IPG2025-XXXX

INTEGRATING LOGISTIC REGRESSION AND AUTOMATION TOOLS FOR RAINFALL-INDUCED LANDSLIDE RISK MANAGEMENT IN PIPELINE NETWORKS

Pedro Victor Serra Mascarenhas
Transpetro
São Sebastião, Brazil

Hudson Régis de Oliveira
Transpetro
Rio de Janeiro, Brazil

Thiago da Costa Santos
Transpetro
Guararema, Brazil

Douglas de Andrade Neve
Transpetro
Rio de Janeiro, Brazil

Wanderley Camargo Russo
Transpetro
Rio de Janeiro, Brazil

Rodrigo Alves Spagnolo
Transpetro
Rio de Janeiro, Brazil

Célio Roberto Campos Piedade Júnior
Transpetro
Guaramirim, Brazil

Fernando da Silva Oliveira
Transpetro
Rio de Janeiro, Brazil

ABSTRACT

Pipelines right-of-way (ROW) traversing geotechnically sensitive terrains are particularly vulnerable to rainfall-induced landslides and debris flows. In Brazil, the Serra do Mar mountainous region presents significant geohazard challenges for pipeline integrity management. To address this, Transpetro implemented a rainfall monitoring protocol using empirical rainfall thresholds to inform operational decisions. However, recent operational demands and the expansion of the monitored network highlighted limitations in the existing protocol — particularly regarding its conservative nature and static decision boundaries. This study presents the development and application of a logistic regression model to enhance the rainfall-based decision-making protocol for pipeline response actions. The model replaces threshold curves with probabilistic susceptibility zones derived from logistic regression, using accumulated rainfall data over 24 and 72 hours as predictors. By leveraging a database of rainfall events and associated geotechnical incidents, we updated the susceptibility curves to reflect probability-based zones, allowing for more realistic and dynamic responses to increasing rainfall risk. The updated protocol enables tailored responses depending on the intersecting susceptibility class of each pipeline segment and the real-time rainfall-based probability zone. This approach exemplifies how geotechnical risk management can benefit from data science tools in pipeline operations.

Keywords: Logistic regression, Rainfall thresholds, Landslide susceptibility, Pipeline geohazard management, Probabilistic risk assessment, Rainfall-induced landslides

1. INTRODUCTION

Rainfall-induced landslides and debris flows are among the most critical natural threats to pipeline infrastructure, especially in geologically sensitive and high-rainfall environments such as Brazil's Serra do Mar. These mass movement events, often triggered by intense or prolonged rainfall, can lead to severe consequences, including ruptures in pipeline systems, environmental degradation, service interruptions, and costly emergency interventions. In this context, managing the geotechnical risks associated with rainfall has become a central challenge for infrastructure operators like Transpetro, the pipeline logistics branch of PETROBRAS.

In the Serra do Mar region, where pipeline corridors frequently traverse steep slopes with high geotechnical susceptibility, rainfall-induced landslides represent a significant challenge for risk management [1] These landslides can cause not only physical damage to the pipelines and ancillary infrastructure but also pose serious environmental and social risks due to potential product leakage in ecologically sensitive areas [2]. Additionally, the combination of rugged terrain, dense vegetation, and limited access hinders rapid response and inspection activities, increasing the importance of reliable predictive models for proactive decision-making. Managing these risks requires tools that can anticipate hazardous conditions with precision, allowing for the timely implementation of preventive measures while minimizing disruptions to operations.

To mitigate these risks, Transpetro has historically employed a rainfall threshold-based protocol to trigger field inspections and operational responses. These thresholds were

based on accumulated rainfall over short durations (24 and 72 hours) and defined conservative action limits [3]. This protocol combines a susceptibility map along with rainfall threshold curves to pinpoint where risk control actions are to be taken. While effective in avoiding incidents, this deterministic and static approach often led to unnecessary field mobilizations, excessive resource allocation, and even temporary pipeline shutdowns in the absence of actual slope instability. The limitations of such rigid frameworks become especially pronounced in large-scale pipeline networks where regional variability in rainfall response and slope susceptibility demand more flexible and location-specific decision-making tools.

In response to this need, the present work introduces a probabilistic and data-driven methodology to refine and improve the existing rainfall-based risk management protocol. By applying logistic regression models to a comprehensive historical database of rainfall and landslide events along pipeline corridors, the project aims to derive statistically robust rainfall thresholds that quantify the probability of slope failure as a continuous function of rainfall accumulation [4]. Other modifications based on experience of the application of the previous model were also implemented. The model outputs are used to define zones of landslide susceptibility, which are then integrated into a decision-support system guiding operational actions in real time.

This approach builds upon and contributes to a growing body of literature on rainfall thresholds and probabilistic modeling of landslides [5, 6, 7]. It aligns with international recommendations for quantitative landslide risk assessment [8] and leverages recent advances in geospatial data management and machine learning integration in geotechnical applications [9, 10].

The remainder of this paper presents the methodology adopted, including data processing, model construction and validation, and implementation in an operational setting. Results are discussed in terms of model performance, improvements in operational efficiency, and implications for broader risk.

1. Methodology

The methodological framework consisted of five main stages designed to address the operational need for improved, location-specific predictions of rainfall-induced landslide susceptibility along pipeline corridors: data collection, data preprocessing, model construction, model validation, and operational integration.

1.1 Data Collection and Processing

Rainfall and landslide data were collected from Transpetro's geotechnical monitoring program in three major operational regions: Southern Region, Rio de Janeiro/Minas Gerais (RJMG), and São Paulo. FIGURE 1 through FIGURE 3 show the ROWs for which the study was performed.
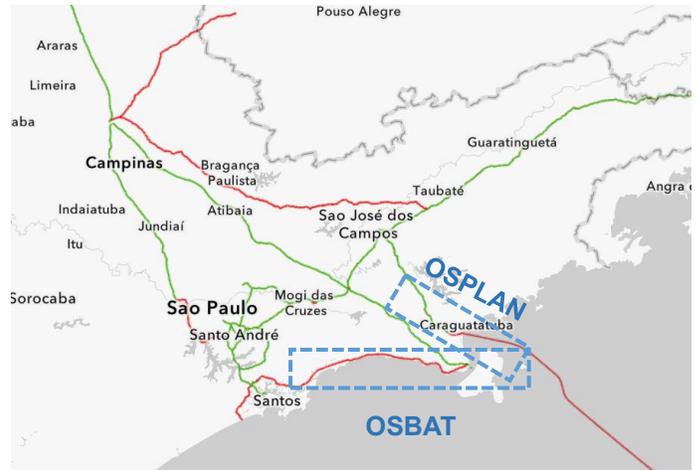

FIGURE 1 – OSBAT AND OSPLAN ROWS IN SÃO PAULO.

FIGURE 2 - RJMG ROWS.

FIGURE 3 – SOUTH ROWS.

Separate models were calibrated for each operational region to reflect local geomorphological, climatic, and operational characteristics. Four diagrams were plotted for these regions: two for São Paulo (OSBAT and OSPLAN ROWs, one for the southern region and one for Rio de Janeiro/Minas Gerais region. This division was established considering geological/geotechnical similarities or differences ROWs in each operational region, and data quality.

Events were classified as landslide occurrences (1) or non-occurrences (0). Rainfall parameters included accumulated precipitation over 24 hours and 72 hours. Only georeferenced and time-referenced events were considered.

Separate models were calibrated for each operational region to reflect local geomorphological, climatic, and operational characteristics.

2.2 Logistic Regression Model

A binary logistic regression model was used to estimate the probability of landslide occurrence as a function of the antecedent accumulated rainfall in 24h and 72h, namely A_{24h} and A_{72h}. The model has the following form:

$$p = \frac{1}{1 + \exp\left[-(\beta_0 + \beta_1 A_{72h} + \beta_2 A_{24h})\right]} \qquad (1)$$

where:
- P(Landslide) is the probability of a landslide event;
- β_0 is the intercept;
- β_1 and β_2 are coefficients for A_{24h} and A_{72h}, respectively.

The binary logistic regression implies that A_{24h} and A_{72h} are related through the usual empirical power law relation usually adopted when studying rainfall thresholds (Safeland, 2012):

$$A_{24h} = \alpha \, A_{72h}^{-\gamma} \qquad (2)$$

Where α and γ are defined by Giannechinni et al. [4]:

$$\alpha = \exp\left(-\frac{\beta_0}{\beta_2}\right)\left(\frac{1-p}{p}\right)^{-\frac{1}{\beta_2}} \qquad (3)$$

and

$$\gamma = -\frac{\beta_1}{\beta_2} \qquad (4)$$

The model is fitted using the rainfall database. Fixing the fitted γ value, we choose failure probabilities p according to the operational parameters and use the resulting parallel curves as thresholds to rainfall susceptibility zones. Each of these zones define operational response actions according to the probability of failure that it poses to the slopes in the region of the pipelines. The definition of 4 landslide susceptibility zones follows the same approach as done in Mascarenhas et al. [3].

The continuous probability output from the model was divided into four zones:

- Zone 1: < 05% (Low Susceptibility)
- Zone 2: 05–15%
- Zone 3: 15–60%
- Zone 4: <= 60% (High Susceptibility)

1.2 Model Validation

Model performance was evaluated using ROC curves and the Area Under the Curve (AUC) metric. Sensitivity and specificity values were analyzed for each region. Data visualization was performed using *Wolfram Mathematica* [11]

Model performance was evaluated using Receiver Operating Characteristic (ROC) curves and Area Under the Curve (AUC) metrics, which quantify the trade-off between true positive rate (sensitivity) and false positive rate (1 - specificity). Models achieving AUC values above 0.85 were considered satisfactory for operational deployment. Additionally, the performance assessment of the models is based on the number of true positive (TP), false negative (FN), false positive (FP), and true negative (TN) events. From these, various metrics are derived, such as the probability of detection (POD), probability of false detection (POFD), and probability of false alarm (POFA). Additional indicators include the efficiency score (Ef), the Hanssen and Kuipers score, the Euclidean distance to the perfect classifier (δ).

2.5 Operational Integration

The logistic model was incorporated into an automated data acquisition and decision-support system. Field reports are expected to be submitted via Microsoft Forms, feeding a SharePoint-based database that updates weekly rainfall-likelihood curves in Power BI dashboards.

A daily bulletin is sent via e-mail through automation using Microsoft Power Automate to engineers, managers and other workers involved with the response actions' process. The corresponding actions to each rainfall level has not been changed, and they fully described in Mascarenhas et al. [3].

2. Results and Discussion

The logistic regression models developed for the three operational regions demonstrated strong predictive capacity and practical utility. Each model achieved an Area Under the ROC Curve (AUC) greater than 0.85, confirming their effectiveness in distinguishing between rainfall events that did and did not result in landslides. Table 1 summarizes the performance metrics.

TABLE 1: AUC OF LOGISTIC REGRESSION MODELS BY REGION.

Region	AUC
Southern Brazil	0.86
Rio de Janeiro/MG	0.96
São Paulo (OSBAT ROW)	0.95
São Paulo (OSPLAN ROW)	0.97

More performance metrics are shown in TABLE 1 through TABLE 4

TABLE 1 – GENERAL PERFORMANCE INDICATORS FOR OSBAT ROW.

Probability	TP	FN	FP	TN	POD	POFD	POFA	Ef	HK	δ
5%	9	0	23	131	1.	0.149351	0.71875	0.858896	**0.850649**	1
10%	7	2	14	140	0.777778	0.0909091	0.666667	0.90184	0.686869	**0.149351**
15%	6	3	13	141	0.666667	0.0844156	0.684211	0.90184	0.582251	0.240098
20%	6	3	11	143	0.666667	0.0714286	0.647059	0.91411	0.595238	0.343856
25%	6	3	7	147	0.666667	0.0454545	0.538462	0.93865	0.621212	0.340901
30%	5	4	6	148	0.555556	0.038961	0.545455	0.93865	0.516595	0.336418
35%	5	4	4	150	0.555556	0.025974	0.444444	0.95092	0.529582	0.446149
40%	4	5	2	152	0.444444	0.012987	0.333333	**0.957055**	0.431457	0.445203
45%	3	6	2	152	0.333333	0.012987	0.4	0.95092	0.320346	0.555707
50%	2	7	2	152	0.222222	0.012987	0.5	0.944785	0.209235	0.666793
55%	2	7	1	153	0.222222	0.00649351	0.333333	0.95092	0.215729	0.777886
60%	2	7	0	154	0.222222	0.	0.	**0.957055**	0.222222	0.777805
65%	2	7	0	154	0.222222	0.	0.	**0.957055**	0.222222	0.777778
70%	2	7	0	154	0.222222	0.	0.	**0.957055**	0.222222	0.777778
75%	2	7	0	154	0.222222	0.	0.	**0.957055**	0.222222	0.777778
80%	2	7	0	154	0.222222	0.	0.	**0.957055**	0.222222	0.777778
85%	1	8	0	154	0.111111	0.	0.	0.95092	0.111111	0.777778
90%	1	8	0	154	0.111111	0.	0.	0.95092	0.111111	0.888889
95%	1	8	0	154	0.111111	0.	0.	0.95092	0.111111	0.888889

TABLE 2 – GENERAL PERFORMANCE INDICATORS FOR OSPLAN ROW.

Probability	TP	FN	FP	TN	POD	POFD	POFA	Ef	HK	δ
5%	7	0	12	150	1.	0.0740741	0.631579	0.928994	0.925926	1
10%	7	0	9	153	1.	0.0555556	0.5625	0.946746	**0.944444**	0.0740741
15%	6	1	8	154	0.857143	0.0493827	0.571429	0.946746	0.80776	**0.0555556**
20%	6	1	8	154	0.857143	0.0493827	0.571429	0.946746	0.80776	0.151152
25%	5	2	8	154	0.714286	0.0493827	0.615385	0.940828	0.664903	0.151152
30%	5	2	6	156	0.714286	0.037037	0.545455	0.952663	0.677249	0.289951
35%	5	2	5	157	0.714286	0.0308642	0.5	**0.95858**	0.683422	0.288105
40%	3	4	5	157	0.428571	0.0308642	0.625	0.946746	0.397707	0.287376
45%	2	5	4	158	0.285714	0.0246914	0.666667	0.946746	0.261023	0.572261
50%	1	6	2	160	0.142857	0.0123457	0.666667	0.952663	0.130511	0.714712
55%	0	7	2	160	0.	0.0123457	1.	0.946746	−0.0123457	0.857232
60%	0	7	1	161	0.	0.00617284	1.	0.952663	−0.00617284	1.00008
65%	0	7	0	162	0.	0.	0	**0.95858**	0.	1.00002
70%	0	7	0	162	0.	0.	0	**0.95858**	0.	1.
75%	0	7	0	162	0.	0.	0	**0.95858**	0.	1.
80%	0	7	0	162	0.	0.	0	**0.95858**	0.	1.
85%	0	7	0	162	0.	0.	0	**0.95858**	0.	1.
90%	0	7	0	162	0.	0.	0	**0.95858**	0.	1.
95%	0	7	0	162	0.	0.	0	**0.95858**	0.	1.

TABLE 3 – GENERAL PERFORMANCE INDICATORS FOR RJMG REGION ROWS.

Probability	TP	FN	FP	TN	POD	POFD	POFA	Ef	HK	δ
5%	19	2	46	371	**0.904762**	0.110312	0.707692	0.890411	**0.79445**	1
10%	17	4	25	392	0.809524	0.059952	0.595238	0.93379	0.749572	**0.145736**
15%	16	5	17	400	0.761905	0.0407674	0.515152	0.949772	0.721137	0.199688
20%	14	7	12	405	0.666667	0.028777	0.461538	0.956621	0.63789	0.24156
25%	13	8	10	407	0.619048	0.0239808	0.434783	0.958904	0.595067	0.334573
30%	10	11	9	408	0.47619	0.0215827	0.473684	0.954338	0.454608	0.381706
35%	10	11	8	409	0.47619	0.0191847	0.444444	0.956621	0.457006	0.524254
40%	9	12	7	410	0.428571	0.0167866	0.4375	0.956621	0.411785	0.524161
45%	9	12	6	411	0.428571	0.0143885	0.4	0.958904	0.414183	0.571675
50%	8	13	6	411	0.380952	0.0143885	0.428571	0.956621	0.366564	0.57161
55%	8	13	4	413	0.380952	0.00959233	0.333333	**0.961187**	0.37136	0.619215
60%	7	14	3	414	0.333333	0.00719424	0.3	**0.961187**	0.326139	0.619122
65%	6	15	3	414	0.285714	0.00719424	0.333333	0.958904	0.27852	0.666705
70%	5	16	1	416	0.238095	0.00239808	0.166667	**0.961187**	0.235697	0.714322
75%	5	16	1	416	0.238095	0.00239808	0.166667	**0.961187**	0.235697	0.761909
80%	4	17	0	417	0.190476	0.	0.	**0.961187**	0.190476	0.761909
85%	3	18	0	417	0.142857	0.	0.	0.958904	0.142857	0.809524
90%	2	19	0	417	0.0952381	0.	0.	0.956621	0.0952381	0.857143
95%	0	21	0	417	0.	0.	0.	0.952055	0.	0.904762

TABLE 4 – GENERAL PERFORMANCE INDICATORS FOR SOUTH REGIO ROWS.

Probability	TP	FN	FP	TN	POD	POFD	POFA	Ef	HK	δ
5%	10	0	31	35	1.	0.469697	0.756098	0.592105	0.530303	1
10%	9	1	21	45	0.9	0.318182	0.7	0.710526	**0.581818**	0.469697
15%	7	3	13	53	0.7	0.19697	0.65	0.789474	0.50303	**0.333526**
20%	6	4	10	56	0.6	0.151515	0.625	0.815789	0.448485	0.358883
25%	4	6	7	59	0.4	0.106061	0.636364	0.828947	0.293939	0.427735
30%	4	6	5	61	0.4	0.0757576	0.555556	0.855263	0.324242	0.609302
35%	4	6	4	62	0.4	0.0606061	0.5	0.868421	0.339394	0.604764
40%	4	6	2	64	0.4	0.030303	0.333333	0.894737	0.369697	0.603053
45%	4	6	2	64	0.4	0.030303	0.333333	0.894737	0.369697	0.600765
50%	4	6	2	64	0.4	0.030303	0.333333	0.894737	0.369697	0.600765
55%	4	6	0	66	0.4	0.	0.	**0.921053**	0.4	0.600765
60%	3	7	0	66	0.3	0.	0.	0.907895	0.3	0.6
65%	3	7	0	66	0.3	0.	0.	0.907895	0.3	0.7
70%	3	7	0	66	0.3	0.	0.	0.907895	0.3	0.7
75%	1	9	0	66	0.1	0.	0.	0.881579	0.1	0.7
80%	1	9	0	66	0.1	0.	0.	0.881579	0.1	0.9
85%	0	10	0	66	0.	0.	Indeterminate	0.868421	0.	0.9
90%	0	10	0	66	0.	0.	Indeterminate	0.868421	0.	1.
95%	0	10	0	66	0.	0.	Indeterminate	0.868421	0.	1.

The models produced smooth and interpretable probability surfaces that align well with expert knowledge of regional rainfall patterns and slope behavior. FIGURE 4 through FIGURE 7 presents the probabilistic curves, showing the gradient of landslide probability. Additionally, FIGURE 8 through FIGURE 11 show the ROC curve for each fit.
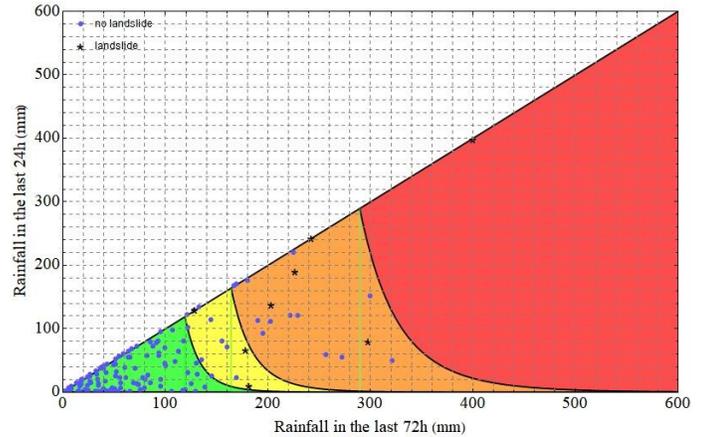

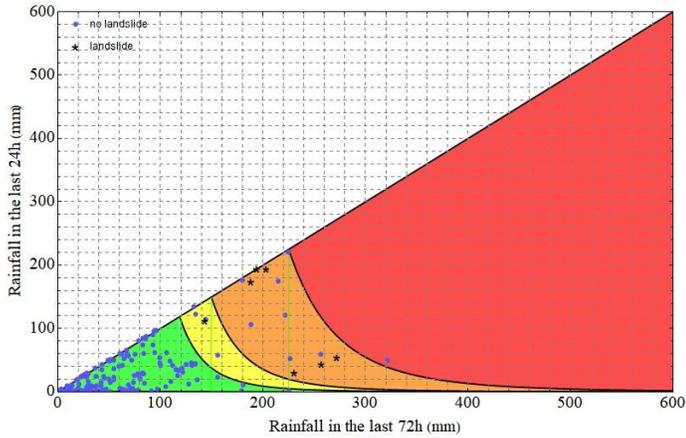FIGURE 4 – LOGISTIC REGRESSION DECISION CURVES FOR OSBAT
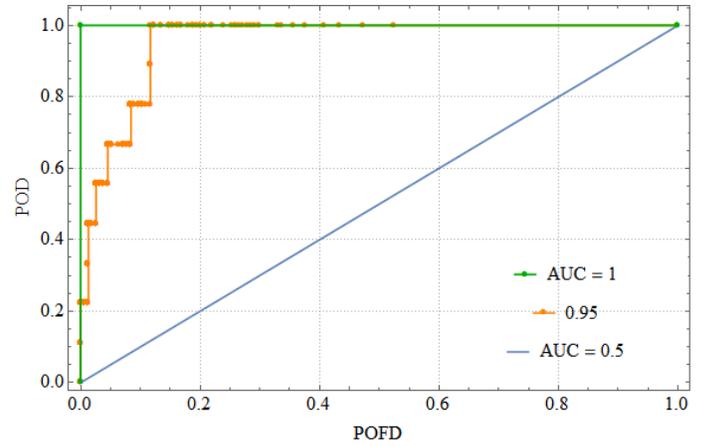
FIGURE 5 – LOGISTIC REGRESSION DECISION CURVES FOR OSPLAN
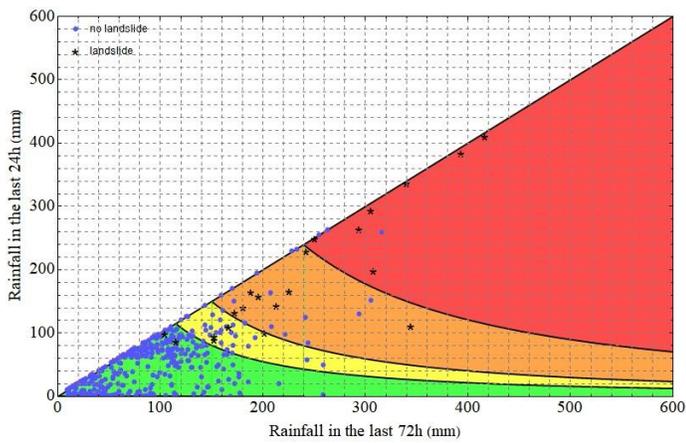

FIGURE 8 – LOGISTIC REGRESSION ROC CURVE FOR OSBAT

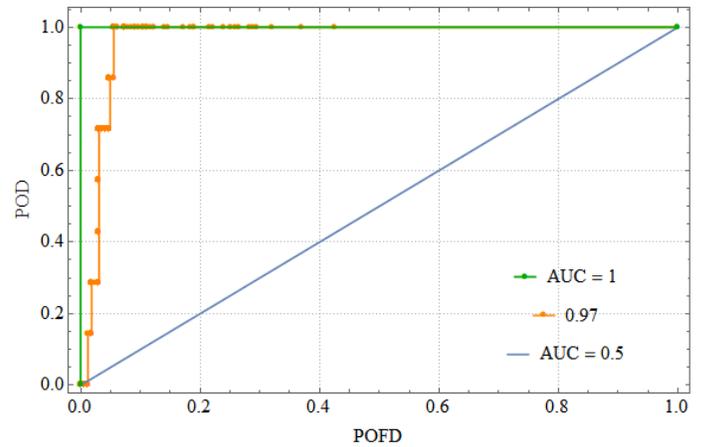FIGURE 6 – LOGISTIC REGRESSION DECISION CURVES FOR RJMG REGION
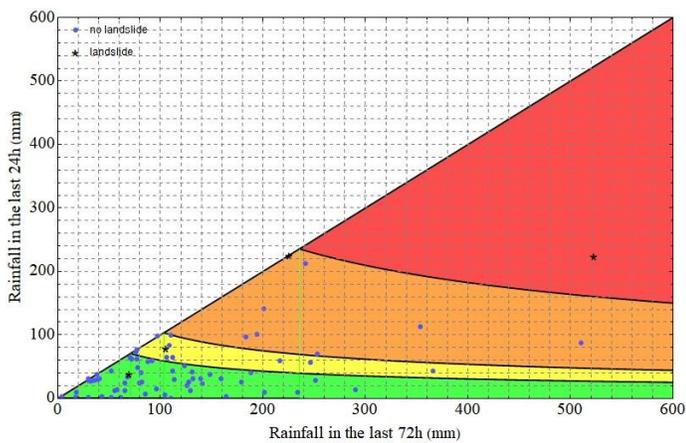

FIGURE 9 - – LOGISTIC REGRESSION ROC CURVE FOR OSPLAN

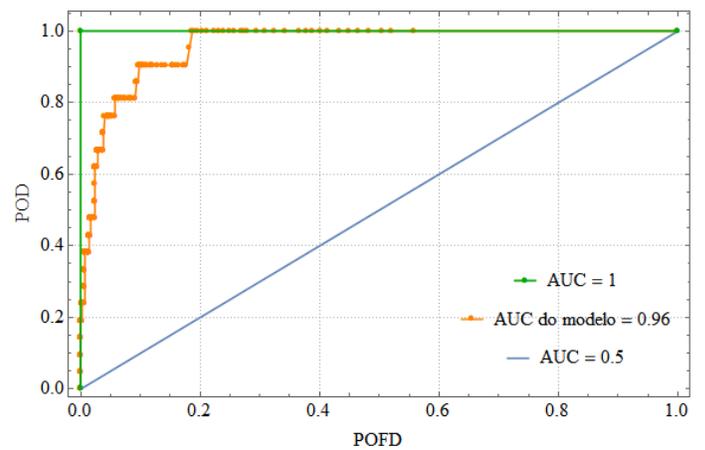FIGURE 7 – LOGISTIC REGRESSION DECISION CURVES FORFOR SOUTH REGION ROWS

FIGURE 10 -– LOGISTIC REGRESSION ROC CURVE FOR RJMG REGION ROWS
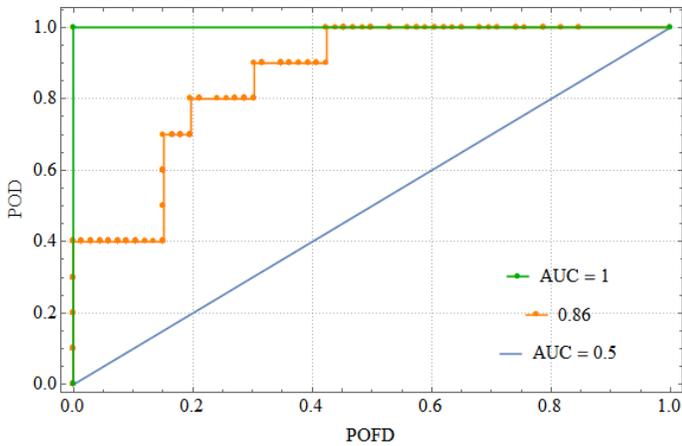
5

© Courtesy of ASME

FIGURE 11 – LOGISTIC REGRESSION ROC CURVE FOR SOUTH REGION ROWS

The São Paulo (OSPLAN) ROW showed the highest model performance. In contrast, the Southern region exhibited slightly lower predictive accuracy. This could be due to the quality of the data set available for fitting the logistic regression.

From TABLE 1 (OSBAT ROW), The best balance between Probability of Detection (POD) and Probability of False Detection (POFD) occurs at the 5% threshold, where POD is maximum (1.0) and POFD is moderate (0.149). This leads to the highest Hanssen & Kuipers score (HK = 0.85) and the smallest Euclidean distance (δ = 0.149), indicating optimal model performance at this point. At 10%, the model maintains a strong POD (0.777778) with slightly improved POFD (0.09901), but HK decreases (0.685869), and the δ distance increases. From 15% to 35%, POD steadily declines (from 0.666667 to 0.444444), and although POFD also improves, the Hanssen & Kuipers score falls significantly. For instance, at 35%, POD is 0.444444, POFD is 0.019802, but HK drops to 0.4246419, reflecting the diminishing utility of the classifier at higher thresholds. Starting from the 45% threshold, POD drops to 0.333333 and continues decreasing. POFD remains low (as low as 0.00649351), but HK becomes close to or less than 0.2, indicating increased missed detections (FN). From 65% onwards, the model becomes too conservative: POD stabilizes at 0.222222 or lower, and although POFD becomes 0.0, the classifier fails to capture many real events.

From TABLE 2 (OSPLAN), At the 10% threshold, the classifier achieves its best overall balance. This threshold represents the optimal operating point, balancing high sensitivity with low false detection. At the 5% threshold, although the model still detects all true positives (POD = 1.0), the false positive rate increases slightly (POFD = 0.074074), resulting in a slightly lower HK = 0.925926. Between 15% and 25%, the model maintains high detection rates (POD = 0.8571) and moderate POFD values (ranging from 0.03 to 0.07), indicating the model still performs well, but with gradually declining effectiveness. Beyond the 30% threshold, POD begins to drop noticeably, and HK also declines — reflecting a growing number of missed events. At thresholds above 55%, POD approaches zero, and POFD also drops to zero, which indicates that the model makes almost no positive predictions, and it fails to detect true events (TP = 0 from 55% onward). Efficiency (Ef) remains above 0.94 in almost all thresholds, but this metric alone can be misleading since it is affected by the large number of TNs and does not fully capture the cost of missed landslide detections. The efficiency indicator (Ef) remains consistently high (above 0.90) across thresholds between 25% and 80%, showing that the model maintains strong general performance even with reduced sensitivity.

From TABLE 3, the 5% threshold yields the highest POD (0.904762) and the highest Hanssen & Kuipers score (HK = 0.79445). Although the POFD is 0.110312, the model maintains the best balance between detecting true events and avoiding false alarms. Additionally, this threshold also has the lowest δ value (0.145736), confirming its proximity to the behavior of a perfect classifier. As the threshold increases, POD steadily decreases (from 0.81 at 10% down to 0.09 at 90%), POFD improves (approaching 0), but this comes at the cost of higher false negatives (undetected landslide events). Furthermore, HK drops sharply beyond 15–20%, reaching zero at 95%, indicating loss of discriminative power as the threshold increases.

From TABLE 4, at 5% the model achieves the maximum POD (1.0), but at the cost of a high POFD (0.47) and moderate HK (0.53). δ = 1.0 confirms it is distant from optimal balance. The optimal threshold in this configuration appears to be 10%. Thresholds between 15–25% maintain moderate POD (0.6–0.7) with improved POFD (0.10–0.20), but HK is slightly lower than at 10%. For intermediate thresholds (30–50%), the balance begins to shift: POD stabilizes at 0.4 from 30% to 50%, indicating increasing false negatives; the POFD improves continuously, dropping to just 0.03 at 50%; the HK score gradually rises to 0.366997 at 50%, but it's still below the peak observed at 10%; and Ef reaches its maximum at 50% (0.894737), yet this is largely influenced by the high number of TNs and does not reflect sensitivity performance. From 65% upward, the classifier effectively stops identifying any true positives. At this range of probability, TP = 0, POD = 0, HK = 0, Ef remains high (~0.86–0.88) only because of the overwhelming TNs. POFD and POFA = 0, but this is because no alerts are issued at all, including for true events. Results are labeled "Indeterminate" for Ef and HK in this region, reinforcing that these thresholds are not viable for operational use.

3. Conclusions

This work demonstrates that logistic regression models can effectively improve rainfall-induced landslide risk management for pipeline systems. By transitioning from rigid empirical thresholds to probability-based zones, the system provides more accurate and operationally efficient decisions.

This methodology ensures that rainfall risk is continuously reassessed, updated, and translated into actionable information with minimal latency, enhancing the overall resilience and efficiency of pipeline operations in landslide-prone areas.

Across all operational regions and pipeline segments analyzed, the logistic classifiers exhibited a consistent and

interpretable performance pattern. Classifiers demonstrated their best performance at low probability thresholds, particularly around 10%, where sensitivity was maximized, false detection rates remained within acceptable limits, and the balance between detection and specificity was optimal, as indicated by the highest Hanssen & Kuipers (HK) scores and the lowest Euclidean distances to the perfect classifier.

As probability thresholds increased beyond this range, classifier performance degraded progressively. Although false alarms decreased, so did the ability to detect true landslide-triggering events, compromising operational safety. At high thresholds (above 60%), classifiers ceased to function effectively, failing to identify any relevant events while maintaining artificially high efficiency due to the prevalence of true negatives

These findings highlight the need to avoid overly conservative thresholds in operational settings. In practice, this is accomplished by adopting 4 different threshold curves with different response actions.

The results confirm that logistic regression is a robust and valuable methodology for defining dynamic, probabilistic rainfall thresholds in support of risk-based decision-making. The consistent identification of 10% as the optimal threshold further supports its adoption within inspection and emergency protocols to enhance the resilience and reliability of pipeline infrastructure exposed to intense rainfall.

While the logistic models have shown strong performance, several limitations must be acknowledged:

- The models depend on the quality and completeness of reported landslide data, which may underrepresent smaller or undocumented events.
- Terrain-specific variables (e.g., slope angle, land use, lithology) are not directly included in the logistic function and are instead addressed indirectly via static susceptibility maps.
- Threshold sensitivity must be recalibrated periodically, especially in the context of climate change trends and land use modifications that alter hydrological responses.

Further enhancements could include ensemble learning methods, dynamic rainfall-triggering indicators derived from remote sensing, or integration with soil moisture sensor data to refine near-real-time predictions.

In summary, the logistic regression models provided a practical and scientifically sound means of improving landslide risk decision-making for pipeline corridors under variable rainfall conditions. Their probabilistic nature allows for more nuanced and proportionate operational responses, ultimately contributing to safer and more resilient infrastructure management.

Future work may include integrating soil moisture sensors, machine learning ensembles, or satellite data to further refine predictions.

4. References

[1] Amarasinghe, M.P., Kulathilaka, S.A.S., Robert, D.J., Zhou, A., and Jayathissa, H.A.G., "Risk Assessment and Management of Rainfall-Induced Landslides in Tropical Regions: A Review," *Natural Hazards*, Vol. 120, No. 3, pp. 2179–2231, 2024.

[2] Cunha, S.B., "A Review of Quantitative Risk Assessment of Onshore Pipelines," *Journal of Loss Prevention in the Process Industries*, Vol. 44, pp. 282–298, 2019.

[3] Mascarenhas, P.V.S., Santos, T.C., Peixoto, V.C., and Guimarães Neto, J.D., "A Statistical Model for Pipeline Rainfall Triggered Landslides Risk Management," *Proceedings of the Pipeline Technology Conference 2023*, Berlin, Germany, 2023.

[4] Giannecchini, R., Galanti, Y., Avanzi, G.D.A., and Barsanti, M., "Probabilistic Rainfall Thresholds for Triggering Debris Flows in a Human-Modified Landscape," *Geomorphology*, Vol. 257, pp. 94–107, 2016.

[5] Caine, N., "The Rainfall Intensity-Duration Control of Shallow Landslides and Debris Flows," *Geografiska Annaler*, Vol. 62, No. 1, pp. 23–27, 1980.

[6] Guzzetti, F., et al., "Rainfall and Landslide Initiation," in *Rainfall: Modeling, Measurement and Applications*, Elsevier, pp. 427–444, 2022.

[7] Giannecchini, R., et al., "An Update on Rainfall Thresholds for Rainfall-Induced Landslides Using Different Statistical Methods," *Water*, Vol. 16, No. 5, Article 624, 2024.

[8] Corominas, J., et al., "Recommendations for the Quantitative Analysis of Landslide Risk," *Bulletin of Engineering Geology and the Environment*, Vol. 73, pp. 209–263, 2014.

[9] Pourghasemi, H.R., et al., "Prediction of the Landslide Susceptibility: Which Algorithm, Which Precision?," *Catena*, Vol. 162, pp. 177–192, 2020.

[10] Alcântara, E., et al., "Machine Learning Approaches for Mapping and Predicting Landslide-Prone Areas in São Sebastião (Southeast Brazil)," *Natural Hazards Research*, in press, 2024.

[11] Wolfram Research, Inc., *Mathematica, Version 11.0* [Computer software], Wolfram Research, Inc., Champaign, IL, 2016.