Predictability of rainfall induced-landslides: The case study of Western Himalayan Region

Swadhi Ritumbara Das¹, Poulomi Ganguli¹,*

¹Agricultural and Food engineering department, Indian Institute of Technology Kharagpur, Kharagpur India

*Correspondence to: Poulomi Ganguli (pganguli@agfe.iitkgp.ac.in)

Abstract

Landslides are one of the natural hazards that are most prominent in tectonically active regions, such as mountainous terrains of the Himalayas. The Himalayan region is vulnerable to landslides due to its fragile lithology, steep slopes, geology, rainfall patterns, and high topographical roughness. Among several factors responsible for slope instability, rainfall is one of the significant drivers that cause a maximum number of landslides. However, very few studies have explored precipitation-induced landslide susceptibility of the Himalayan region due to observational constraints. This study attempts to fill the gaps in the literature by developing a power-law relationship between the maximum rainfall intensity that potentially triggers landslides and the corresponding event duration of selected ‘hotspot’ locations across the Western Himalayan Region (hereafter WHR) that are highly susceptible to landslides. We identified more than 500 landslide events between 2007 and 2016 based on the landslide inventory database, which suggests more than 70% of landslide events are clustered during the southwest monsoon season. Further, we show an increase in rainfall in recent decades (2007-16) over low elevated areas of the WHR compared to the long-term climatology (1988-2006), revealing intensification of rain events, which could amplify landslide occurrence. Our observational assessment suggests around 28% of landslides events are followed by within a week of occurrence of triggering rain events. The regionalization of the maximum intensity of triggering rain events versus the corresponding event duration shows that the tail of the triggering rain events tends to follow a power-law relationship with a robust positive exponent of more than one, suggesting synchronicity between two variables. The derived insights would aid in the predictability of shallow-to-deep rain-induced landslide events and inform climate adaptations in steep-slope areas.
1 Introduction

Landslides are typical hydro-geological cascading hazard, which often results from incessant rainfall followed by typical mass movement processes, such as slides and debris flow, triggering several tertiary disasters (Nguyen et al., 2013; de Ruiter et al., 2020). In India, out of the 0.42 million km$^2$ (12.6%) that are prone to landslide, 0.32 million km$^2$ falls in the Himalayan range (www.gsi.gov.in). Climate change is expected to alter regional patterns of weather systems, shifting the High Mountain Asian water cycle, which has a consequence on the stability of natural and engineered slopes that trigger shallow-to-deep landslides (Gariano and Guzzetti, 2016; Masson-Delmotte et al., 2021). Rainfall is one of the primary causal drivers’ contributing to landslides. The sequence of landslide triggering rainstorms increases the pore water pressure and volumetric water content changes due to changes in mechanical, hydraulic, and physical properties of soil and vegetation characteristics. Furthermore, due to climate change, many mountainous regions have expected more precipitation as rain rather than snow, which increases the likelihood of slope instabilities, especially in densely populated localities (Davenport et al., 2020). A typical landslide event leads to immense loss of life and property. Between 2007 and 2016, the Western Himalayan region experienced 547 landslides, out of which 508 landslide events have caused due to heavy rainfall, tropical cyclone, and snowfall-snowmelt (Kirschbaum et al., 2015). Moreover, slope failure could trigger or exacerbate land degradation owing to soil erosions, resulting in permanent damage to agricultural land (Tarolli et al., 2021). Therefore, investigating such consecutive disasters, i.e., precipitation-induced landslides/mass movement, is crucial in unexplored areas of the globe. Mapping spatiotemporal patterns of rain-induced landslides would facilitate sustainable land management over complex terrains ensuring food security and ecosystem resilience (Tarolli et al., 2021).

A large body of the literature investigated rainfall thresholds for shallow to deep landslides. Manual fitting (Giannecchini et al., 2012), probabilistic approaches (Gariano and Guzzetti, 2016; Abraham et al., 2020; Berti et al., 2012), and empirical methods (Dahal and Hasegawa, 2008; Kanungo and Sharma, 2014) are some of the approaches used by various studies to determine ID threshold values. Few studies (Abraham et al., 2020, 2019; Guzzetti et al., 2007; Kanungo and Sharma, 2014; Mathew et al., 2014) have shown the importance of antecedent rainfall in contributing to slope instability. Globally,
several studies have reported rain-induced shallow and deep landslides and derived the ID threshold curve; however, generalization may not be possible for all climate types, for instance, mild temperate regions of western Himalaya. A few studies lack standardized database to archive landslide information and relate it with rainfall and geographic data (Brunetti et al., 2010). Also few studies disregard spatial heterogeneity and topographical variability (Berti et al., 2012) as multiple rain gauge data were merged to one in homogenous regions. The spatiotemporal variability in rainfall thresholds for landslide occurrences of Indian Himalayas are not fully explored. Very few observational assessments are available for the Western Himalayan Region (WHR) analysing rainfall threshold curves, and they are either based on a limited number of gauge-based observations (Dahal and Hasegawa, 2008; Kanungo and Sharma, 2014) or satellite-based products (Martha et al., 2021). An update of the rainfall ID threshold considering recent landslide inventory is of utmost importance for building climate-resilient infrastructure and vulnerability mapping, aiding early warning systems. Second, although several studies focused on landslides triggered by heavy precipitation, no studies have shown space-time changes in precipitation patterns, and investigated its potential linkage with landslide occurrences over the WHR.

To fill the gaps in the literature, we investigated the following research questions: (i) Is there any change in regional rainfall pattern in recent (2007-2016) versus historical (1988-2006) time window that may trigger the landslide occurrences? (ii) Can we develop a mathematical relationship between triggering rain intensity and rainfall duration for selected landslide-prone sites across the WHR? (iii) Can we regionalize the ID threshold relationships based on regional rainfall distribution? We selected six meteorological stations across the WHR (Figure 1), prone to landslide events (Ram et al., 2020; shah et al., 2022; Pathak, 2016; Devi and Patra, 2015; The Indian Express, 2021; Daily Excelsior, 2022). Leveraging the station-based observational records, we derive the power-law relationship between maximum rainfall intensity and corresponding duration that potentially trigger the shallow-to-deep landslides. While we generate at-site ID threshold curves from individual station-based observations, we establish regional ID threshold relationship by grouping rain-gauge stations based on similar spatial distributions of rainfall regions and geospatial properties (see Methods: section 2.2.4 and detailed workflow in Figure 2). Our preliminary analyses show an increased frequency of landslides events within two to three days of heavy rainfall across selected sites. For regionalization, we apply the Principal
Component Analysis (PCA) coupled with Fuzzy c-means (FCM) clustering (Gautam, 2006; Mosavi et al., 2021; Sadri and Burn, 2011). Despite a sparse observational network (e.g., Joshimath and Katra, see table 1), we establish a power-law relationship by reconstructing the missing gaps in the daily rainfall time series through Regularized Expectation Maximization method (RegEM; Schneider, 2001). In the subsequent section, we organize the paper as follows: In section 2, we describe the data and the methods, Section 3 discusses the performance statistics for infilling gaps in the rainfall time series, section 4 explains the results obtained from the analysis; finally, section 5 presents the summary and conclusion of the study.

2 Data and Methods

2.1 Data

We selected six rain gauge stations across the WHR, namely Banihal and Katra located in Jammu and Kashmir, Mandi and Solan in Himachal Pradesh, Dehradun and Joshimath in Uttarakhand (see figure 1). First, we collect station-based daily precipitation time series from the India Meteorological Department (IMD)’s Data Supply Portal (https://dsp.imdpune.gov.in/) for the time period of 1970-2017 (See Table 1). All stations have varying lengths of records. However, few stations have sparse temporal coverage and available records vary from three to five years (see table 1). Next, we obtain gridded daily rainfall records from the IMD at a spatial resolution of 0.25° (Srivastava et al., 2009). We infill the missing observations in the station-based records using the gridded rainfall time series (See Methods). Finally we retrieved the historical landslide inventory for the WHR from the NASA Cooperative Open Online Landslide Repository (COOLR), which archived the global landslide database (Kirschbaum et al., 2015).

2.2 Methods

2.2.1 Treatment of Daily Rainfall Time Series

The gauge-based rainfall time series often contain incomplete observations. Machine learning approaches are often used in the literature (Schneider, 2001; Kim and Pachepsky, 2010) to reconstruct missing precipitation records. To infill the missing gaps in the time series, we have used a robust machine learning...
algorithm, the Regularized Expectation Maximization (RegEM), as proposed by Schneider, (2001). This is the extension of the conventional EM algorithm and is specially designed to fill missing gaps in rainfall time series. It consists of an E-step (Expectation step) that attempts to estimate the missing variables, and the M-step (Maximization step) optimizes the parameters to best fit the data. The RegEM is a machine learning tool that provides an efficient avenue to impute missing gaps in rainfall records compared to conventional regression-based approaches (Schneider, 2001; Tsidu, 2012; Kalteh and Hjorth, 2009). Instead of considering the whole year as single time series, we infill the missing gaps in the rainfall time series considering the seasonal stratifications. We consider four seasons: winter (January to March), summer (April to May), monsoon (June to September), and fall (October to December) seasons. Typically the gridded records often underestimate the frequency of extreme events (Muñoz et al., 2011; Rajeevan et al., 2021). We adopt a Statistical Post Processing (SPP) approach using the Quantile Mapping (QM; Li et al., 2010) method to address this issue. In this method, we map the distributions (Cumulative Distribution Function, CDF) of the observed time series and the infilled time series, which is designed to adjust the distributions of infilled time series such that it matches the observed climatologies of station-based rainfall time series (Li et al., 2010). We implemented a month-wise-CDF matching for the daily time series to preserve the seasonal variability for the period of analysis (1970-2019). Since several studies (Martinez-Villalobos and Neelin, 2019) showed that daily to monthly precipitation tends to follow the Gamma probability distribution, for CDF matching, we assume that non-zero monthly rainfall follows the Gamma distribution for each site. We apply the QM at a 21-year moving window, shifted by one year at each time step to account for temporal variability in daily rainfall time series (Brunetti et al., 2001; Kharin et al., 2007; Blöschl et al., 2017).

Next, we evaluate the performance of reconstructed time series versus the available observed time series. For this, we compare the observed records versus (i) the reconstructed time series without any statistical post-processing scheme implemented (hereafter referred to as SPP-WQM) (ii) the series with statistical post-processing (hereafter referred to as SPP-QM). We evaluate the performance of both series for individual sites through a series of performance statistics developed by Beck et al. (2017). To this end, we employed an integrated test score, which checks for rainfall magnitude and its temporal signatures individually, and their combinations (Beck et al., 2017). Here we use three measures of rainfall
signatures i.e., difference in Annual Average Rainfall ($\Delta AAR$), the difference in 90% exceedance probability ($\Delta Q$) of rainfall, and the difference in the timing of the centre of mass ($\Delta CT$) of the observed versus simulated rain events. We compute $\Delta CT$ using circular statistics (Stewart et al., 2005). We normalize the rainfall signatures using the standard deviation of the observed time series to provide more weightage to smaller values. The overall score of the rainfall signature is the median of the three signatures. Second, to consider the temporal coherency, we compare the nonparametric spearman’s rank correlations of observed versus simulated time series for daily ($r_{dly}$) and 5-day ($r_{5\text{-day}}$) rainfall records. We consider Spearman’s rank correlation to minimize the effect of outliers. The final performance score is the median of the rainfall and temporal signatures.

2.2.2 Identification of Landslide Triggering Rain Events

Next, we identify the triggering rain events that cause landslides. First, we consider the rain event when the minimum rainfall is 1 mm for seasons with low rainfall, such as summer, winter, and fall. To emphasize the seasonal pattern of the monsoon, we determine the rain threshold for the monsoon season following equation 1. For this, we calculate the weighted average of rain by weighing the length of the rain events of a particular season (in days) that has $>1$ mm rain magnitude:

$$\text{mean seasonal rainfall} = \frac{\text{sum}(\text{Rain sum}_{\text{seasons}} \times \text{length}_{\text{seasons}})}{\text{sum}(\text{length}_{\text{seasons}})}$$

(1)

In Eq. 1, the variable $\text{Rain sum}_{\text{seasons}}$ indicates sum of the rain events that has magnitude greater than 1 mm during a particular season, whereas $\text{length}_{\text{seasons}}$ denotes the number of events that have rain magnitude of more than 1 mm. We consider a threshold of 1% of the mean June-September rainfall for the monsoon months to account for the seasonal variability of rain events. This is because the southwest summer monsoon season is the most significant contributor to total annual rainfall in India (Soman and Kumar, 1990). To maintain consistency in seasonal rainfall thresholds across the sites, we consider the spatial average (50th percentile) of the threshold values for the monsoon season for all stations. Considering the spatial average, the rainfall threshold in monsoon is approximately 3 mm. Among six
stations, Banihal showed the highest rainfall magnitude in the winter months, unlike other stations. Therefore, we consider the rainfall threshold during winter for Banihal as 3 mm, while the rain threshold is 1 mm for the other three seasons.

2.2.3 Developing the Power-law Relation: ID Threshold Curves

To identify the triggering rain event for a specific landslide event, we use the NASA Cooperative Open Online Landslide Repository (COOLR) (Kirschbaum et al., 2015). Then we consider the landslide events within a 50 and 100 km radii of each station. As the number of events within the 50 km radius is less in count, we expand the radius to 100 km encompassing the maximum number of landslide events (Figure S1). We consider a time window of up to a month preceding the date of occurrence of landslides (Bertola et al., 2021; Kanungo and Sharma, 2014; Johnston et al., 2021). While temporal clustering of moderate to heavy rain events often amplifies the likelihood of landslides (Bevacqua et al., 2021), we consider the triggering rain event, which produces the maximum rain intensity at most 30 days before the landslide. Once we obtain the triggering rain events, we develop the at-site ID threshold curve for the six stations. We derive the at-site power-law relation by fitting the event's triggering rainfall intensity ($I_m$) and corresponding event duration ($d$) to a polynomial function, represented on a logarithmic scale (base 10).

$$I_m = \alpha d^\lambda, \quad \lambda > 0, \ d > d_{\text{min}} > 1$$

(2)

Where $\alpha$ and $\lambda$ are constants and have been derived for specific sites. The value of $\lambda > 1$ indicates that the intensity amplifies with an increase in duration, whereas $\lambda < 1$ indicates decrease in rainfall intensity with an increase in duration. The rainfall intensity of a storm event is computed using rainfall accumulated ($W$) over the duration, $d$; where $I_{\text{max}} = W/d$. The slope of the least-square fitting corresponds to $\lambda$ and the starting point of the linear fitting is $d_{\text{min}}$.

2.2.4 Regionalization of ID Threshold Relationship

Next, we identify spatial contiguity of rainfall ID threshold by applying a clustering algorithm (Sadri and Burn, 2011) over six rain gauges across the WHR based on seven attributes (see table 1): (i) latitude and
longitude of rainfall stations, (ii) elevation of the stations, (iii) average seasonal rainfall of four seasons summer, winter, monsoon, and fall. The regionalization of ID thresholds based on various attributes involves the Principal Component Analysis (PCA) followed by fuzzy c-means (FCM) clustering (Sadri and Burn, 2011). We normalize the multiple attributes to a 0 to 1 scale before performing the regionalization operation. We discovered that the first three Principal Components (PCs) together explain 95% of the variability in records (Figure 3a). Therefore, we apply the first three PCs to identify the homogenous rainfall regions. Our analyses suggests that rainfall has substantial contributions to the PCs (Figure 3b). Therefore we divide the WHR into two regions: (i) region 1: Banihal as a separate region, and (ii) region 2: remaining stations collectively considered part of region 2. Finally, for region 2, we derive the regional ID threshold by combining landslide triggering rain intensity and duration for the five stations.

3 Results and Discussions

3.1 Spatial Variation in Seasonal Rainfall across the WHR

The rainfall in India is highly seasonal. The Southwest monsoon season (June – September) is the most significant contributor to rains in most parts of India (Soman and Kumar, 1990). Further, the western disturbances (WD) are the primary source of precipitation (rain and snow) in northern India during the winter (Hunt et al., 2018). The seasonal rainfall variability across the gauged sites shows the occurrence of peak rainfall during the monsoon season for most of the locations except Banihal (Figure 4). For Banihal, the peak appears in the winter between February and March. This is because the Southwest monsoon is weak over this region (Dad et al., 2021). The whole of Kashmir valley receives heavy precipitation (comprises rain and snow) in winter due to the evolution of WDs from the Mediterranean sea (IMD, 2014; Hunt et al., 2018). The fuzzy c-means clustering algorithm suggests two distinct regions over the WHR: Region 1 comprises Banihal, which show peak rainfall during the winter season; region 2 include the remaining five stations with peak rain event during the monsoon season. The temporal evolution of monthly precipitation versus the frequency (number) of landslide events (Figure 5) suggests strong synchronicity between the two: the number of landslide events is higher for seasons with high
rainfall. In particular, the peak landslides occurred between 2010 and 2013, when the monthly rainfall often exceeded 200 mm. This is in agreement with an earlier study (Jones et al., 2013) over the Nepal Himalaya that showed a strong relationship between rain-triggered mass-wasting phenomena versus extreme precipitation using 30-year mass-wasting inventory records.

### 3.2 Clustering of Landslides during the Monsoon

Across the WHR, we identified more than 500 registered landslide events resulting from heavy rainfall, tropical cyclone-induced storm, and snowfall-snowmelt from 2007 to 2016. However, we could not identify any landslides from 1988-2006 in the region, possibly due to the lack of credible records. The propensity of landslide events were pronounced in low elevated areas of Uttarakhand (Figure 6a). Record breaking landslide events are typically clustered around 2010-2013, out of which 2013 was the most devastating in terms of fatality (Figure 6b). The June 2013 flash floods and landslides events in Kedarnath town is well documented in the literature (Martha et al., 2015; Houze et al., 2017; Singh et al., 2014; https://theworld.org/stories/2013-11-18/himalayan-tsunami-climate-change-triggers-deadly-floods-among-world-s-highest). Substantial damages and fatalities during 2013 landslides events are contributed to heavy rains in June 2013 resulting in multiple landslides followed by glacial lake outburst flood (GLOFs; Martha et al., 2015; Houze et al., 2017; Singh et al., 2014). The interaction between an impending trough of westerlies and the strong south easterly monsoon wind flow, compounded by a low pressure system established over North India, resulted in extreme rainfall in this region (Singh et al., 2014; Martha et al., 2015).

The spatial trends in rainfall over the WHR (Figure 7a) show, in the two out of three states, such as Jammu and Kashmir and Himachal Pradesh, the high average rainfall occurs in the western half, and this pattern gradually tends to decrease towards the eastern half. This is possibly due to the eastern half is a high elevated polar region, where more precipitation falls as a snow event than the rain (figure 7a-b). Unlike the other two states, the spatial trend in rainfall over Uttarakhand shows an opposite pattern, where the eastern half of the state shows high average rainfall, which gradually decreases from the east to the west. Comparative assessment of the spatial trends in rainfall pattern (Figure 7c) in the recent (2007-2016) versus the past decades (1988-2006) suggests an increase in rainfall in the recent years, especially
over the low elevated terrains that possibly led to an increase in landslide events (see Figure 7b). Our analysis of historical landslide events shows ~77% of medium to large-sized landslide events clustered around the monsoon season, whereas the remaining landslide events are distributed in other seasons. To further investigate whether the ‘local’ (i.e., at individual rain grids) changes in rainfall pattern are statistically significant, we perform the non-parametric Wilcoxon rank-sum test at 5% significance level at each grid points. The areas showing a significant increase in rainfall mostly lie in the latitudinal belts of 29º N to 30º N and 32º N to 34º N (Figure 7c). The increase in rainfall in recent decades is in agreement with a recent study (Meher and Das, 2022) over the WHR, in which, using climate model simulations, authors have suggested intensification of the southwest monsoon with an increase in percentage changes of rainfall. Moreover, the review of the literature confirms an increase in the number of wet days over the WHR with warming during the past few decades (Sabin et al., 2020).

3.3 Development of ID Threshold Curve and Power-law Relationship

First, we infilled the missing gaps in rainfall time series at each station using the EM algorithm. The RegEM method provides the reasonable performance in filling missing gaps with performance ranges from 0.95 (excellent) to 0.43 (moderate). The RegEM approach provides an excellent performance measure for gauges with fewer counts of missing observations. However, the method offers a modest performance for sites with sparse temporal coverage. The heat map comparison of the time series, SPP-WQM versus SPP-QM, shows (Figure 8) little to no improvement in applying the SPP-QM method against observations across the selected time windows at each station. This could be primarily due to the sparse temporal coverage of the rain gauges, which may have affected the performance of the statistical post-processing technique. The other possible issues could due to the parameterization of the Gamma distribution, which may not be able to preserve the moment properties of observed time series credibly (Walter and Piet-Lahanier, 1990). Therefore, for further analyses, we consider the SPP-WQM time series.

We identify the triggering rain events with maximum rain intensity at most 30 days preceding the landslide events to derive the power-law relation. As a first step, within this time window, we identify all rain events with a pre-defined threshold (see Methods: section 2.2.2) that may cause landslides. Among these selected storm events, we consider the event with the maximum rainfall intensity as the triggering
event, whereas the other associated events with modest to low intensity are considered non-triggering events (Figure 9). Considering both regions, ~28% of cases show the occurrence of landslides within a week of the maximum rain intensity. Figure 9 maps the occurrence of landslides preceded by the triggering and non-triggering precipitation events during the peak-rain seasons during the winter season for region 1 (Figure 9a) and monsoon season for region 2 (Figure 9b). Overall, Figure 9 suggests 22-26% of landslides have occurred within a week of triggering rain events during the peak rain seasons of both regions between 2007 and 2016. This is alarming since an earlier assessment has reported only 8-17% of landslides are linked to 3-5 days preceding heavy rainfall over Darjeeling, the eastern Himalayan range (Mandal and Sarkar, 2021). In contrast, our analysis suggests a strong susceptibility of preceding heavy rain events on landslides over the WHR. We compare the triggering versus non-triggering rain events and find that triggering rain events overlaps the non-triggering rain events and has higher magnitudes overall (Figure 10). We find (Figure 10a-f) that the maximum rainfall intensity (triggering) ranges from 6.24 mm/hr to 0.05 mm/hr across all sites, with an average $I_m$ of 0.75 mm/hr. The durations corresponding to these maximum intensities range from 24- to 432 hours (or around 18 days). During these events, the rain magnitudes are as high as 748.8 mm and as low as 1.2 mm with an average of 72 mm. Further, the number of landslide events tends to decrease with an increase in duration (Figure 11). This implies landslides are triggered by short-duration high intensity rainfall.

Next, we fit a polynomial curve considering the triggering rain events and find the power-law relation for each site individually (Figure 11a-f) following Eq. 2. The positive Kendall’s tau estimates for these sites demonstrate a robust positive dependence between rainfall magnitude and duration (p-value < 0.05). For each of these at-site empirical ID threshold relations, we find a robust fit with a coefficient of determination ($R^2$) values of more than 0.9. The derived ID threshold curves show a positive scaling exponent $\lambda$ that varies between 0.8 and 1.8 with an intercept, $\alpha$ parameter in the range of 0.0005 to 0.02.

Next, we present the regional empirical ID threshold curve for the WHR: since we have considered Banihal as region 1, the regional and at-site ID threshold relation remains the same for this gauged site. For region 2, the empirical ID threshold relationship follows the power-law relation with an $R^2$ estimate of 0.62 (see Figure 12):

$$I_m = 0.0034d^{1.19}$$ (3)
The 95% confidence bounds of the empirical ID threshold curves suggest a notable uncertainty towards the lower tail as indicated by the larger spread of the bounds. In contrast, the uncertainty envelope tends to narrow down towards the upper tail. As a mechanistic interpretation, Eq. 3 suggests a rainfall intensity of 0.36 mm/hr will require duration of around two days (i.e., 50 hours) to trigger a landslide event in the Region 2 of the WHR. Our derived positive scaling exponents in empirical ID threshold relations corroborate with an earlier study (Abraham et al., 2020), which showed a positive scaling relationship between the cumulative total rainfall preceding the landslides to the corresponding event duration. However, our study slightly differs from Abraham et al. (2020): while in later, cumulative rainfall preceding the landslide events are considered in developing the ID threshold curves, we consider the maximum triggering rain intensity to derive the empirical ID threshold relations. Second, study areas are different: Abraham et al., 2020 considered the Darjeeling region, which belongs to the eastern Himalayan range, whereas we focus our analysis on the western Himalayan range. Finally, we show that around 28% of landslides have occurred within a week of occurrence of triggering rain events. Our results are in agreement with earlier studies that show landslide hazard is most sensitive to preceding precipitation events and, in particular robust from 1-day to up to 30-day periods of precipitation accumulation (Bevacqua et al., 2021; Johnston et al., 2021).

Conclusions

Precipitation-induced landslides pose a severe threat to life and property in several regions across the globe. Landslides are part of cascade hazard processes, which may result from several interacting processes, such as heavy rainfall or snowmelt, and seismic activity, followed by a series of consequences, such as debris flow and GLOFs. This study contributes to understanding precipitation trends in the WHR and identifying triggering and non-triggering rain events over a 51-year instrumental period from 1970-2019 to analyse extreme rain events leading to landslides. Since rain gauges in the WHR often contain large gaps in observational time series, we investigate the credibility of statistical post-processing technique coupled with regularized EM algorithm to impute missing gaps in observed rainfall time series. Next, we establish a power-law relationship between triggering maximum rain intensity and corresponding event duration based on the decadal landslides archive available for the WHR region.
Further, we offer at-site as well as regional \( ID \) threshold curves depicting the rainfall control on shallow-to-deep landslides. Our main conclusions are as follows:

- The rainfall pattern across the WHR shows a notable spatial heterogeneity. While in most sites, peak rain events are concentrated around the southwest monsoon (June-September) period, in high latitude area, \( e.g. \), Banihal, the peak rain events are clustered around the late winter (between February-March) season. The peak rain events in Banihal during winter is primarily linked to increased winter season precipitation due to synoptic western disturbances.

- An apparent increase in landslide frequency in the post-2010s is possibly linked to an increase in rainfall in low elevated areas of the WHR during the post-2007s. While more than 50% of medium-sized landslide events occurred in the monsoon season, a significant increase in rainfall is spatially clustered around the latitudinal belts of \( 29^\circ N \) to \( 30^\circ N \), and \( 32^\circ N \) to \( 34^\circ N \).

- The observational evidence shows around 28% of the landslides occurred within a week of occurrence of triggering rain events. The maximum triggering rain intensity varies from 6.24 mm/hr to 0.05mm/hr, while the triggering rain magnitude varies from 748.8 mm to 1.2 mm.

- The regionalization of \( I_{\text{max}}D \) rainfall threshold relations show a simple scaling relation with a power-law dependence of \( I_{\text{max}} \) versus \( D \). While at a regional scale, we observe a robust positive scaling exponent of 1.19, at a local scale, the strengths of positive scaling exponents often ranges from 0.8 to 1.8 with values close to 1 indicates a strong influence of triggering rain events.

Overall, we highlight that the regularized EM algorithm is very skillful in capturing precipitation variability and its extremes at individual station locations. The added value from the analysis is threefold: First, the developed empirical \( I_{\text{max}}D \) rainfall threshold relations would help to predict landslide occurrences at regional and local scales. The derived insights based on rainfall thresholds would aid in the development of early warning systems. Second, identifying rainfall thresholds and detecting suitable time lags in triggering landslides would help in developing landslide forecast models based on hydroclimatic control. Finally, the derived insights inform climate adaptations in steep-slope agricultural systems, especially in densely populated areas (Tarolli et al., 2021; Wang et al., 2022). Furthermore, the obtained insights have direct implications towards rain-induced landslide susceptibility mapping of Asian
water towers in a changing climate (Kirschbaum et al., 2019). As a caveat, we acknowledge that observational constraints limit us to assess the power-law behaviour of triggering rain events based on only ten years of recent records (2007-16). Second, the analysis is solely based on rainfall at a daily temporal resolution; an extension to a sub-daily timescale would capture the role of short-duration intense convective storm events in modulating landslides initiation. As a future effort, we plan to assess the role of climate and topographic controls, for instance, impact of antecedent soil wetness, soil properties, and terrain factors in triggering landslides, and develop a probabilistic ensemble forecast model using hydroclimatic and topographic forcings.

**Author Contributions.** Formal analysis: SRD; Visualization: SRD; Software: SRD and PG; Conceptualization: PG and SRD; Data curation: PG and SRD; Investigation: SRD and PG; Methodology: PG and SRD; Initial computer code preparation: PG; Supervision: PG; Writing – Original draft: SRD; Writing – writing, review and editing: PG.

**Competing Interests.** The author declare no competing interests.

**Funding.** The work did not receive any significant funding. SRD is supported through MHRD scholarship from the government of India to carry out the Master’s thesis at IIT Kharagpur.

**Code and Data Availability.** RegEM code is available in public repository ([https://climate-dynamics.org/software/](https://climate-dynamics.org/software/)). Station-based daily rainfall time series is retrieved from the India Meteorological Department (IMD)’s Data Supply Portal ([https://dsp.imdpune.gov.in/](https://dsp.imdpune.gov.in/)). Gridded rainfall time series for the WHR region is obtained from the archived gridded records of IMD, available at: [https://www.imdpune.gov.in/Clim_Pred_LRF_New/Grided_Data_Download.html](https://www.imdpune.gov.in/Clim_Pred_LRF_New/Grided_Data_Download.html). The MATLAB codes for this study can be made available upon reasonable requests from the authors. The historical landslide inventory was obtained from the NASA Cooperative Open Online Landslide Repository (COOLR) available at: [https://catalog.data.gov/dataset/global-landslide-catalog-export](https://catalog.data.gov/dataset/global-landslide-catalog-export).
Table 1: Site details, availability of rainfall records and seasonal distribution of rain for selected sites

<table>
<thead>
<tr>
<th>WMO ID</th>
<th>Station Name</th>
<th>District</th>
<th>State</th>
<th>Latitude (ºN)</th>
<th>Longitude (ºE)</th>
<th>Elevation (m)</th>
<th>Record Availability</th>
<th>Annual Average Rainfall (mm)</th>
<th>Seasonal Rainfall (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>42045</td>
<td>Banihal</td>
<td>Ramban</td>
<td>Jammu and Kashmir</td>
<td>33.50</td>
<td>75.17</td>
<td>1623.7</td>
<td>1985-2019</td>
<td>1391.1</td>
<td>632.23 229.64 339.0 190.21</td>
</tr>
<tr>
<td>42078</td>
<td>Mandi</td>
<td>Mandi</td>
<td>Himachal Pradesh</td>
<td>31.72</td>
<td>76.97</td>
<td>760</td>
<td>1985-2010</td>
<td>1178.7</td>
<td>171.31 81.16 872.34 53.91</td>
</tr>
<tr>
<td>42106</td>
<td>Solan</td>
<td>Solan</td>
<td>Himachal Pradesh</td>
<td>30.84</td>
<td>77.14</td>
<td>1256.4</td>
<td>1988-2019</td>
<td>737.8</td>
<td>158.17 68.02 459.33 52.29</td>
</tr>
<tr>
<td>42111</td>
<td>Dehradun</td>
<td>Dehradun</td>
<td>Uttarakhand</td>
<td>30.31</td>
<td>78.03</td>
<td>620</td>
<td>1970-2017</td>
<td>2278.9</td>
<td>162.16 89.95 1955.05 71.76</td>
</tr>
<tr>
<td>42116</td>
<td>Joshimath</td>
<td>Chamoli</td>
<td>Uttarakhand</td>
<td>30.55</td>
<td>79.57</td>
<td>1875</td>
<td>1985-1987</td>
<td>1003.8</td>
<td>144.4 144.01 645.26 70.14</td>
</tr>
</tbody>
</table>
Figure 1: Spatial locations of rain gauges across the WHR (See Table 1 for details). The elevational profile shows the locations of high hills across the northern part of the WHR, which gradually decreases from the north to the south. The digital elevation model of 1-Arc second (approximately 30 m) spatial resolution was derived from the SRTM-1 Arc Second Global data product archived at USGS Earth explorer (https://earthexplorer.usgs.gov/). The elevation map of India is projected using spatial analysis software Arc GIS Desktop version 10.8.1. The inset shows the location of the WHR over the Indian subcontinent.
Figure 2: Detailed flowchart of the methodology. The India map showing landslide inventory was retrieved from Bhuvan portal, hosted by the Indian Space Research Organization (https://bhuvan.nrsc.gov.in/home/index.php).
Figure 3: Identification of homogenous regions based on geographical and climatic attributes

(a) Explained variance by different principal components (PCs) (b) Biplot of the PCs. Red colour dots indicate the clusters of the selected sites. Blue arrows indicate the weightage of different attributes in the PCA space.
Figure 4: Standardized rainfall anomaly plot showing seasonal variation of rainfall for all sites. The temporal distribution of rainfall anomaly time series shows two distinct climatological regions.
Figure 5: Temporal evolution of monthly precipitation versus the number of landslides for (a) Region 1 (b) Region 2. The landslide frequency (the number of events) are smoothed using the Locally weighted Scatter plot smoothing (LOWESS) regression, shown in solid thick lines in black with a span length of 0.75.
Figure 6: Spatial distribution of historical landslides versus the mortality. (a) Landslide inventory map for the period 2007-2016 and (b) temporal distribution of fatality.
Figure 7: Spatial distributions of rainfall versus the distribution of landslides over WHR. Spatial distribution of rainfall during (a) 1988-2006 (b) 2007-2016 and (c) Changes in annual rainfall climatology during the recent (2007-2016) versus retrospective (1988-2006) periods. The grids with significant positive median changes (2007-2016 vs 1988-2006) are marked in ‘+‘ symbol.
Figure 8: Heat map showing performance statistics of sites across selected time windows for observed versus (a) SPP-WQM and (b) SPP-QM. The colour bar shows the performance metric obtained from the median of the rainfall and temporal signatures of RegEM simulated time series relative to the observed time series as suggested by Beck et al. (2017).
Figure 9: Temporal contiguity of triggering and non-triggering rain events followed by landslide events for each region (a) Region 1 showing winter season and (b) Region 2 depicting monsoon months. The day of landslide is marked in ‘x’.
Figure 10: Intensity-duration ($I_{\text{max}}D$) plots of triggering versus non-triggering rain events. The storm event with maximum rain intensity, $I_m$ preceding 30-days of landslide event is considered for the determination of triggering event.
Figure 11: Development of at-site power-law relation for selected gauges (a-f) across WHR. The ID threshold relationship was derived considering the maximum rain intensity versus corresponding event duration. The confidence bounds correspond to 95% confidence interval obtained from the least-square regression fit.
Figure 12: Regional ($I_{\text{max}} D$) threshold curve for region 2 of the WHR.
References


