Relationship of Landslide and Rainfall in Hong Kong

Name: Henry Chan
Academic Program or Faculty: Major in Environmental Engineering, Faculty of Engineering

Summary (150-200 words)

Hong Kong has been threatening by landslide, especially landslide that is cause by rainfall. To develop the city while protecting citizen's life and property, it is essential to prevent and predict landslide. In this essay, the relationship of landslide and rainfall is examined by data driven method. Three public datasets, containing historic information of landslide incidents, daily rainfall and rainstorm warning signal posted are used to perform exploratory data analysis, in the aim to solve the question that "How does rainfall increase the chance of landslide". The analysis is conducted by answering four subsequent questions regarding to the relationship between landslide and rainfall pattern, rainfall and type of slope, rainfall and scale of failure, as well as landslide and intensive rainstorm, respectively. The result confirms that rainfall pattern has close relationship with landslide, while intensive rainfall may act as the trigger point of a landslide. Also, man-made slope is confirmed to be more vulnerable to withstand rainfall from landslide. However, there is not any clear correlation found between rainfall and scale of failure.

Background and introduction

Hong Kong is the fourth densely populated place in the world, with population density of about 6800 people per kilometers (1). With the shortage of plains suitable for residential usage, it is common to see tall buildings located along the hillside in Hong Kong. This introduced many man-made slopes, for example, retaining walls, cut slopes, and filled slopes, to provide the houses a concrete base even though there are not supposed to be. Man-made slope is said to be more sensitive to change of water content inside the rocks and soil. As a result, they are more vulnerable to rainfall-induced landslide (2). This could be a major threat to citizen's life and properties. Also, roads might be shut down due to debris from landslide, causing secondary impact on transportation, social infrastructure, and economy. Government departments like Geotechnical Engineering Office (GEO) under Civil Engineering and Development Department (CEDD), and Hong Kong Observatory (HKO) have been investigating in prevention and prediction of landslide after several severe landslides occurred in the 1970s, causing hundreds of causalities (3). However, it is difficult to obtain a clear relationship between landslide and rainfall since there many hidden factors like underground water level that may contribute to the incident. In this essay, the relationship between landslide and rainfall will be investigated using public datasets of landslide and rainfall record.

Description of the source(s) of the data

Three datasets are used for the analysis. The first one is Location of Landslide Incidents provided by CEDD. The landslide dataset contains records of all landslide incidents occurred between October 1983 and December 2020. It has 17 columns with over 9800 records, providing details of a landslide, for example, date of failure, location, type of slope, and number of consequences (death, injury, etc.), which is the major target of the analysis.

The second dataset used is Daily Total Rainfall All Year - Hong Kong Observatory provided by HKO. The rainfall dataset contains records of daily rainfall measured at HKO from March 1884 to June 2022. Although there are datasets obtained from other measuring stations, for the purpose of this project, the above dataset will be used as the representative daily rainfall in Hong Kong.

The third dataset used is Rainstorm Waring Signal Record provided by HKO, and it is uploaded onto Kaggle as a .csv file by author. The signal dataset contains posting date, ending date, duration, color of rainstorm warning signal posted from April 1998 to July 2022. This could be a fair indicator of how intensive the rainfall is taking place.

Question(s) or hypothesis

The main question of this project is "How does rainfall increase the chance of landslide". To answer the question, the analysis is carried out by focusing on four parts:

Q1) The pattern of rainfall and landslide;

Q2) Relationship between rainfall and types of slope;

Q3) Relationship between rainfall and scale of failure;

Q4) Relationship between rainstorm warning signal and landslide.

Methods used

R is the programming language used throughout the analysis, specifically using its tidyverse, lubridate and janitor packages. First, the above three raw datasets are downloaded and imported into R. For each dataset, a R script, naming "landslide.r", "rf.r" and "signal.r" respectively, is used for cleaning and tidying the dataset. By referring to the reference sheets and comments inside the original .csv file, the variables in datasets are transformed into proper data types while not available (NAs) and NA-equivalent records are treated with regards to reality. Also, duplicated records, for example, due to capitalization, are also unified to keep simplicity. At the end of the cleaning script, cleaned dataset is exported as .csv for further analysis.

For analysis, another R scripted named "intpre.r" is created to code for answering Q1 to Q4. This will at first read in the cleaned datasets processed as mentioned above. Then, the cleaned landslide dataset and rainfall dataset is merged to conduct exploratory data analysis (EDA).

In Q1, the pattern of rainfall and landslide is examined by plotting their distribution. It is important to include previous rainfall (in this project, past 5 days, past 10days, past 14 days and total=past 14 days + rainfall on day of incident, are chosen to be the indicators) as landslide is found to have connection with prior factor (4). Empirical cumulative distribution function (ECDF) is used to contrast how different categories of rainfall relate to distribution of landslide. Besides, histogram of average monthly rainfall and occurrence of landslide per month is used to investigate the trend of the two. If a positive correlation is observed, it is concluded that rainfall is a crucial factor for landslide. This will provide solid evidence for the later analysis.

In Q2, relationship between rainfall and types of slope is examined by plotting boxplots. While feature of slope (e.g. retaining wall, natural hillside, disturbed terrain, etc.) is given in the raw datasets, type of slope, naming man-made slope, natural slope and other, is not provided. Therefore, the merged dataset is first manipulated to classify the three kinds of slope. Boxplots are then used to compare rainfall in different genre of slopes. It is done to check if man-made slope were actually more vulnerable to rainfall, i.e. lead to landslide more frequently. Mean of rainfall in each type of slope is used as an indicator for the purpose. In Q3, relationship between rainfall and scale of failure is examined by plotting jitter point graph. It is noticed that the value in scale of failure is polarized, i.e. most of the values are withing single-digit while a few of them are over thousands, due to several major incident. To cope with the above problem, a suitable range of value is selected by referring to the summary statistics of scale of failure. Correlation coefficient is used to confirm the visualized result. This is to check if heavier rainfall is causing greater landslide, which is normally precepted by most of the people.

In Q4, relationship between rainstorm warning signal and landslide is examined by computing odds and odds ratio. Odds is used to compare the chance of landslide under different levels of rainstorm warning signal. Odds ratio is used for a more general investigation on chance of landslide under the posting of any rainstorm warning signal. The cleaned signal dataset is introduced and merged with cleaned landslide dataset, to obtain the contingency table for calculation of odds. For calculation of odds ratio, cleaned rainfall dataset is filtered to join landslide dataset and signal dataset together for obtaining the contingency table. By doing so, it is able to check if intensity of rainfall in short term is having impact of landslide.

Results: description of findings and results

Figure 1 shows the ECDF of landslide against different categories of rainfall. It is observed that when previous rainfall is considered, the curve shifts to the right, suggesting a higher median and mean of rainfall.

Figure 2 shows the distribution of average monthly rainfall and total number of landslides from 1983 to 2020. It is obvious that both share the same pattern, having greater amount in summer and smaller amount in winter.

Figure 3 shows boxplots of total rainfall against features of slope. Natural hillside has the highest median among other features, yet it is not a great difference.

Table 1 shows the mean of total rainfall among different features of slope. It is observed that disturbed terrain has the highest mean, followed by natural hillside.

Figure 4 shows boxplots of total rainfall against types of slope. Average total rainfall of man-made slope seems to be lower than that of natural slope, yet again it is not a great difference.

Table 2 shows the mean of total rainfall among different types of slope. It is observed that natural slope has the highest mean, about 6% greater than that of man-made slope.

Figure 5 shows the jitter-point graph of total rainfall against scale of failure. Though a slightly concentration of point near the bottom left-hand side of the graph, generally points are evenly distributed.

Table 3 shows the correlation coefficient of different kind using the data from Figure 5. Most of them are closed to zero, suggesting that there might not be correlation between total rainfall and scale of failure.

Table 4 shows the odds and probability of landslide under different level of rainstorm warning signals. Higher the level the warning is, the greater the odds and probability of landslide are.

Table 5 shows the contingency table of landslide and rainstorm warning signal. From this, odds ratio is calculated as 8.79.

Insights into the data

Figure 1 and figure 2 provides strong evidence that landslide is closely related to landslide. From figure 1, most of the landslide occurred even though there is no or very little rainfall on the day of incident. However, as previous rainfall is included, the reality is revealed: accumulation of water content due to previous rainfall is acting an important role of causing a landslide. From figure 2, it is deduced that the amount of rainfall is positively related to the occurrence of landslide. However, there are a few landslides occurred in winter which has relatively small amount of rainfall.

Figure 3 and 4, table 1 and 2 may suggest the answer to Q2 that man-made slope is more vulnerable to rainfall. From table 1, although disturbed terrain has a higher mean than natural hillside, it has a relatively much smaller sample size. It might be the outcome due to this small sample size. Therefore, it is more appropriate to state that natural hillside has a higher mean among other features of slope. From table 2, it is once again confirmed that natural slope has a higher mean than man-made slope. With the large sample size of over 81% of landslide results from man-made slope, it is possible to claim that man-made slope is subject to landslide, because many of them are not well-designed and poorly maintained. Figure 5 depicts a different story of what people have used to believe. Instead of a positive linear relationship between total rainfall and scale of failure, it looks more like a randomly plotted graph. Table 3 suggests ignorable positive correlation coefficient, with the spearman's being the highest, it does not align with the visualization of figure 5. Therefore, it is hard to claim any clear correlation between total rainfall and scale of failure. This might be attributed to the great number of minor incidents, which have scale below 5 m^3, causing both visual and mathematical investigation nearly impossible.

Table 4 and 5 are considered supporting materials for Q4. Odds become larger when the level is rising, from amber to red, and from red to black. All of them guarantee a chance of over 80% of landslide when they are issued. From table 5, the odds ratio is greater than 8, ensuring the claim that intensive rainfall is triggering landslide more frequently than normal rainfall.

Discussion of results and perspective

The above results are induced without any statistical testing. To make claims with confidence, for example, t-test should be carried in Q2, to ensure the higher mean of natural slope is not induced by random chance. Also, random sampling should also be used to obtain a more general trend, for example, to plot the jitter-point graph in Q3. It will also be useful in calculation of odds and odds ratio in Q4.

Throughout this analysis, location of incident and rainfall is ignored (or in other words, normalized). However, it is obvious that rainfall measured at HKO cannot completely represent the instant rainfall pattern in other places in Hong Kong. Some place might have sudden rainstorm in short time which might not necessarily causing HKO to post a rainstorm warning signal. As a result, these may promote a misleading insight s into data, especially in Q1, it is seen that many landslides have little rainfall on day of incident, while the reality might not be the same. More advanced technique, for example, map coordinate processing and text(address) matching, can be the solution to this problem.

Main conclusion and outcome

In this essay, it is confirmed that rainfall is increasing the chance of landslide in multiple ways. Firstly, seasonal rainfall pattern matches the distribution of occurrence of landslide. This is because in rainy summer, amount of water accumulates in the soil, and a sudden intensive rainstorm can be the trigger point for landslide. Secondly, man-made slope is confirmed to be more vulnerable than that of a natural slope when it comes to a landslide. Fewer amount of total rainfall can trigger a landslide in man-made slope. However, the scale of failure does not show a correlation with total rainfall.

References

1. Population density. Wikipedia. https://en.wikipedia.org/wiki/Population_densitydietary 2. Malone, A. (2011). Hong Kong's 35-Year War on Landslides. https://www.researchgate.net/publication/279954130_Hong_Kong's_35-Year_War_on_Landslides
3. Landslide risk in Hong Kong. Slope safety in Hong Kong. Civil engineering and development department. https://hkss.cedd.gov.hk/hkss/en/slope-safety-in-hong-kong/landslide-risk/index.html
4. Naranjo, L. (2007). Connecting rainfall and landslides. Earthdata. NASA. https://www.earthdata.nasa.gov/learn/sensing-our-planet/connecting-rainfall-and-landslides

Figures and tables

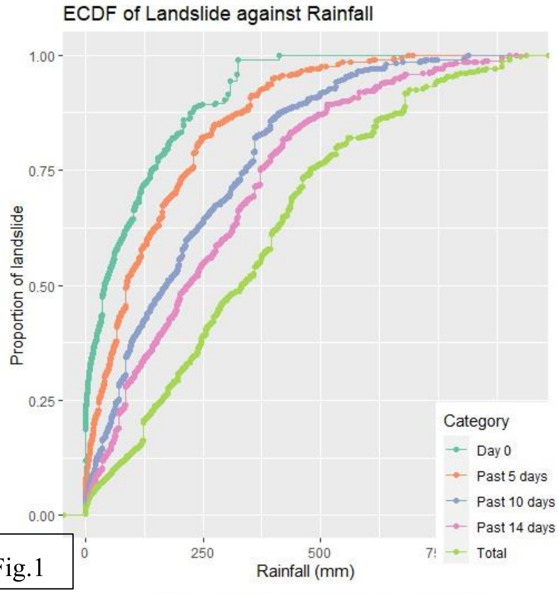

Fig.1

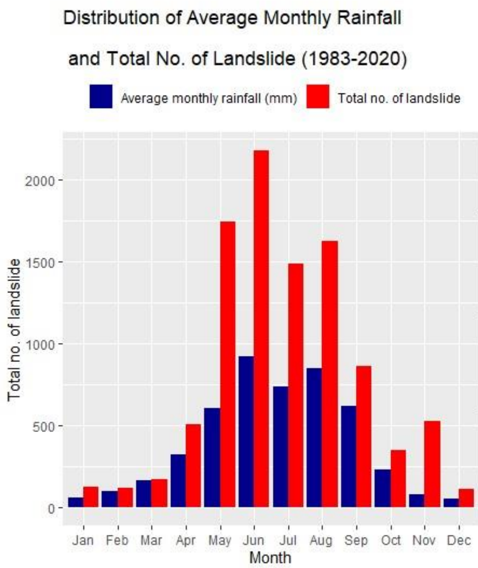

Fig.2

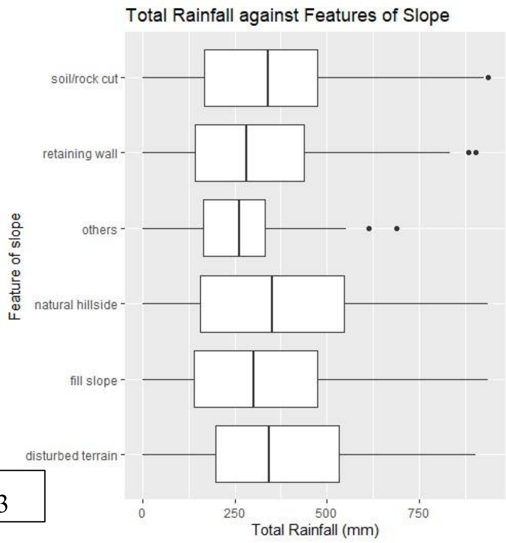

Fig.3

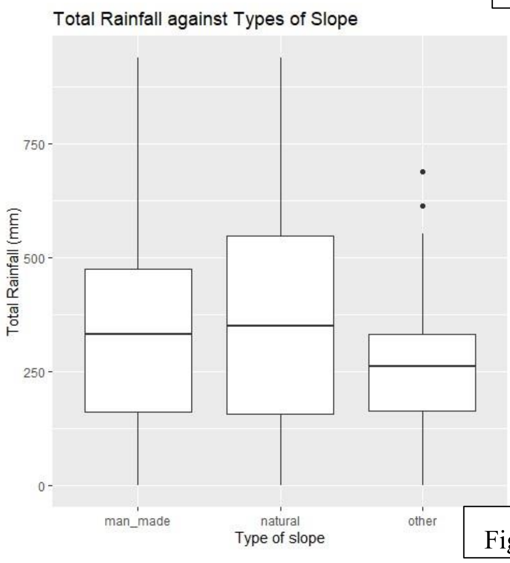

Fig.4

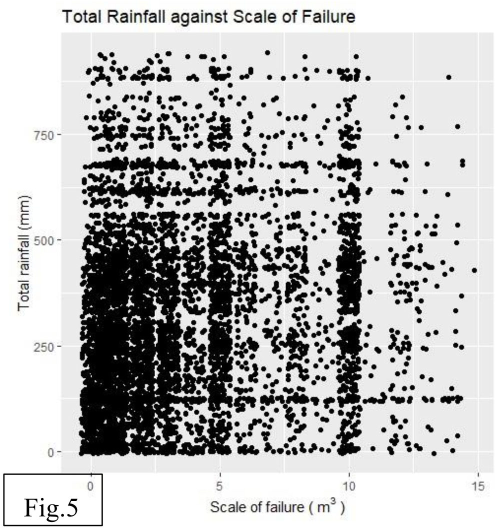

Fig.5

Tab.1

	feattype	n	mean
	<chr>	<int>	<dbl>
1	disturbed terrain	96	371.
2	fill slope	674	335.
3	natural hillside	1792	368.
4	others	61	272.
5	retaining wall	596	309.
6	soil/rock cut	6611	349.

Tab.2

	stype	mean
	<chr>	<dbl>
1	man_made	345.
2	natural	368.
3	other	272.

	pearson	kendall	spearman
	<dbl>	<dbl>	<dbl>
1	0.144	0.125	0.179

	rep_color	odds	prob
	<chr>	<dbl>	<dbl>
1	Amber	5.39	0.844
2	Black	492.	0.998
3	Red	47.7	0.979

	landslide	
signal	No	Yes
No	6371	1483
Yes	150	307