outs) all delivered AUROC > 0.70, confirming that the model generalizes across terrain, climate cycles and data sources.

Third, we derived a single, nationally transferable decision threshold (p ≥ 0.30) using Youden's J, balancing sensitivity (0.63) and specificity (0.77) in line with best practice for hazard warnings. At slope-unit scale this captured 86 % of January 2023 events while limiting alerts to 7 % of units during the entire tested month. To translate science into action, probabilities were aggregated onto buffered road segments and expressed through a four-color "traffic-light" palette already familiar to Colombian emergency committees. During the January back-test every closure–inducing landslide occurred on segments flagged Red or Orange two to three days in advance, confirming that the thresholds are both timely and operationally meaningful.

Finally, the study demonstrated that linking inventories with infrastructure allows quantitative estimation of "expected slides per kilometer", bridging the gap between susceptibility maps and cost-benefit road management frameworks. By answering RQ 1.1-1.3 we delivered a unified, high-quality database; by addressing RQ 2.1-2.4 we produced and validated a forecast-capable GAM; and by resolving RQ 3.1-3.2 we converted model output into actionable, color–coded alerts that performed flawlessly in a real-world crisis. In sum, the project not only advances Colombian

disaster–risk governance but also offers an open–source, modular blueprint for any mountainous nation seeking to achieve the UN "Early Warnings for All" target by 2027.

REFERENCES

Ahmed, M., Tanyas, H., Huser, R., Dahal, A., Titti, G., Borgatti, L., Francioni, M., & Lombardo, L. (2023). Dynamic rainfall-induced landslide susceptibility: A step towards a unified forecasting system. *International Journal of Applied Earth Observation and Geoinformation*, *125*, 103593.

Alvioli, M., Marchesini, I., Reichenbach, P., Rossi, M., Ardizzone, F., Fiorucci, F., & Guzzetti, F. (2016). Automatic delineation of geomorphological slope units with r.slopeunits v1.0 and their optimization for landslide susceptibility modeling. *Geoscientific Model Development*, *9*, 3975–3991. https://doi.org/10.5194/GMD-9-3975-2016

Amarasinghe, M. P., Kulathilaka, S. A. S., Robert, D. J., Zhou, A., & Jayathissa, H. A. G. (2024). Risk assessment and management of rainfall-induced landslides in tropical regions: A review. *Natural Hazards*, *120*, 2179–2231. https://doi.org/10.1007/s11069-023-06277-3

Anbalagan, R. (1992). Landslide hazard evaluation and zonation mapping in mountainous terrain. *Engineering Geology*, *32*, 269–277. https://doi.org/10.1016/0013-7952(92)90053-2

Aristizábal, E., López, S., Sánchez, O., Vásquez, M., Rincón, F., Ruiz-Vásquez, D., Restrepo, S., & Valencia, J. S. (2019). Evaluación de la amenaza por movimientos en masa detonados por lluvias para una región de los andes colombianos estimando la probabilidad espacial, temporal, y magnitud. *Boletin de geologia*, *41*(3), 85–105.

Aristizábal, E., Riaño, F., & Jiménez-Ortiz, J. (2022). Umbrales de lluvia como detonante de movimientos en masa en el piedemonte de la cordillera central en los andes colombianos. *Boletín de Geología*, *44*(2), 183–197.

Aristizábal, E. V., Correa, D., Cadavid, C. G., Londoño, A. M. V., & Benavides, E. A. Z. (2025). Geohazard: Base de datos por movimientos en masa y avenidas torrenciales en colombia y el departamento de antioquia. *Ciencia e Ingeniería Neogranadina*, *35*(1), 131–147.

Ayalew, L., & Yamagishi, H. (2005). The application of gis-based logistic regression for landslide susceptibility mapping in the kakuda-yahiko mountains, central japan. *Geomorphology*, *65*(1-2), 15–31.

Bai, J., Guo, K., Liu, M., & Jiang, T. (2023). Spatial variability, evolution, and agglomeration of eco-environmental risks in the yangtze river economic belt, china. *Ecological Indicators*, *152*, 110375.

Bai, L., Shi, C., Li, L., Yang, Y., & Wu, J. (2018). Accuracy of chirps satellite-rainfall products over mainland china. *Remote sensing*, *10*(3), 362.

Barnes, R. (2018). *Richdem: High-performance terrain analysis* (tech. rep.). PeerJ Preprints.

Baum, R. L., Savage, W. Z., & Godt, J. W. (2008). *Trigrs: A fortran program for transient rainfall infiltration and grid-based regional slope-stability analysis, version 2.0.* US Geological Survey Reston, VA, USA.

Berti, M., Pizziolo, M., Scaroni, M., Generali, M., Critelli, V., Mulas, M., Tondo, M., Lelli, F., Fabbiani, C., Ronchetti, F., et al. (2025). Rer2023: The landslide inventory dataset of the may 2023 emilia-romagna meteorological event. *Earth System Science Data*, *17*(3), 1055–1074.

Brenning, A., Schwinn, M., Ruiz-Páez, A., & Muenchow, J. (2015). Landslide susceptibility near highways is increased by 1 order of magnitude in the andes of southern ecuador, loja province. *Natural Hazards and Earth System Sciences*, *15*(1), 45–57.

Brenning, A. (2012). Spatial cross-validation and bootstrap for the assessment of prediction rules in remote sensing: The r package sperrorest. *2012 IEEE international geoscience and remote sensing symposium*, 5372–5375.

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly weather review*, *78*(1), 1–3.

Brunetti, M. T., Peruccacci, S., Rossi, M., Luciani, S., Valigi, D., & Guzzetti, F. (2010). Rainfall thresholds for the possible occurrence of landslides in italy. *Natural Hazards and Earth System Sciences*, *10*, 447–458. https://doi.org/10.5194/nhess-10-447-2010

Caine, N. (1980). The rainfall intensity-duration control of shallow landslides and debris flows. *Geografiska annaler: series A, physical geography*, *62*(1-2), 23–27.

Carocol Radio. (2023). *Cerrada la vía que comunica a manizales con medellín*. Retrieved June 17, 2025, from https://caracol.com.co/2023/01/13/cerrada-la-via-que-comunica-a-manizales-con-medellin/

Carrara, A., Cardinali, M., Detti, R., Guzzetti, F., Pasqui, V., & Reichenbach, P. (1991). Gis techniques and statistical models in evaluating landslide hazard. *Earth Surface Processes and Landforms*, *16*, 427–445. https://doi.org/10.1002/esp.3290160505

Carvajal, J. H. (2012). *Propuesta de estandarización de la cartografía geomorfológica en colombia*. Servicio Geológico Colombiano.

Casagli, N., Intrieri, E., Tofani, V., Gigli, G., & Raspini, F. (2023). Landslide detection, monitoring and prediction with remote-sensing techniques. *Nature Reviews Earth & Environment*, *4*, 51–64. https://doi.org/10.1038/s43017-022-00373-x

Chae, B.-G., Park, H.-J., Catani, F., Simoni, A., & Berti, M. (2017). Landslide prediction, monitoring and early warning: A concise review of state-of-the-art. *Geosciences Journal*, *21*, 1033–1070.

Cheng, D., Cui, Y., Su, F., Jia, Y., & Choi, C. E. (2018). The characteristics of the mocoa compound disaster event, colombia. *Landslides*, *15*, 1223–1232.

Corominas, J., van Westen, C., Frattini, P., Cascini, L., Malet, J.-P, Fotopoulou, S., Catani, F., Eeck-haut, M. V. D., Mavrouli, O., Agliardi, F., Pitilakis, K., Winter, M. G., Pastor, M., Ferlisi, S., Tofani, V., Hervás, J., & Smith, J. T. (2013). Recommendations for the quantitative analysis of landslide risk. *Bulletin of Engineering Geology and the Environment*, *73*, 209–263. https://doi.org/10.1007/s10064-013-0538-8

Correa, O., García, F., Bernal, G., Cardona, O. D., & Rodriguez, C. (2020). Early warning system for rainfall-triggered landslides based on real-time probabilistic hazard assessment. *Natural Hazards*, *100*, 345–361. https://doi.org/10.1007/s11069-019-03815-w

Crippen, R., Buckley, S., Agram, P., Belz, E., Gurrola, E., Hensley, S., Kobrick, M., Lavalle, M., Martin, J., Neumann, M., et al. (2016). Nasadem global elevation model: Methods and progress. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *41*, 125–128.

Cruden, D., & Varnes, D. (1996). Landslide types and processes. *Landslides, Investigation and Mitigation: Special Report*, *247*, 36–75. https://www.researchgate.net/publication/269710331_CrudenDMVarnes_DJ1996_Landslide_Types_and_Processes_Transportation_Research_Board_US_National_Academy_of_Sciences_Special_Report_247_36-75

Departamento Nacional de Planeación. (2021). *Política para el desarrollo de la infraestructura vial terciaria: Documento conpes 4023* (tech. rep.). Consejo Nacional de Política Económica y Social – DNP. Bogotá. https://www.dnp.gov.co/Conpes/Documents/Conpes_4023_2021.pdf

Desinventar. (2022). Disaster inventory system. https://www.desinventar.org/

Domínguez-Calle, E., Lozano-Báez, S., & Conceptual, M. (2014). Estado del arte de los sistemas de alerta temprana en colombia ciencias de la tierra. *Revista Academia Colombiana de Ciencias*, *38*, 321–332.

Donnini, M., Napolitano, E., Salvati, P., Ardizzone, F., Bucci, F., Fiorucci, F., Santangelo, M., Cardinali, M., & Guzzetti, F. (2017). Impact of event landslides on road networks: A statistical analysis of two italian case studies. *Landslides*, *14*, 1521–1535. https://doi.org/10.1007/S10346-017-0829-4/TABLES/7

Dotta, G., Fornaciai, A., Bertolini, G., Isola, I., Nannipieri, L., Favalli, M., Burrato, P., Devoti, R., Gigli, G., Mucchi, L., et al. (2023). Geomorphology of the upper sector of the roncovetro active landslide (emilia-romagna region, italy). *Journal of Maps*, *19*(1), 1–11.

Eeckhaut, M. V. D., Reichenbach, P., Guzzetti, F., Rossi, M., & Poesen, J. (2009). Combined landslide inventory and susceptibility assessment based on different mapping units: An example from the flemish ardennes, belgium. *Natural Hazards and Earth System Sciences*, *9*, 507–521. https://doi.org/10.5194/nhess-9-507-2009

El Pereirano. (2023). *Vía entre risaralda y chocó, cada vez más peligrosa*. Retrieved June 17, 2025, from https://elpereirano.com/2023/01/12/via-entre-risaralda-y-choco-cada-vez-mas-peligrosa/

Fang, Z., Tanyas, H., Gorum, T., Dahal, A., Wang, Y., & Lombardo, L. (2023). Speech-recognition in landslide predictive modelling: A case for a next generation early warning system. *Environmental Modelling & Software*, *170*, 105833.

Fathani, T. F., Karnawati, D., & Wilopo, W. (2016). An integrated methodology to develop a standard for landslide early warning systems. *Natural Hazards and Earth System Sciences*, *16*, 2123–2135. https://doi.org/10.5194/nhess-16-2123-2016

Fernández-Raga, M., García-Díez, I., Campo, J., Viejo, J., & Palencia, C. (2021). Effectiveness of a new drainage system for decreasing erosion in road hillslopes. *Air, Soil and Water Research*, *14*, 1178622120988722.

Ferrario, M. F., & Livio, F. (2023). Rapid mapping of landslides induced by heavy rainfall in the emilia-romagna (italy) region in may 2023. *Remote Sensing*, *16*(1), 122.

Froude, M. J., & Petley, D. N. (2018). Global fatal landslide occurrence from 2004 to 2016. *Natural Hazards and Earth System Sciences*, *18*, 2161–2181. https://doi.org/10.5194/nhess-18-2161-2018

Funk, C., Peterson, P., Landsfeld, M., et al. (2023). *Chirps–gefs: Ensemble mean 5- to 15-day rainfall forecasts* [Version accessed 17 June 2025].

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., & Michaelsen, J. (2015). The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific Data*, *2*, 150066. https://doi.org/10.1038/sdata.2015.66

Gaidzik, K., & Ramírez-Herrera, M. T. (2021). The importance of input data on landslide susceptibility mapping. *Scientific reports*, *11*(1), 19334.

Galli, M., Ardizzone, F., Cardinali, M., Guzzetti, F., & Reichenbach, P. (2008). Comparing landslide inventory maps. *Geomorphology*, *94*(3-4), 268–289.

Garcia-Delgado, H., Petley, D. N., Bermúdez, M. A., & Sepúlveda, S. A. (2022). Fatal landslides in colombia (from historical times to 2020) and their socio-economic impacts. *Landslides*, *19*, 1689–1716. https://doi.org/10.1007/s10346-022-01870-2

García-Delgado, H., Machuca, S., & Medina, E. (2019). Dynamic and geomorphic characterizations of the mocoa debris flow (march 31, 2017, putumayo department, southern colombia). *Landslides*, *16*(3), 597–609.

Gariano, S. L., & Guzzetti, F. (2016). Landslides in a changing climate. *Earth-Science Reviews*, *162*, 227–252. https://doi.org/10.1016/j.earscirev.2016.08.011

Gini, C. (1921). Measurement of inequality of incomes. *The economic journal*, *31*(121), 124–125.

Glade, T. (2003). Landslide occurrence as a response to land use change: A review of evidence from new zealand. *Catena*, *51*(3-4), 297–314.

Glade, T., & Nadim, F. (2014). Early warning systems for natural hazards and risks. *Natural Hazards*, *70*, 1669–1671. https://doi.org/10.1007/s11069-013-1000-8

Goetz, J. N., Guthrie, R. H., & Brenning, A. (2011). Integrating physical and empirical landslide susceptibility models using generalized additive models. *Geomorphology*, *129*(3-4), 376–386.

Gómez, D., Aristizábal, E., García, E. F., Marín, D., Valencia, S., & Vásquez, M. (2023a). Landslides forecasting using satellite rainfall estimations and machine learning in the colombian andean region. *Journal of South American Earth Sciences*, *125*, 104293. https://doi.org/10.1016/j.jsames.2023.104293

Gómez, D., García, E. F., & Aristizábal, E. (2023b). Spatial and temporal landslide distributions using global and open landslide databases. *Natural Hazards*, *117*(1), 25–55.

Gómez, J., Montes Ramírez, N. E., & Mateus-Zabala, D. (2025). Mapping the geology of colombia. *Geological Society, London, Special Publications*, *558*(1), SP558–2024.

Gupta, S. K., & Shukla, D. P. (2023). Handling data imbalance in machine learning based landslide susceptibility mapping: A case study of mandakini river basin, north-western himalayas. *Landslides*, *20*(5), 933–949.

Guzzetti, F., Cardinali, M., Reichenbach, P., & Carrara, A. (2000). Comparing landslide maps: A case study in the upper tiber river basin, central italy. *Environmental management*, *25*(3).

Guzzetti, F., Carrara, A., Cardinali, M., & Reichenbach, P. (1999). Landslide hazard evaluation: A review of current techniques and their application in a multi-scale study, central italy. *Geomorphology*, *31*, 181–216. https://doi.org/10.1016/S0169-555X(99)00078-1

Guzzetti, F., Gariano, S. L., Peruccacci, S., Brunetti, M. T., Marchesini, I., Rossi, M., & Melillo, M. (2020). Geographical landslide early warning systems. *Earth-Science Reviews*, *200*, 102973. https://doi.org/10.1016/j.earscirev.2019.102973

Guzzetti, F., Mondini, A. C., Cardinali, M., Fiorucci, F., Santangelo, M., & Chang, K.-T. (2012). Landslide inventory maps: New tools for an old problem. *Earth-Science Reviews*, *112*(1-2), 42–66.

Hallegatte, S., Vogt-Schilb, A., Bangalore, M., & Rozenberg, J. (2017). *Unbreakable: Building the resilience of the poor in the face of natural disasters*. Washington, DC: World Bank. https://doi.org/10.1596/978-1-4648-1003-9

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology*, *143*(1), 29–36.

Hastie, T. J. (2017). Generalized additive models. *Statistical models in S*, 249–307.

Herd, D. G. (1986). The 1985 ruiz volcano disaster. *Eos, Transactions American Geophysical Union*, *67*, 457–460. https://doi.org/10.1029/EO067i019p00457-03

Herrera, G., Mateos, R. M., García-Davalillo, J. C., Grandjean, G., Poyiadji, E., Maftei, R., Filipciuc, T.-C., Jemec Auflič, M., Jež, J., Podolszki, L., et al. (2018). Landslide databases in the geological surveys of europe. *Landslides*, *15*, 359–379.

Herrera-Coy, M. C., Calderón, L. P., Herrera-Pérez, I. L., Bravo-López, P. E., Conoscenti, C., Delgado, J., Sánchez-Gómez, M., & Fernández, T. (2023). Landslide susceptibility analysis on the vicinity of bogotá-villavicencio road (eastern cordillera of the colombian andes). *Remote sensing*, *15*(15), 3870.

Hervás, J., & Bobrowsky, P. (2009). Mapping: Inventories, susceptibility, hazard and risk. In *Landslides– disaster risk reduction* (pp. 321–349). Springer.

Hoyos, I., Dominguez, F., Cañón-Barriga, J., Martínez, J. A., Nieto, R., Gimeno, L., & Dirmeyer, P. A. (2018). Moisture origin and transport processes in colombia, northern south america. *Climate Dynamics*, *50*, 971–990. https://doi.org/10.1007/s00382-017-3653-6

Huang, Y., & Zhao, L. (2018). Review on landslide susceptibility mapping using support vector machines. *Catena*, *165*, 520–529. https://doi.org/10.1016/j.catena.2018.03.003

IDEAM. (2008). *Protocolo para la emisión de los pronósticos hidrológicos* (tech. rep.). IDEAM.

Infobae. (2023). *Cierre total de la vía quibdó–pereira por derrumbes*. Retrieved June 17, 2025, from https://www.infobae.com/america/colombia/2023/01/10/cierre-total-de-la-via-quibdo-pereira-por-derrumbes/

Instituto Nacional de Vías (INVÍAS). (2023). *Boletín estadístico de infraestructura vial 2023* (tech. rep.). INVÍAS. Bogotá. https://www.invias.gov.co/boletin-estadistico-2023

INVIAS. (2023). *Avance 1 - 3:30 p.m. invías habilita paso provisional para vehículos livianos y continúa con trabajos de remoción en la vía quibdó - pereira en el km 16 entre santa cecilia y pueblo rico*. Retrieved June 17, 2025, from https://www.invias.gov.co/index.php/sala/noticias/4950-avance-

1-3-30-p-m-invias-habilita-paso-provisional-para-vehiculos-livianos-y-continua-con-trabajos-de-remocion-en-la-via-quibdo-pereira-en-el-km-16-entre-santa-cecilia-y-pueblo-rico#:~:text=Bogot%C3%A1%2C%20D.%20C.%2C%2010%20de%20enero, se%20sugiere%20como%20v%C3%ADa%20alterna:

IPCC. (2012). *Managing the risks of extreme events and disasters to advance climate change adaptation. a special report of working groups i and ii of the intergovernmental panel on climate change* (C. B. Field, V. Barros, T. F. Stocker, Q. Dahe, D. J. Dokken, K. L. Ebi, M. D. Mastrandrea, K. J. Mach, G.-K. Plattner, S. K. Allen, M. Tignor, & P. M. Midgley, Eds.). Cambridge University Press.

IPCC. (2023, July). *Ipcc, 2023: Climate change 2023: Synthesis report. contribution of working groups i, ii and iii to the sixth assessment report of the intergovernmental panel on climate change* (P. Arias, M. Bustamante, I. Elgizouli, G. Flato, M. Howden, C. Méndez-Vallejo, J. J. Pereira, R. Pichs-Madruga, S. K. Rose, Y. Saheb, R. S. Rodríguez, D. Ürge-Vorsatz, C. Xiao, N. Yassaa, J. Romero, J. Kim, E. F. Haites, Y. Jung, R. Stavins, … C. Péan, Eds.; tech. rep.). Intergovernmental Panel on Climate Change. https://doi.org/10.59327/IPCC/AR6-9789291691647

Jakob, M., & Weatherly, H. (2003). A hydroclimatic threshold for landslide initiation on the north shore mountains of vancouver, british columbia. *Geomorphology*, *54*(3-4), 137–156.

Jasiewicz, J., & Stepinski, T. F. (2013). Geomorphons—a pattern recognition approach to classification and mapping of landforms. *Geomorphology*, *182*, 147–156.

Kargel, J. S., Leonard, G. J., Shugar, D. H., Haritashya, U. K., Bevington, A., Fielding, E. J., Fujita, K., Geertsema, M., Miles, E., Steiner, J., et al. (2016). Geomorphic and geologic controls of geohazards induced by nepal's 2015 gorkha earthquake. *Science*, *351*(6269), aac8353.

Kirschbaum, D. B., Adler, R., Hong, Y., Hill, S., & Lerner-Lam, A. (2010). A global landslide catalog for hazard applications: Method, results, and limitations. *Natural Hazards*, *52*, 561–575.

Klose, M., Damm, B., & Terhorst, B. (2015). Landslide cost modeling for transportation infrastructures: A methodological approach. *Landslides*, *12*, 321–334.

Kruschel, S., Hambauer, N., Weinzierl, S., Zilker, S., Kraus, M., & Zschech, P. (2025). Challenging the performance-interpretability trade-off: An evaluation of interpretable machine learning models. *Business & Information Systems Engineering*, 1–25.

Lien, G.-Y., Kalnay, E., Miyoshi, T., & Huffman, G. J. (2016). Statistical properties of global precipitation in the ncep gfs model and tmpa observations for data assimilation. *Monthly Weather Review*, *144*(2), 663–679.

Liu, L., Deng, J., & Tang, Y. (2023). A dynamic management and integration framework for models in landslide early warning system. *ISPRS International Journal of Geo-Information*, *12*(5), 198.

Lizarazo, S. C., Sagiya, T., & Mora-Páez, H. (2021). Interplate coupling along the caribbean coast of colombia and its implications for seismic/tsunami hazards. *Journal of South American Earth Sciences*, *110*, 103332.

Lombardo, L., Opitz, T., Ardizzone, F., Guzzetti, F., & Huser, R. (2020). Space-time landslide predictive modelling. *Earth-Science Reviews*, *209*, 103318. https://doi.org/10.1016/j.earscirev.2020.103318

López-Bermeo, C., Montoya, R. D., Caro-Lopera, F. J., & Díaz-García, J. A. (2022). Validation of the accuracy of the chirps precipitation dataset at representing climate variability in a tropical mountainous region of south america. *Physics and Chemistry of the Earth, Parts A/B/C, 127*, 103184.

Malamud, B. D., Turcotte, D. L., Guzzetti, F., & Reichenbach, P. (2004). Landslide inventories and their statistical properties. *Earth Surface Processes and Landforms*, *29*, 687–711. https://doi.org/10.1002/esp.1064

Marc, O., Behling, R., Andermann, C., Turowski, J. M., Illien, L., Roessner, S., & Hovius, N. (2019). Long-term erosion of the nepal himalayas by bedrock landsliding: The role of monsoons, earthquakes and giant landslides. *Earth Surface Dynamics*, *7*(1), 107–128.

Marin, R. J., & Marin-Sanchez, J. C. (2024). Landscient-ews: Real-time monitoring of rainfall thresholds for landslide early warning - a case study in the colombian andes. *The Korean Society of Engineering Geology*, *34*, 173–191. https://doi.org/https://doi.org/10.9720/kseg.2024.2.173

Marin, R. J., Velásquez, M. F., García, E. F., Alvioli, M., & Aristizábal, E. (2021). Assessing two methods of defining rainfall intensity and duration thresholds for shallow landslides in data-scarce catchments of the colombian andean mountains. *Catena*, *206*, 105563.

McAdoo, B. G., Quak, M., Gnyawali, K. R., Adhikari, B. R., Devkota, S., Rajbhandari, P. L., & Sudmeier-Rieux, K. (2018). Roads and landslides in nepal: How development affects environmental risk. *Natural Hazards and Earth System Sciences*, *18*(12), 3203–3210.

Michaelis, L., Menten, M. L., et al. (1913). Die kinetik der invertinwirkung. *Biochem. z*, *49*(333-369), 352.

Ministerio de Transporte. (2019). *Plan maestro de transporte intermodal 2015-2035 (pmti)* (tech. rep.) (Strategic master plan summarising national road classes and investment needs). República de Colombia, Ministerio de Transporte. Bogotá. https://www.mintransporte.gov.co/pmti

Ministerio de Transporte. (2023). *Transporte en cifras 2022 - anuario nacional de transporte* (tech. rep.). Ministerio de Transporte. www.mintransporte.gov.co

Montgomery, D. R., & Dietrich, W. E. (1994). A physically based model for the topographic control on shallow landsliding. *Water Resources Research*, *30*, 1153–1171. https://doi.org/10.1029/93WR02979

Moreno, M., Lombardo, L., Crespi, A., Zellner, P. J., Mair, V., Pittore, M., van Westen, C., & Steger, S. (2024). Space-time data-driven modeling of precipitation-induced shallow landslides in south tyrol, italy. *Science of The Total Environment*, *912*, 169166. https://doi.org/10.1016/J.SCITOTENV.2023.169166

Neußner, O. (2021). Early warning alerts for extreme natural hazard events: A review of worldwide practices. *International Journal of Disaster Risk Reduction*, *60*, 102295.

Nguyen, B. Q., Vo, N. D., Le, M.-H., Nguyen, Q.-D., Lakshmi, V., Bolten, J. D., et al. (2023). Quantification of global digital elevation model (dem)–a case study of the newly released nasadem for a river basin in central vietnam. *Journal of Hydrology: Regional Studies*, *45*, 101282.

Nocentini, N., Rosi, A., Segoni, S., & Fanti, R. (2023). Towards landslide space-time forecasting through machine learning: The influence of rainfall parameters and model setting. *Frontiers in Earth Science*, *11*, 1152130.

Osanai, N., Shimizu, T., Kuramoto, K., Kojima, S., & Noro, T. (2010). Japanese early-warning for debris flows and slope failures using rainfall indices with radial basis function network. *Landslides*, *7*, 325–338.

Pack, R. T., Tarboton, D. G., & Goodwin, C. N. (1998). The sinmap approach to terrain stability mapping. *8th Congress of the International Association of Engineering Geology*, *2*, 1157–1166.

Parker, R. N., Rosser, N. J., & Hales, T. C. (2017). Spatial prediction of earthquake-induced landslide probability. *Natural Hazards and Earth System Sciences Discussions*, *2017*, 1–29.

Pedersen, E. J., Miller, D. L., Simpson, G. L., & Ross, N. (2019). Hierarchical generalized additive models in ecology: An introduction with mgcv. *PeerJ*, *2019*, e6876. https://doi.org/10.7717/PEERJ.6876/SUPP-1

Peñas, V. (2017). La catástrofe de mocoa (colombia): Un desastre anunciado previsible que podía haberse evitado. *Recuperado: https://www. cienciasagrarias. medellin. unal. edu. co.*

Peruccacci, S., Brunetti, M. T., Luciani, S., Vennari, C., & Guzzetti, F. (2012). Lithological and seasonal control on rainfall thresholds for the possible initiation of landslides in central italy. *Geomorphology*, *139*, 79–90.

Piciullo, L., Calvello, M., & Cepeda, J. M. (2018). Territorial early warning systems for rainfall-induced landslides. *Earth-Science Reviews*, *179*, 228–247. https://doi.org/10.1016/j.earscirev.2018.02.013

Piciullo, L., Gariano, S. L., Melillo, M., Brunetti, M. T., Peruccacci, S., Guzzetti, F., & Calvello, M. (2017). Definition and performance of a threshold-based regional early warning model for rainfall-induced landslides. *Landslides*, *14*(3), 995–1008.

Pollock, W., & Wartman, J. (2020). Human vulnerability to landslides. *GeoHealth*, *4*. https://doi.org/10.1029/2020GH000287

Poveda, G., Álvarez, D. M., & Rueda, Ó. A. (2011). Hydro-climatic variability over the andes of colombia associated with enso: A review of climatic processes and their impact on one of the earth's most important biodiversity hotspots. *Climate Dynamics*, *36*, 2233–2249. https://doi.org/10.1007/s00382-010-0931-y

Pulido, N. (2003). Seismotectonics of the northern andes (colombia) and the development of seismic networks. *Bulletin of the International Institute of Seismology and Earthquake Engineering*, 69–76.

Ramampiandra, E. C., Scheidegger, A., Wydler, J., & Schuwirth, N. (2023). A comparison of machine learning and statistical species distribution models: Quantifying overfitting supports model interpretation. *Ecological Modelling*, *481*, 110353.

Reichenbach, P., Rossi, M., Malamud, B. D., Mihir, M., & Guzzetti, F. (2018). A review of statistically-based landslide susceptibility models. *Earth-Science Reviews*, *180*, 60–91. https://doi.org/10.1016/J.EARSCIREV.2018.03.001

Rey, S. J., & Smith, R. J. (2013). A spatial decomposition of the gini coefficient. *Letters in Spatial and Resource Sciences*, *6*(2), 55–70.

Roback, K., Clark, M. K., West, A. J., Zekkos, D., Li, G., Gallen, S. F., Chamlagain, D., & Godt, J. W. (2018). The size, distribution, and mobility of landslides caused by the 2015 mw7. 8 gorkha earthquake, nepal. *Geomorphology*, *301*, 121–138.

Rossi, M., Guzzetti, F., Salvati, P., Donnini, M., Napolitano, E., & Bianchi, C. (2019). A predictive model of societal landslide risk in italy. *Earth-Science Reviews*, *196*, 102849. https://doi.org/10.1016/j.earscirev.2019.04.021

Sahrane, R., Bounab, A., & Kharim, Y. E. (2023). Investigating the effects of landslides inventory completeness on susceptibility mapping and frequency-area distributions: Case of taounate province, northern morocco. *Catena*, *220*, 106737.

Schlögl, M., Spiekermann, R., & Steger, S. (2025). Towards a holistic assessment of landslide susceptibility models: Insights from the central eastern alps. *Environmental Earth Sciences*, *84*(4), 113.

Schumann, A., & Polygon, T. (1998). Encyclopedia of hydrology and lakes. encyclopedia of earth science.

Segoni, S., Battistini, A., Rossi, G., Rosi, A., Lagomarsino, D., Catani, F., Moretti, S., & Casagli, N. (2015). An operational landslide early warning system at regional scale based on space–time-variable rainfall thresholds. *Natural Hazards and Earth System Sciences*, *15*(4), 853–861.

Sepúlveda, S. A., & Petley, D. N. (2015). Regional trends and controlling factors of fatal landslides in latin america and the caribbean. *Natural Hazards and Earth System Sciences*, *15*, 1821–1833. https://doi.org/10.5194/nhess-15-1821-2015

SGC. (2024). Sistema de información de movimientos en masa. http://simma.sgc.gov.co/

Shrestha, N. K., Qamer, F. M., Pedreros, D., Murthy, M., Wahid, S. M., & Shrestha, M. (2017). Evaluating the accuracy of climate hazard group (chg) satellite rainfall estimates for precipitation based drought monitoring in koshi basin, nepal. *Journal of Hydrology: Regional Studies*, *13*, 138–151.

SIATA. (2024, July). Sistema de alerta temprana de medellín y el valle de aburrá.

Sinčić, M., Bernat Gazibara, S., Rossi, M., & Mihalić Arbanas, S. (2024). Comparison of conditioning factors classification criteria in large scale statistically based landslide susceptibility models. *Natural Hazards and Earth System Sciences Discussions*, *2024*, 1–42.

Steger, S., Brenning, A., Bell, R., & Glade, T. (2017). The influence of systematically incomplete shallow landslide inventories on statistical susceptibility models and suggestions for improvements. *Landslides*, *14*, 1767–1781.

Steger, S., & Glade, T. (2017). The challenge of "trivial areas" in statistical landslide susceptibility modelling. *Advancing Culture of Living with Landslides: Volume 2 Advances in Landslide Science*, 803–808.

Steger, S., Moreno, M., Crespi, A., Gariano, S. L., Brunetti, M. T., Melillo, M., Peruccacci, S., Marra, F., de Vugt, L., Zieher, T., Rutzinger, M., Mair, V., & Pittore, M. (2024). Adopting the margin of stability for space–time landslide prediction – a data-driven approach for generating spatial dynamic thresholds. *Geoscience Frontiers*, *15*, 101822. https://doi.org/10.1016/j.gsf.2024.101822

Steger, S., Moreno, M., Crespi, A., Zellner, P. J., Gariano, S. L., Brunetti, M. T., Melillo, M., Peruccacci, S., Marra, F., Kohrs, R., et al. (2023). Deciphering seasonal effects of triggering and preparatory precipitation for improved shallow landslide prediction using generalized additive mixed models. *Natural Hazards and Earth System Sciences Discussions*, *2022*, 1–38.

Tanyas, H., & Lombardo, L. (2020). Completeness index for earthquake-induced landslide inventories. *Engineering Geology*, *264*, 105331. https://doi.org/10.1016/j.enggeo.2019.105331

Tanyaş, H., van Westen, C. J., Allstadt, K. E., Jessee, M. A. N., Görüm, T., Jibson, R. W., Godt, J. W., Sato, H. P., Schmitt, R. G., Marc, O., & Hovius, N. (2017). Presentation and analysis

of a worldwide database of earthquake☐induced landslide inventories. *Journal of Geophysical Research: Earth Surface*, *122*, 1991–2015. https://doi.org/10.1002/2017JF004236

Tapias, J. G., Schobbenhaus, C., Ramírez, N. E. M., Gutiérrez, F. A. A., & Zabala, D. M. (2023). Mapping the geology of south america. *Episodes Journal of International Geoscience*, *46*(4), 537–549.

Terrell, G. R., & Scott, D. W. (1992). Variable kernel density estimation. *The Annals of Statistics*, 1236–1265.

Thirugnanam, H., Ramesh, M. V., & Rangan, V. P. (2020). Enhancing the reliability of landslide early warning systems by machine learning. *Landslides*, *17*, 2231–2246. https://doi.org/10.1007/s10346-020-01453-z

UNGRD. (2023a). *Reporte diario situacional – 14 enero 2023*. Retrieved June 17, 2025, from https://portal.gestiondelriesgo.gov.co/Documents/Manejo/Reportes/2023-01-14-REPORTE-DIARIO-SITUACIONAL.pdf

UNGRD. (2023b). *Reporte diario situacional – 6 enero 2023*. Retrieved June 17, 2025, from https://portal.gestiondelriesgo.gov.co/Documents/Manejo/Reportes/2023-01-06-REPORTE-DIARIO-SITUACIONAL.pdf

UNISDR. (2015, March). *Sendai framework for disaster risk reduction 2015 - 2030* (tech. rep.). United Nations.

Urrea, V., Ochoa, A., & Mesa, O. (2019). Seasonality of rainfall in colombia. *Water Resources Research*, *55*, 4149–4162. https://doi.org/10.1029/2018WR023316

Uuemaa, E., Ahi, S., Montibeller, B., Muru, M., & Kmoch, A. (2020). Vertical accuracy of freely available global digital elevation models (aster, aw3d30, merit, tandem-x, srtm, and nasa-dem). remote sens 12 (21): 3482.

Van Den Eeckhaut, M., & Hervás, J. (2012). State of the art of national landslide databases in europe and their potential for assessing landslide susceptibility, hazard and risk. *Geomorphology*, *139*, 545–558.

Van Westen, C. J., Castellanos, E., & Kuriakose, S. L. (2008). Spatial data for landslide susceptibility, hazard, and vulnerability assessment: An overview. *Engineering geology*, *102*(3-4), 112–131.

Vargas-Cuervo, G., Rotigliano, E., & Conoscenti, C. (2019). Prediction of debris-avalanches and-flows triggered by a tropical storm by using a stochastic approach: An application to the events occurred in mocoa (colombia) on 1 april 2017. *Geomorphology*, *339*, 31–43.

Vega, J., Barco, J., & Hidalgo, C. (2024). Space-time analysis of the relationship between landslides occurrence, rainfall variability and enso in the tropical andean mountain region in colombia. *Landslides*, *21*(6), 1293–1314.

Wang, T., Dahal, A., Fang, Z., van Westen, C., Yin, K., & Lombardo, L. (2024). From spatio-temporal landslide susceptibility to landslide risk forecast. *Geoscience Frontiers*, *15*(2), 101765.

Wieczorek, G. F., & Glade, T. (2005, December). Climatic factors influencing occurrence of debris flows. In *Debris-flow hazards and related phenomena* (pp. 325–362). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-27129-5_14

Winter, M. G., Shearer, B., Palmer, D., Peeling, D., Harmer, C., & Sharpe, J. (2016). The economic impact of landslides and floods on the road network. *Procedia Engineering*, *143*, 1425–1434. https://doi.org/10.1016/j.proeng.2016.06.168

WMO. (2022). *Early warnings for all: The un global early warning initiative for the implementation of climate adaptation - executive action plan 2023-2027* (tech. rep.). United Nations. https://library.wmo.int/idurl/4/58209

WMO. (2024). *State of the global climate 2023* (tech. rep.). World Metereological Organization.

Wood, S. N. (2017). *Generalized additive models: An introduction with r*. chapman; hall/CRC.

World Bank. (2016). *Logistics, infrastructure and competitiveness in colombia: Diagnostic for resilient transport* (tech. rep.) (Highlights landslide disruption costs on primary corridors). World Bank Group. Washington, DC. https://openknowledge.worldbank.org/handle/10986/25095

Wu, W., Li, Y., Luo, X., Zhang, Y., Ji, X., & Li, X. (2019). Performance evaluation of the chirps precipitation dataset and its utility in drought monitoring over yunnan province, china. *Geomatics, Natural Hazards and Risk*, *10*(1), 2145–2162.

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, *3*(1), 32–35. https://doi.org/10.1002/1097-0142(1950)3:1<32::aid-cncr2820030106>3.0.co;2-3

Yue, H., Gebremichael, M., & Nourani, V. (2022). Performance of the global forecast system's medium-range precipitation forecasts in the niger river basin using multiple satellite-based products. *Hydrology and Earth System Sciences*, *26*(1), 167–181.

Zakaria, M. T., Muztaza, N. M., Zabidi, H., Salleh, A. N., Mahmud, N., & Rosli, F. N. (2022). Integrated analysis of geophysical approaches for slope failure characterisation. *Environmental Earth Sciences*, *81*, 299. https://doi.org/10.1007/s12665-022-10410-z

Zanaga, D., Van De Kerchove, R., Daems, D., De Keersmaecker, W., Brockmann, C., Kirches, G., Wevers, J., Cartus, O., Santoro, M., Fritz, S., et al. (2022). Esa worldcover 10 m 2021 v200.

KEYWORDS

- Landslide Data Management
- Disaster Risk Reduction (DRR)
- Data Maturity Framework
- Landslide Inventory
- Landslide Damage & Loss
- Early Warning Systems (EWS)
- Geospatial Analysis
- Multi-Hazard Integration
- Data Governance
- Colombia
- Stakeholder Collaboration
- Institutional Capacity
- Community Engagement
- SIMMA
- Disaster Resilience

ABBREVIATIONS AND SYMBOLS

- **General Abbreviations:**

  - **API:** Application Programming Interface
  - **CSV:** Comma-Separated Values
  - **DEM:** Digital Elevation Model
  - **DOI:** Digital Object Identifier
  - **DRM:** Disaster Risk Management
  - **DRR:** Disaster Risk Reduction
  - **DDRRMM:** Digital Disaster Risk Reduction Maturity Model
  - **EWS:** Early Warning System
  - **GIS:** Geographic Information System
  - **ICT:** Information and Communications Technology
  - **IoT:** Internet of Things
  - **ISO:** International Organization for Standardization
  - **JSON:** JavaScript Object Notation
  - **MoU:** Memorandum of Understanding
  - **NGO:** Non-Governmental Organization
  - **PPP:** Public-Private Partnership
  - **QA/QC:** Quality Assurance/Quality Control
  - **SMS:** Short Message Service
  - **SOPs:** Standard Operating Procedures
  - **URL:** Uniform Resource Locator

- **Colombian-specific abbreviations:**

  - **SGC:** Servicio Geológico Colombiano (Colombian Geological Service)
  - **SIMMA:** Sistema de Información de Movimientos en Masa (Landslide Information System, SGC)
  - **IDEAM:** Instituto de Hidrología, Meteorología y Estudios Ambientales (Institute of Hydrology, Meteorology, and Environmental Studies)

- **UNGRD:** Unidad Nacional para la Gestión del Riesgo de Desastres (National Unit for Disaster Risk Management)
- **INVIAS:** Instituto Nacional de Vías (National Roads Institute)
- **SNGRD:** Sistema Nacional de Gestión del Riesgo de Desastres (National Disaster Risk Management System)
- **SIATA:** Sistema de Alerta Temprana de Medellín y el Valle de Aburrá (Medellin and Aburrá Valley Early Warning System)
- **DesInventar:** Disaster Inventory Database (UN-supported disaster database widely used in Latin America)

- **Institutional Abbreviations:**

  - **GFDRR:** Global Facility for Disaster Reduction and Recovery (World Bank-supported initiative)
  - **ITC:** Faculty of Geo-Information Science and Earth Observation, University of Twente, Netherlands
  - **UGLD:** Unified Global Landslide Database
  - **UNDP:** United Nations Development Programme
  - **UNDRR:** United Nations Office for Disaster Risk Reduction (formerly UNISDR)
  - **UNISDR:** United Nations International Strategy for Disaster Reduction (previous name of UNDRR)

- **Platforms and Portals:**

  - **Geohazards:** Colombian geological hazard research group with a geoportal with Antioquia landslide occurrences. Part of the National University in Medellín
  - **Colombia en Mapas:** Colombian Geospatial Data Hub portal.

A   APPENDIX A: INVENTORIES CHARACTERIZATION

Table A.1: Provenance and custodianship of Colombian landslide inventories.

Inventory	Ownership	Key data sources	Last updated	Regular updated?	Format(s)	Access / link
Gómez et al. (2023a)	University of Antio-quia/Medellin/National	SIMMA, DesInventar, GeoHazard	2023	Yes	SHP	e-mail on request
GGarcia-Delgado et al. (2022)	Academic review	15 historical studies, SIMMA	2020	No	XLSX	https://data.mendeley.com/datasets/xbrc8gvby9/1
GeoHazards (Ant.)	National University of Colombia (Medellín)	Municipal reports, media, AMVA	2025	Irreg.	SHP/GeoJSON	https://geohazards.com.co/visor-geohazard.html
GeoHazards (Col.)	National University of Colombia (Medellín)	fire brigades, media, DESINVENTAR	2025	Irreg.	SHP/GeoJSON	same as above

(continued on next page)

Inventory	Ownership	Key data sources	Last updated	Regular updated?	Format(s)	Access / link
SIMMA (Inventory)	SGC	SGC, Universities, CARs, IDEAM	2023	No	PDF, XLSX, SHP	https://simma.sgc.gov.co/#/public/basic/
SIMMA (Catalog)	SGC	Risk units, NGOs, Responders, Media, Government organizations	2025	Irreg.	PDF, XLSX, SHP	same as above
INVIAS (Critical)	INVIAS	Road engineers	2023	No	CSV	e-mail on request
INVIAS (Emerg.)	INVIAS	Road emergency portal	2023	No	CSV/SHP/GEOJSON	https://hermes2.invias.gov.co/SIV/

Table A.2: Spatial definition of each inventory.

Inventory	Region covered	Geometry type(s)
Gómez et al. (2023a)	Andean region	Points
Garcia-Delgado et al. (2022)	Nationwide	Point
GeoHazards (Ant.)	Antioquia	Points
GeoHazards (Col.)	Nationwide	Points
SIMMA (Inventory)	Nationwide	Points + polygons
SIMMA (Catalog)	Nationwide	Points
INVIAS (Critical)	Nationwide	Points
INVIAS (Emerg.)	Nationwide	Line strings (roads)

Table A.3: Temporal coverage and granularity.

Inventory	Start-end years	Granularity	Auxiliary time fields (if any)
Gómez et al. (2023a)	1981-2023	Day	Report dates
GGarcia-Delgado et al. (2022)	1743-2020	Day	Report dates
GeoHazards (Ant.)	1880-2025	Day	Report dates
GeoHazards (Col.)	1880-2025	Day	Input, Survey, Report dates
SIMMA (Catalog)	1492-2024	Day	Survey, Report dates
INVIAS (Critical)	2018-2023	Day	Report dates
INVIAS (Emerg.)	2019-2023	Hour	Interruption dates

Table A.4: Thematic content captured by each inventory.

Inventory	Descriptive variables	Mapping technique	Trigger info	Impact indicators
Gómez et al. (2023a)	Types, sub-types; alt.; state	—	Mixed / unknown	Deaths, generic losses
Garcia-Delgado et al. (2022)	Municipality only	—	Unknown	Fatalities
GeoHazards (Ant.)	Types	Third party reports, remote sensing	Unknown for some entries	Deaths, econ. loss
GeoHazards (Col.)	Types	Third party reports	Unknown for some entries	Deaths, econ. loss
SIMMA (Inventory)	Lithology; land-use; photos	RS + geomorph. + Field validation	Trigger + inherent + contr.	Econ., env., USD value
SIMMA (Catalog)	Material; photos; alt.	RS + geomorph.	None	Damage class, value
INVIAS (Critical)	Road name; photo	Field validation	None	Road impacts
INVIAS (Emerg.)	Slope class	Field validation	None	Road impacts

Table A.5: Data-quality and completeness information.

Inventory	Quality / completeness notes	Metadata available?
Gómez et al. (2023a)	Uncertainty flags (5 235 events) - type not specified	No
Garcia-Delgado et al. (2022)	Confidence ratio	Partial (methods)
GeoHazards (Ant.)	Uncertainty flag every record; unclear kind	No
GeoHazards (Col.)	Same as Antioquia version	No
SIMMA (Inventory)	Date uncert. flagged for 3rd-party events	Format and instructions
SIMMA (Catalog)	Same as Inventory	Format and instructions
INVIAS (Critical)	No uncertainty info	No

(continued)

Inv.	Data-quality note	Meta?
INVIAS (Emerg.)	No uncertainty info	No

Table A.6: Size, spatial and temporal completeness metrics.

Inventory	Records	Characterisation (%)	Spatial (%)	A–TUCS	Temporal scope
SIMMA (Inventory)	9 175	100.0	31.8	0.734	1900–2023
SIMMA (Catalog)	35 658	83.8	29.4	0.887	1724–2024
GeoHazards (Comb.)	7 519	66.7	10.8	0.515	1880–2025

Table A.7: Impact information and data-quality metrics for the source inventories.

Inventory	Binary impact info (%)	Economic-loss info (%)	Trigger info (%)	Accuracy attributes (%)	Impact attribute description
GeoHazards (Combined)	0	27.1	59.3	54.6	Death + economic losses; some road/field details
SIMMA (Inventory)	65.3	9.9	86.6	35.0	Infrastructure, economic, environmental; quantity + USD value
SIMMA (Catalog)	32.4	2.3	0	48.1	Infrastructure, economic, environmental; quantity + USD value

Table A.8: Completeness metrics for the unified inventory before and after the post-2000 filter.

Inventory version	Records	Characterisation (%)	Spatial (%)	A-TUCS	Temporal scope	Trigger (%)	Accuracy (%)
Combined - no filter	50 156	83.0	28.8	0.831	1724-2024	66.8	50.6
Combined - post-2000 filter	17 824	87.9	29.0	0.767	2000-2024	21.8	29.8

B   APPENDIX B: SUPPLEMENTARY MATERIAL FOR MODELING AND VALIDATION STEPS
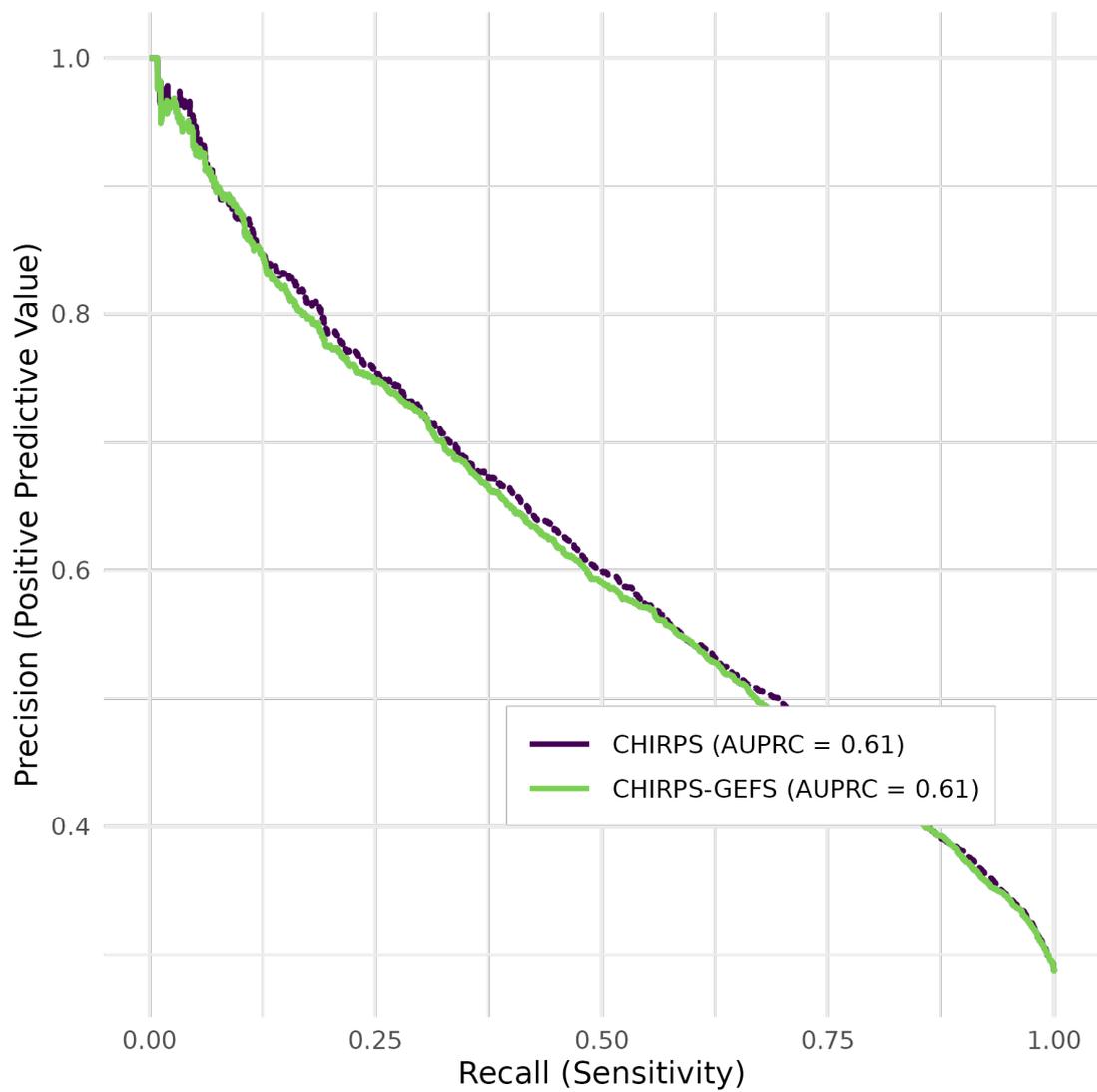

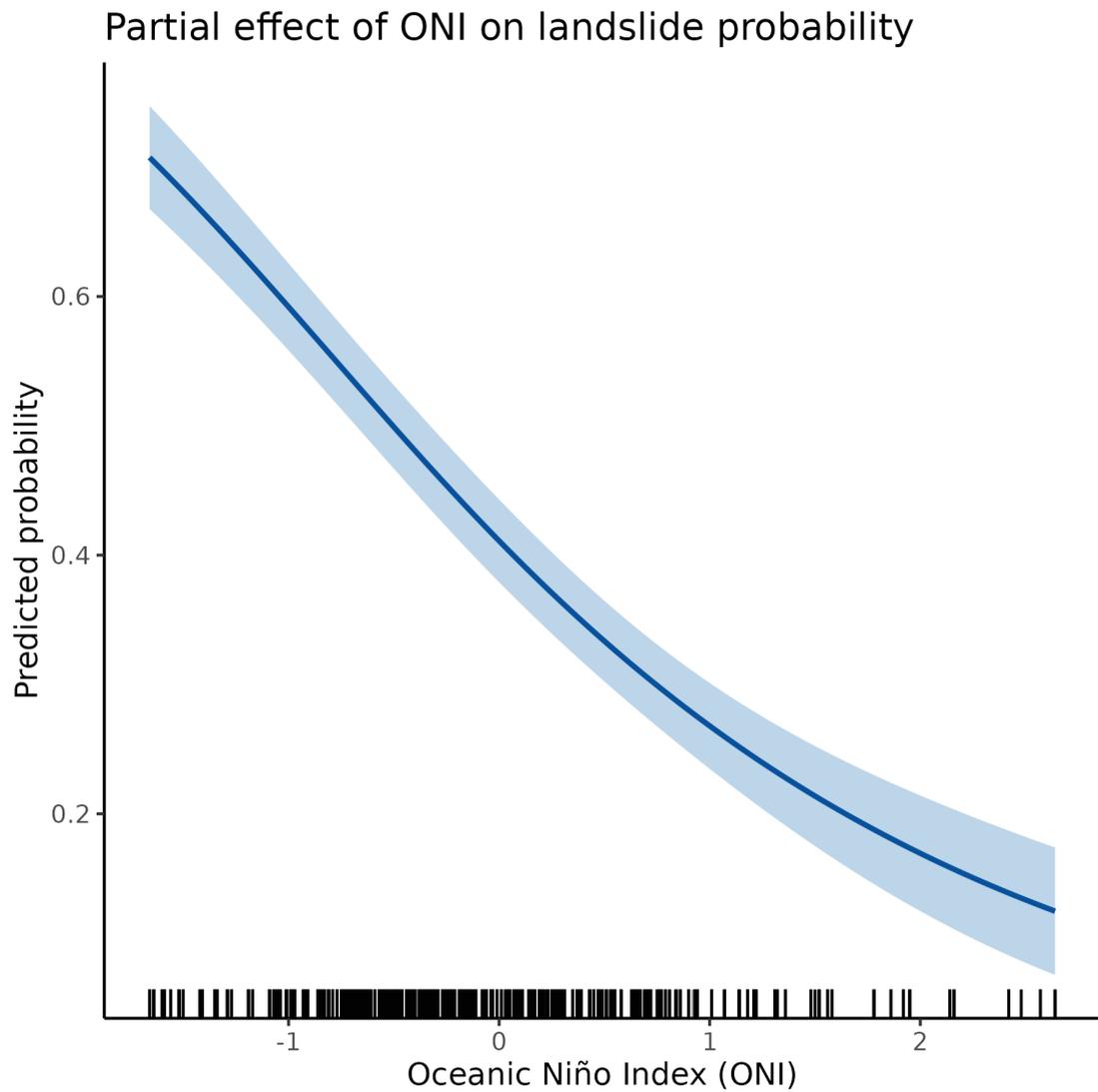Figure B.1: Precision-Recall Positive Predictive Values comparison dynamic models

Figure B.2: GAM ONI Index linear effect

B.1 Categorical contribution in GAM modeling