Accepted manuscript doi:
10.1680/jgeot.21.00160

Submitted: 25 June 2021

Published online in 'accepted manuscript' format: 07 January 2022

Manuscript title: Predicting spatio-temporal man-made slope failures induced by rainfall in Hong Kong using machine learning techniques

Authors: Te Xiao*, Li Min Zhang*, Raymond Wai Man Cheung[†] and Suzanne Lacasse[‡]

Affiliations: *Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, China; [†]Geotechnical Engineering Office, Civil Engineering and Development Department, The Government of the Hong Kong Special Administrative Region, Hong Kong, China and [‡]Norwegian Geotechnical Institute, Oslo, Norway

Corresponding author: Li Min Zhang, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, China.

E-mail: cezhangl@ust.hk

Accepted manuscript doi:
10.1680/jgeot.21.00160

Abstract

Rain-induced man-made slope failures pose great threats to public safety as most man-made slopes are formed in densely populated areas. A critical step in managing landslide risks is to predict the time, locations, and consequences of slope failures in future rainstorms. Based on comprehensive databases of in-service man-made slopes, rainstorms, and landslides in Hong Kong in the past 35 years, a spatio-temporal landslide forecasting model for man-made slopes is developed in this study within a unified machine learning framework. With a storm-based data integration strategy and multiclass classification on landslide scales, the framework incorporates landslide time and consequences into landslide susceptibility mapping to successfully achieve spatio-temporal landslide forecasting. The machine learning-based landslide forecasting model is validated against historical landslide incidents both temporally and spatially and through a case study of the June 2008 storm, and significantly outperforms the prevailing statistical rainfall-landslide correlations in the prediction accuracy. The model can predict the real-time evolution of probabilities, scales, and spatial distribution of landslides during the progression of a rainstorm, which can never be achieved by statistical methods. It can serve as an essential module for state-of-the-art landslide risk assessment and early warning.

Keywords: artificial intelligence; landslides; slopes; statistical analysis; uncertainty, reliability & risk

Accepted manuscript doi:
10.1680/jgeot.21.00160

INTRODUCTION

Slope safety is an essential issue in Hong Kong owing to its mountainous topography and subtropical monsoon climate. The long rainy season lasts from April to October, and the rainfall intensity can be high, with 50 to 100 mm per hour and 250 to 350 mm in 24 hours being not uncommon. On average about 300 landslide incidents were reported annually, and the majority of them were triggered by rainstorms (e.g., Au, 1998). Compared with natural terrain landslides in remote areas, man-made slope failures in urban areas usually pose greater threats to public safety. A vast number of man-made slopes formed before 1977 without much engineering input are particularly vulnerable to rainstorms. Records show that hundreds of man-made slopes could fail in a single severe rainstorm. For example, the 6-9 June 2008 rainstorm, whose 4-h rainfall corresponded to a return period of about 1000 years, triggered 221 man-made slope failures. The Geotechnical Engineering Office (GEO) of the Hong Kong SAR Government has taken many engineering and non-engineering measures to mitigate landslide risks, among which establishing a regional landslide early warning system (Kong et al., 2020; Cheung, 2021) is of great importance. The success of such a warning system highly relies on a reliable landslide forecasting model.

Most regional landslide early warning systems in the world adopt rainfall thresholds (e.g., intensity-duration relations) as landslide indicators, while only a few take statistical rainfall-landslide correlations or physically-based distributed models to predict landslides (Piciullo et al., 2018; Guzzetti et al., 2020). Although physically-based models (e.g., Crosta and Frattini, 2003; Baum et al., 2010; Chen and Zhang, 2014; Shen et al., 2018) can simulate the whole spatio-temporal process of landslides by incorporating the physical mechanisms of slope failure, such models may be less suitable for prompt prediction of impending landslides because of high computational demands. Data-driven approaches such as rainfall thresholds

(e.g., Guzzetti et al., 2007; Segoni et al., 2018) and rainfall-landslide correlations (e.g., Dai and Lee, 2001; Yu et al., 2004; Ko and Lo, 2016; Gao et al., 2018) are able to forecast landslides in real time, providing the temporal variation of the total number of landslides. Hong Kong took the lead in applying multiple statistical correlations between rainfall intensity and landslide frequency regarding different types of man-made slopes to its landslide early warning system. From the perspective of landslide risk assessment (e.g., Fell, 1994; Dai et al., 2002; Cheung and Tang, 2005; Lacasse and Nadim, 2009; Corominas et al., 2014; Li et al., 2016; Xiao et al., 2016), the total number of landslides may not be a perfect measure of landslide risk due to the lack of spatial distribution and consequences; for instance, a large landslide occurring in a densely populated area certainly can pose a higher risk than a small landslide occurring in a remote area. Therefore, the desired model shall be a spatio-temporal forecasting model that predicts not only the temporal variation of the number of landslides but also the landslide location, type, scale, runout path, and elements at risk so that the landslide consequences can be properly assessed.

Machine learning has recently emerged as a powerful data-driven tool in landslide science. From the simplest logistic regression to emerging deep learning, the machine learning methods have broad applications in landslide displacement prediction (e.g., Lian et al., 2014; Krkač et al., 2017; Zhou et al., 2018; Yang et al., 2019), landslide detection/identification (e.g., Stumpf and Kerle, 2011; Ghorbanzadeh et al., 2019; Su et al., 2021; Wang et al., 2021b), and landslide susceptibility mapping (e.g., Dai and Lee, 2002; Frattini et al., 2010; Ching et al., 2011; Van Den Eeckhaut et al., 2012; Goetz et al., 2015; Bui et al., 2016; Lombardo et al., 2020; Huang et al., 2020; Merghadi et al., 2020; Wang et al., 2021a). As represented by the landslide susceptibility mapping, machine learning methods are good at predicting the spatial distribution of landslides by fusing multi-source

data of rainfall and various geographic factors governing slope stability. Opposite to statistical correlations, previous machine learning applications did not determine the time of landslides in different rainstorms, hence were rarely ultilised for landslide early warning. To achieve a spatio-temporal prediction of landslides, it is necessary to integrate the advantages of statistical correlations in temporal prediction and machine learning techniques in spatial mapping.

This study aims to develop a spatio-temporal landslide forecasting model for man-made slopes in Hong Kong within a unified machine learning framework. Comprehensive databases of man-made slopes, rainstorms, and landslides in the past 35 years are compiled and integrated storm by storm to connect man-made slope failures to their triggering rainstorms. A multiclass machine learning model is trained on the storm-based integrated dataset to simultaneously predict landslide time, locations, and consequences. The machine learning-based landslide forecasting model is validated against historical landslide incidents temporally and spatially and compared with statistical rainfall-landslide correlations. The most influential factors controlling man-made slope failures are also identified.

COMPREHENSIVE DATABASES FOR HONG KONG

Machine learning emerges as an attractive alternative to statistical correlations because a large amount of data on in-service man-made slopes (Fig. 1, since 1977), rainfall (Fig. 2, since 1884), and landslides (Fig. 3, since 1984) are available in Hong Kong. For consistency, the time span is selected from 1984 to 2017 for all databases in the present study.

The Slope Information System (SIS) managed by GEO had registered more than 60,000 sizeable man-made slopes, as shown in Fig. 1. In Hong Kong, man-made slopes refer to cut slopes, fill slopes, and retaining walls; A retaining wall could be an independent feature or associated with a fill/cut slope. After data cleansing, a total of 59,763 man-made slopes are

available with complete information on man-made features, geometry, location, formation time, slope materials, slope covers, and drainage conditions. The majority of the slope population are cut slopes (59%), followed by retaining walls (37%) and fill slopes (22%). There are 81% of slopes made up of soil material only, 16% of both soil and rock materials, and 3% of rock material only. As counted in Fig. 4(a), 86% of slopes are lower than 15 m, 81% are sloping between 30° and 70°, and more than 90% of retaining walls are lower than 10 m and nearly vertical (i.e., wall angle $\geq$ 80°).

The spatial distribution of the mean annual rainfall amount in Hong Kong (1990-2010) is illustrated in Fig. 2, ranging from about 1400 mm in the northeast to more than 3000 mm in Tai Mo Shan. The latest rain gauge network comprises 50 HKO (Hong Kong Observatory) rain gauges and 91 GEO rain gauges, with an average density of 10 km^2/gauge. In the period of 1984-2017, 419 major rainstorm events (see Appendix 1) have been identified from the annual reports on rainfall and landslides published by GEO, as shown in Fig. 5. In the same period, 7933 man-made slope failure incidents were reported to GEO, of which 5063 were triggered by the 419 rainstorms according to their failure time (i.e., amid the rainstorm or one day after the end of rainstorm). Among these landslides, 2127 man-made slopes had been registered in SIS at the time of failure. Spatially distributed as shown in Fig. 3, these failed slopes are sorted in a sequence of triggering storms, as shown in Fig. 5. Only these cases will be considered in this study, as complete information on both slope and rainfall features is available.

The catalogue of landslide incidents includes feature ID (if in SIS), failure location, time, type, and volume, as well as landslide consequences to persons, roads, and buildings. Figure 6 illustrates an example of a landslide incident that was a major failure of a retaining wall (6SW-C/CR797) during the 6-9 June 2008 storm (Lam et al., 2012). It involved the

collapse of a 25 m long, 2.8 m high concrete retaining wall and part of the slope behind the wall, leading to two fatalities. The landslides are categorised into four failure scales according to their volumes (Fig. 4(b)), i.e., very minor (landslide volume $V < 5$ m^3) (45.0%), minor (5 m^3 $\leq V <$ 50 m^3) (44.0%), major (50 m^3 $\leq V <$ 500 m^3) (9.4%), and very major ($V \geq$ 500 m^3) (1.6%) failures. Major failure types include sliding, washout, rockfall, and wall failure. Sliding (65.2%) is the primary failure type, followed by washout (16.4%) and rockfall (11.0%); the frequencies of wall and other failures are relatively low. Washout and rockfall are usually associated with much smaller failure volumes compared to sliding and wall failure.

In addition to slope, rainfall, and landslide databases, geographic databases of terrain, geology, land cover, and infrastructure are also utilised in this study. As shown in Table 1 and Appendix 1, 43 input learning features and one output landslide feature are eventually extracted from the abovementioned databases to form a record of landslide or non-landslide. Features like slope information, terrain, geological condition, land cover, location, and annual rainfall (i.e., SF1-AR, 29 features) change little temporally, while those storm-related rolling and antecedent rainfall features and landslide scale (i.e., R1h-LS, 15 features) vary significantly from one storm to another. They are referred to as static data and dynamic data, respectively. The various uncertainties associated with geomaterials (e.g., El-Ramly et al., 2005; Lloret-Cabot et al., 2014; Wang and Zhao, 2017; Xiao et al., 2018, 2021; Hicks et al., 2019; De Gast et al., 2021) and rainstorms (e.g., Yuan et al., 2019; Qiang et al., 2020) are not considered.

Accepted manuscript doi:
10.1680/jgeot.21.00160

SPATIO-TEMPORAL LANDSLIDE FORECASTING USING MACHINE LEARNING

Machine learning framework

The Hong Kong landslide database, with the time of man-made slope failures recorded, provides an ideal opportunity to bring landslide time into conventional landslide susceptibility mapping. Lombardo et al. (2020) made progress in this aspect by directly taking the landslide time as one of the prediction features using machine learning techniques. This manner, however, can only predict the time within the provided categories. By contrast, the statistical rainfall-landslide correlations (Fig. 7) (Yu et al., 2004; Wong et al., 2014; Xiao and Zhang, 2020) employed in the landslide early warning system of Hong Kong relate the landslide time indirectly through rainstorms, i.e., the maximum rolling 24-h rainfall amount of the triggering storm. Inspired by this, a unified machine learning framework is proposed for spatio-temporal landslide forecasting, as shown in Fig. 8. It consists of three major components: (1) storm-based data integration to link the pairwise dynamic data, i.e., rainfall and landslide data in different storms, as shown in the upper part of Fig. 8; (2) a machine learning-based landslide forecasting model for predicting a well-chosen landslide feature that characterises landslide consequences, as shown in the lower central part of Fig. 8; and (3) spatio-temporal forecasting of the probabilities, locations, and consequences of landslides over time given a sequence of rainfall forecasts, as shown in the lower part of Fig. 8.

The storm-based data integration plays a pivotal role in connecting each man-made slope failure to its triggering rainstorm so that landslide time can be incorporated into machine learning to realise temporal prediction. The traditional data integration in machine learning (Fig. 9(a)) treats the 2127 landslides as 2127 failure records and then randomly selects the same number of safe examples to form a dataset with 2127×2 records. This

method significantly underestimates the safe records. In fact, every slope has been subjected to numerous repeated ‘landslide tests’ during hundreds of past rainstorms, as shown in Fig. 9(b). More specifically, the 59,763 man-made slopes, having experienced 419 major rainstorms, result in a total of about 22 million ‘testing’ records in the integrated dataset (i.e., 59,763×419, exclusive of invalid records that one slope was not registered at the time of a specific rainstorm). The 22 million records comprise 44 features as shown in Table 1, in which the 29 static slope features need to duplicate 419 times in different storms to match the 15 dynamic rainfall and landslide features. Since only 2127 landslides occur in the 22 million records, the ratio of negative samples (i.e., non-landslides) to positive samples (i.e., landslides) exceeds 10,000, which is extremely imbalanced. The class imbalance is further intensified as these landslides are categorised into four classes with different failure scales.

Figure 9 plots the box-and-whisker statistics of dynamic rainfall features in the storm-based integrated dataset. The median rainfall amounts of landslide records are significantly higher than non-landslide records, particularly in terms of rolling rainfall features. The difference increases as the failure scale enlarges from ‘very minor’ to ‘very major’ category. The ratios of median values converge when the rolling rainfall duration is longer than 12 h for very minor and minor failures, while the ratios increase continuously for major and very major failures. This coincides with the engineering experience that long-duration heavy rainstorms are more likely to trigger man-made slope failures with larger volumes.

Landslide forecasting model

Six candidate machine learning algorithms are adopted to develop potential landslide forecasting models, including logistic regression, neural network, bootstrap aggregating (Bagging), adaptive boosting (AdaBoost), random under-sampling boosting (RUSBoost), and the subspace method. Besides, several statistical rainfall-landslide correlations have been

developed (Fig. 7), such as Yu et al. (2004) using the storm and landslide data in 1984-1995, Wong et al. (2014) in 1996-2010, and Xiao and Zhang (2020) in 1984-2017. This study takes the latest correlations in 1984-2017 to benchmark the machine learning models. After a preliminary comparison among the six machine learning methods, as provided in Appendix 2, the logistic regression is finally chosen as a representative of machine learning methods to develop a spatio-temporal landslide forecasting model within the proposed framework.

The logistic regression, an example of the generalised linear model (e.g., Bishop, 2006; Dobson and Barnett, 2018), is of high interpretability and can provide the probability of landslide occurrence that coincides with the landslide frequency adopted in statistical rainfall-landslide correlations. These advantages allow the machine learning model to replace the statistical model in an existing landslide early warning system without revising warning criteria. In previous studies that deal with only the occurrence of landslides (i.e., two classes, no failure or failure) but do not involve many landslide features, the landslide prediction is taken as a binary classification problem, and the probability of landslide occurrence can be described using a binomial logistic regression as:

$$P\left(y=1\middle|\boldsymbol{x};\boldsymbol{\theta}\right)=\frac{\exp\left(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\theta}\right)}{1+\exp\left(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\theta}\right)} \qquad (1)$$

where $\boldsymbol{x}$ is a vector of 43 learning features (i.e., SF1-A15d) as listed in Table 1; y is the landslide feature (i.e., LS) and y = 0 and 1 stand for non-landslide and landslide, respectively; and $\boldsymbol{\theta}$ is a vector of model parameters.

To empower the prediction of landslide consequences in the meantime, the landslide scale, which is an important indicator of consequences, must also be predicted, leading to a multiclass classification problem with five classes (i.e., no failure and four failure scales).

The probabilities of non-landslide and landslides in the four failure scales can be described, respectively, through a multinomial logistic regression as:

$$P\left(y=i\middle|\boldsymbol{x};\boldsymbol{\Theta}\right)=\begin{cases}\dfrac{1}{1+\sum_{j=1}^{n_s}\exp\left(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\theta}_j\right)}, & i=0\\ \dfrac{\exp\left(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\theta}_i\right)}{1+\sum_{j=1}^{n_s}\exp\left(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\theta}_j\right)}, & i=1,\ 2,\ \ldots,\ n_s\end{cases} \qquad (2)$$

where n_s is the number of failure scales (n_s = 4); y = 0, 1, 2, 3 and 4 stand for no failure, very minor, minor, major, and very major failure scales, respectively; and $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_i, i = 1, 2, \ldots, n_s\}$ is the ensemble of all model parameters. The maximum likelihood estimates of the model parameters can be obtained through the iteratively reweighted least squares method (Dobson and Barnett, 2018) as:

$$\boldsymbol{\Theta}^{(m+1)}=\left[\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}^{(m)}\boldsymbol{X}\right]^{-1}\left[\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}^{(m)}\boldsymbol{Z}^{(m)}\right] \qquad (3)$$

where $\boldsymbol{\Theta}^{(m+1)}$ is the vector of parameter estimates at (m+1)th iteration; $\boldsymbol{X}$ is the matrix form of learning features; $\boldsymbol{W}^{(m)}$ is a $n_s \times n_s$ diagonal weight matrix related to $\boldsymbol{\Theta}^{(m)}$; and $\boldsymbol{Z}^{(m)}$ is a vector depending on $\boldsymbol{\Theta}^{(m)}$, $\boldsymbol{W}^{(m)}$, and the estimated prediction error at mth iteration. The iteratively reweighted least squares method fits the generalised linear model efficiently, typically within tens of iterations.

Eventually, a multinomial landslide forecasting model is trained with all the 22 million records in the storm-based integrated dataset and 44 features shown in Table 1. Figure 11 plots a radar chart to showcase the estimated model parameters corresponding to the 43 learning features and with respect to the four failure scales. The model parameters are standardised by the standard deviations of respective features and then normalised by the norm of parameters regarding the same failure scale. A larger radius indicates a higher impact of a feature. The rolling rainfall amounts (i.e., R1h-R48h) are the most influential features,

followed by man-made features (i.e., SF1-SF2), slope geometries (i.e., SH-SA), slope materials (i.e., SM1), and geological conditions (i.e., GEO1-GEO3). The areas under the receiver operating characteristic curve (AUC), a common measure of model performance in machine learning, for the developed landslide forecasting model range from 0.93 to 1.00 for the non-landslide class and the four landslide classes of different scales (Fig. 20(b) in Appendix 2), indicating its power to distinguish between classes.

Spatio-temporal forecasting

Once a landslide forecasting model has been trained, the computational time for predicting landslides in a future rainstorm can be as short as few seconds. The prediction is performed directly on all the man-made slopes in Fig. 1 rather than the discretised cells as often used in a natural terrain landslide analysis. Those static slope features remain unchanged, and the antecedent rainfall features have been measured at the time of prediction. The only required inputs are the rolling rainfall features evaluated from both measured and predicted rainfall amounts. As illustrated in Fig. 8, the machine learning-based landslide forecasting model can produce the spatio-temporal evolution of man-made slope failures given a sequence of rainfall forecasts. Its preliminary output is the probabilities of each man-made slope falling in one of the four failure scales over time (i.e., Eq. (2)), which inherently provides the information of occurrence time, locations, and consequences of landslides. Although the predicted failure probability of a single slope cannot be verified directly, the failure probabilities of multiple slopes in a certain area can be summarised as the expected number of landslides and validated against historical landslide observations. For this purpose, the territory of Hong Kong is divided into a grid of 3294 cells (Fig. 12), each 750 m from east to west and 600 m from north to south, according to the gridding system of the Survey and Mapping Office. Figure 12 shows the spatial distribution of man-made slopes converted from

Accepted manuscript doi:
10.1680/jgeot.21.00160

Fig. 1 in the grid. There are 2062 cells containing man-made slopes with a maximum number of 350 in one cell. The expected number of landslides in a cell, either total or summed up across different failure scales, is a preferable indicator to characterise the spatial distribution of landslides and to identify the most landslide-prone areas in a storm, as will be illustrated later. Similarly, the predicted failure probabilities of all slopes can also be summarised over the progression of a rainstorm to be comparable with the statistical correlations.

VALIDATION AGAINST HISTORICAL LANDSLIDE INCIDENTS

Validation on spatio-temporal distribution of landslides

The performance of the machine learning-based landslide forecasting model is validated against the historical landslide incidents in the period of 1984-2017 (i.e., Fig. 5) both temporally and spatially. Owing to the class imbalance of the storm-based integrated dataset and the quantity-sensitive nature of landslide forecasting, model performance indices derived from the confusion matrix of classification may not be satisfactory, as illustrated in Appendix 2. Instead, a storm-based index, namely the correlation coefficient (ρ) between the observed and predicted numbers of landslides in all rainstorms, is used to measure the performance of the landslide forecasting model.

Figure 13(a) compares the predicted numbers of landslides by the machine learning model and the observed numbers of landslides in all storms using a bubble chart, in which the bubble size reflects the number of overlapping data points. The correlation coefficient increases from 0.85 for very minor failures to as high as 0.99 for very major failures, indicating a high prediction accuracy of the machine learning model as the landslide risk increases. The predicted total number of landslides in every storm is further compared with those from statistical correlations in Fig. 13(b). The data points of the machine learning model are more concentrated along the 45° line (i.e., perfect prediction) with a much higher

correlation coefficient of ρ = 0.93 than the statistical model (ρ = 0.80). Also included in Fig. 13(b) are the predicted landslides amid the ten major rainstorms in 2018 from an independent dataset. The correlation coefficient of the machine learning model only slightly decreases to 0.91, still much higher than the statistical model.

Figure 14 compares the observed and predicted spatial distributions of landslides in the prescribed grid, indicated by the cumulative number of landslides per cell from 1984 to 2017. The observed and predicted distributions are highly consistent, especially for Hong Kong Island and Kowloon, the two most landslide-prone areas. Such a spatial distribution cannot be obtained via statistical correlations. By comparison, the machine learning-based landslide forecasting model significantly outperforms the prevailing statistical rainfall-landslide correlations, not only in the prediction accuracy but also in the ability to predict both landslide locations and consequences.

Case study of the 6-9 June 2008 storm

The 6-9 June 2008 storm is one of the most severe rainstorms in the history of Hong Kong, with 1-h, 4-h, and 24-h rainfall amounts all falling in the top 5 rainstorms ever recorded. The storm triggered 221 man-made slope failures, including 65 very minor, 75 minor, 15 major, and 7 very major failures registered in SIS. The recorded rainfall amounts of all rain gauges during the June 2008 storm are given in Fig. 15. Taking 00:00 on 6 June as the starting time, the rainfall intensity amplified dramatically from 30 h to 34 h, with a maximum 4-h rainfall amount of 373 mm at gauge N19 located on the western Lantau Island. The total number of landslides predicted by the statistical rainfall-landslide correlations exceeds 15 at 30 h and reaches a maximum value of 162 around 44 h.

With the machine learning-based landslide forecasting model, the spatio-temporal evolution of landslides in the June 2008 storm is presented in Fig. 16. Hardly any landslide occurs before time t = 30 h. The western Lantau Island is the first area to witness many landslides at t = 32 h. Afterwards, landslides rapidly outbreak on the north-western Hong Kong Island and western Kowloon. The shifting of landslide areas is due to the compound effect of the concentration of man-made slopes and the movement of the storm centre. The final predicted landslide distribution at t = 96 h is consistent with the observations as shown in Fig. 16, with a total of 191 predicted landslides, including 64 very minor, 98 minor, 23 major, and 6 very major landslides, which also agree well with the observed man-made slope failures in SIS. Nearly 70% of the landslides are triggered within the most critical four hours (i.e., t = 30-34 h). This highlights the importance of real-time landslide forecasting. A landslide warning was issued by GEO and HKO at t = 25 h, successfully reducing potential losses to landslides.

This case study demonstrates the excellent performance of the landslide forecasting model in predicting the real-time spatio-temporal evolution of landslides to save adequate time for the public to protect themselves from impending landslides and for the responsible authorities to take necessary actions. Landslides occurring in densely populated urban regions (e.g., Hong Kong Island and Kowloon) and sparsely populated remote regions (e.g., Lantau Island) pose different levels of risk. To manage landslide risk better, it is necessary to upgrade the prevailing landslide warning criteria based on the total number of landslides to new criteria based on the landslide risk, with proper consideration of elements at risk and landslide consequences in space. The spatio-temporal landslide forecasting model reported in this paper is the first to provide such information in a unified machine learning framework.

Accepted manuscript doi:
10.1680/jgeot.21.00160

DISCUSSIONS

Importance ranking of influence factors

To facilitate an improved understanding of landslide mechanisms, it is essential to rank the importance of the 43 learning features considered in machine learning (Table 1). The permutation importance technique is favoured to evaluate the feature importance without model retraining but may underestimate the importance of strongly correlated features (e.g., rainfall features). Alternately, a group-based permutation importance technique is proposed, which classifies learning features into several groups according to their physical meanings, such as slope property (i.e., SF1-AR), rolling rainfall (i.e., R1h-R48h), and antecedent rainfall (i.e., A2d-A15d). The group importance, defined as the relative reduction of model score, e.g., correlation coefficient introduced in the previous section, is then evaluated when one specific group is randomly shuffled. Figure 17 presents the permutation importance of the three groups. The rolling rainfall group is far more critical than the other two groups. This is not surprising as the correlation coefficient is a storm-based index but all features in the slope property group are static. These static features are useful to identify landslides spatially in one storm but insufficient to differentiate landslides temporally in different storms.

Within the most influential rolling rainfall group, Fig. 17 further explores the impact of the number of rolling rainfall features on the model performance. The landslide forecasting model is repeatedly trained with only a part of rolling rainfall features (excluding other features for simplicity). Ten of all possible combinations are randomly selected for cases with two, four, and six rolling rainfall features. When the number of rolling rainfall features increases from one to nine, the model performance significantly improves from about 0.7 to 0.9 on average but with a decreasing rate. Further improvement by adding more rolling rainfall features (e.g., 5-, 6- and 7-h, or 72- and 96-h) becomes limited owing to the high

correlation among the rainfall features. Particularly for the case with only one rainfall feature, an apparent boundary exists between the short-duration (i.e., ≤ 8-h) and long-duration (i.e., ≥ 12-h) rainfall features. Features with longer durations on both sides (e.g., R8h and R48h) have greater influences because they inherently contain the information from features with shorter durations (e.g., R1h and R12h). The maximum rolling 24-h rainfall performs second best, supporting GEO's choice to develop statistical correlations based on it. The boundary implies different failure mechanisms of man-made slope under short- and long-duration rainstorms (i.e., shallow failure and deep failure) and highlights the need to utilise both short- and long-duration rainfall features to improve landslide forecasting, which advocates machine learning techniques again.

Suggestions when applied to data-scarce regions

The machine learning framework proposed in this study can be easily applied to other landslide-prone regions worldwide for developing region-specific, spatio-temporal landslide forecasting models. A major obstacle may be the scarcity of data. Compared with the relatively easy-to-obtain rainfall and geographic data, databases for existing slopes and landslide incidents are not always available. However, it is never too late to compile databases in the era of big data. Based on the experience of Hong Kong, the spatial location of slopes and the failure time of landslides are the top two critical data in developing a spatio-temporal landslide forecasting model, and hence should be given higher priority in data collection. The model will gradually converge when more and more data are fed in and after the slopes concerned have suffered several severe rainstorms. Additionally, state-of-the-art machine learning techniques such as transfer learning (e.g., Yang et al., 2020) may provide tools to deal with the data scarcity problem with the help of a pre-trained source model of Hong Kong presented in this paper.

SUMMARY AND CONCLUSIONS

Based on comprehensive databases of man-made slopes, rainstorms, and landslides in the past 35 years, an advanced spatio-temporal landslide forecasting model was developed within a unified machine learning framework to predict the probability, scale, and spatial distribution of rain-induced man-made slope failures over time. The major conclusions are as follows:

(1) With a storm-based data integration strategy and multiclass classification on landslide scales, the proposed machine learning framework incorporates landslide time and consequences into conventional landslide susceptibility mapping to successfully achieve spatio-temporal landslide forecasting.

(2) The machine learning-based landslide forecasting model significantly outperforms the prevailing statistical rainfall-landslide correlations, not only in the prediction accuracy but also in the ability to predict both landslide locations and consequences. Among six candidate machine learning models, logistic regression is the best for landslide forecasting.

(3) Validated against historical landslide incidents both temporally and spatially and through a case study of the June 2008 storm, the machine learning-based landslide forecasting model demonstrates excellent performance in predicting the real-time spatio-temporal evolution of landslides, which provides a powerful tool for state-of-the-art landslide risk assessment and early warning.

(4) The rolling rainfall features are critical factors governing slope stability. It is necessary to utilise both short- and long-duration rainfall features to consider different failure mechanisms of man-made slopes.

Accepted manuscript doi:
10.1680/jgeot.21.00160

Acknowledgements

The authors acknowledge the financial support from the Research Grants Council of the Hong Kong SAR (Project Nos. 16205719, 16203720, and AoE/E-603/18) and the data support from the Geotechnical Engineering Office of the Civil Engineering and Development Department, the Government of the Hong Kong Special Administrative Region. This paper is published with the permission of the Head of the Geotechnical Engineering Office and the Director of the Civil Engineering and Development Department.

APPENDIX 1. FEATURE EXTRACTION IN DATASET

Table 2 provides the sources of data for this study, most of which are open to the public. As shown in Table 1, man-made features, formation time, slope materials, slope covers, slope characteristics, and wall characteristics are directly read from the man-made slope database in Fig. 1. The annual rainfall amounts at locations of man-made slopes and landslide scales are extracted from Figs. 2 and 3, respectively. With respect to the storm-related rainfall features, Fig. 18 lists the 419 major rainstorms according to GEO's criterion, i.e., daily rainfall exceeding 50 mm before 2011 or 100 mm after 2012. To characterise the temporal patterns of a rainstorm in both short- and long-duration (e.g., Gao et al., 2018) and the antecedent rainfall (e.g., Rahardjo et al., 2020), the maximum rolling 1-, 2-, 4-, 8-, 12-, 18-, 24-, 36-, and 48-h rainfall amounts and the cumulative 2-, 4-, 7-, 10-, and 15-day antecedent rainfall amounts are evaluated. High positive correlations exist among these rolling and antecedent rainfall features. Rainfall features at locations of man-made slopes are spatially interpolated from the rain gauges using the inverse distance weighting approach.

In addition to slope, rainfall, and landslide databases, geographic databases of terrain (Fig. 19(a)), geology (Fig. 19(b)), land cover (Fig. 19(c)), and infrastructure (Fig. 19(d)) are also utilised. All feature values are extracted at the locations of man-made slopes as parts of

slope properties. The terrain information, geological conditions, and land cover information are supplements to slope characteristics, materials, and covers, respectively, to avoid oversimplification. Additionally, distances of man-made slopes to faults, waterbodies (e.g., streams, reservoirs, and the sea), roads, and buildings are evaluated as four location features. The distance to faults or waterbodies reflects the potential danger of slopes, while the distance to roads or buildings indirectly reflects the protection level of slopes as man-made slopes near roads or buildings are more likely to be well protected using engineering measures like soil nails and wire meshes.

Last but not least, all numerical variables are scaled to a similar magnitude, and the qualitative features described by categorical variables are converted to binary vectors through a dummy encoding strategy, as shown in Table 1. Taking the geological classification as an example, the four geological conditions are converted to three variables: [1, 0, 0] for granitic rock, [0, 1, 0] for volcanic rock, [0, 0, 1] for superficial deposit, and [0, 0, 0] for others.

APPENDIX 2. COMPARISON OF CANDIDATE MACHINE LEARNING MODELS

This appendix compares the performances of six candidate machine learning models (i.e., logistic regression, neural network, Bagging, AdaBoost, RUSBoost, and the subspace method), as well as the statistical correlations (Xiao and Zhang, 2020), in binary landslide classification (i.e., non-landslide or landslide) without considering the landslide scale. Among the six methods, RUSBoost and the subspace method are particularly designed for class-imbalanced problems. Bagging, AdaBoost, RUSBoost, and the subspace method are ensemble learning methods (e.g., Zhou, 2012) that combine several simple base models to produce one optimal predictive model. The first three take decision trees as weak learners, while the subspace method uses linear discriminant analysis. For simplicity, structures of all the models follow commonly-used options without tuning of hyperparameters; for example,

the neural network has one hidden layer with ten neurons, and the maximal number of branch node splits per tree is 20 in AdaBoost and RUSBoost and 10,000 in Bagging. Five machine learning methods except for RUSBoost can provide posterior probability as a classification score, treated as the failure probability of man-made slopes in a storm.

The storm-based integrated dataset is randomly split by a ratio of 70/30 into training and test sets (i.e., 15 million records with 1489 landslide records in the training set and 7 million records with 638 landslide records in the test set). All six models are trained on the same training set, and their performances on both training and test sets are evaluated. Figure 20(a) demonstrates the receiver operating characteristic curves for the six models. Except for Bagging, the AUC values of all machine learning models are similar, falling in a narrow range from 0.93 to 0.97. This indicates the similar high ability of the machine learning models to distinguish between classes. The AUC of Bagging approaches 1.00 (i.e., perfect classification) for the training set but dramatically falls to only 0.64 for the test set, which means the Bagging model overfits the training data and is unsuitable for such a class-imbalanced problem.

Figure 21 presents the confusion matrices of classification for the six models, in which the left and right matrices are for the training and test sets, respectively, and the four numbers denote the numbers of true negative (TN), false positive (FP), false negative (FN), and true positive (TP), respectively. The logistic regression, neural network, and AdaBoost models cannot identify landslide records well, classifying almost all records into non-landslides. In contrast, RUSBoost and the subspace method can successfully identify 82% and 13% of landslide records, respectively, but at the cost of classifying many non-landslide records into landslides. Such a result is expected for class-imbalanced problems. Numerous false positives (i.e., predicting a landslide when it is not) will trigger a large number of false

alerts in a landslide early warning system, which is unacceptable. Hence, although RUSBoost and the subspace method are designed for general class-imbalanced problems, they may not be suitable for landslide forecasting. In the context that many methods cannot identify landslide records, the posterior probability (i.e., the probability of landslide occurrence) is more useful than the predicted label of classification (i.e., landslide or non-landslide). The summation of landslide probabilities can represent the expected number of landslides over a certain time or area. Note that the extremely high class-imbalance makes many model performance indices derived from the confusion matrix unsatisfactory; for example, the accuracy, defined as (TN+TP)/(TN+TP+FN+FP), of the logistic regression is as high as 1.00, simply because the number of true negatives is far greater than others.

Figure 22 compares the observed and predicted numbers of landslides (i.e., summing up the predicted failure probabilities of all man-made slopes in the same storm) in all 419 rainstorms. The correlation coefficients of logistic regression, neural network, Bagging, AdaBoost, RUSBoost, and the subspace method are 0.92, 0.88, 0.99, 0.87, 0.76, and 0.85, respectively. For reference, the correlation coefficient of statistical correlations is 0.80. Most machine learning models perform better than statistical correlations because much more features are used in machine learning. Bagging is out of consideration due to the overfitting issue, and a significant overestimation of the number of landslides is observed when using RUSBoost and the subspace method. By comparison, logistic regression is the best landslide forecasting model among the six adopted machine learning models.

Notation

AUC	area under the receiver operating characteristic curve
n_s	number of failure scales
N	number of landslides
P	probability of landslide occurrence
$\boldsymbol{W}$	diagonal weight matrix
$\boldsymbol{x}$	vector of learning features, such as slope information and rainfall
$\boldsymbol{X}$	matrix form of learning features

y	landslide feature, such as failure scales of landslides
ρ	correlation coefficient between the observed and predicted numbers of landslides
$\boldsymbol{\theta}$	vector of model parameters
$\boldsymbol{\Theta}$	ensemble of all model parameters

References

Au, S. W. C. (1998). Rain-induced slope instability in Hong Kong. *Eng. Geol.* **51**, No. 1, 1−36.

Baum, R. L., Godt, J. W. & Savage, W. Z. (2010). Estimating the timing and location of shallow rainfall-induced landslides using a model for transient, unsaturated infiltration. *J. Geophys. Res. Earth Surf.* **115**, No. F3, F03013.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Singapore: Springer.

Bui, D. T., Tuan, T. A., Klempe, H., Pradhan, B. & Revhaug, I. (2016). Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides* **13**, No. 2, 361−378.

Chen, H. X. & Zhang, L. M. (2014). A physically-based distributed cell model for predicting regional rainfall-induced shallow slope failures. *Eng. Geol.* **176**, 79−92.

Cheung, R. W. M. (2021). Landslide risk management in Hong Kong. *Landslides* **18**, No. 10, 3457–3473.

Cheung, R. W. M. & Tang, W. H. (2005). Realistic assessment of slope reliability for effective landslide hazard management. *Géotechnique* **55**, No. 1, 85−94.

Ching, J., Liao, H. J. & Lee, J. Y. (2011). Predicting rainfall-induced landslide potential along a mountain road in Taiwan. *Géotechnique* **61**, No. 2, 153−166.

Corominas, J., van Westen, C., Frattini, P., Cascini, L., Malet, J. P., Fotopoulou, S., Catani, F., Van Den Eeckhaut, M., Mavrouli, O., Agliardi, F., Pitilakis, K., Winter, M. G., Pastor,

M., Ferlisi, S., Tofani, V., Hervas, J. & Smith, J. T. (2014). Recommendations for the quantitative analysis of landslide risk. *Bull. Eng. Geol. Environ.* **73**, No. 2, 209−263.

Crosta, G. B. & Frattini, P. (2003). Distributed modelling of shallow landslides triggered by intense rainfall. *Nat. Hazards Earth Syst. Sci.* **3**, 81−93.

Dai, F. C. & Lee, C. F. (2001). Frequency–volume relation and prediction of rainfall-induced landslides. *Eng. Geol.* **59**, No. 3-4, 253−266.

Dai, F. C. & Lee, C. F. (2002). Landslide characteristics and slope instability modeling using GIS, Lantau Island, Hong Kong. *Geomorphology* **42**, No. 3-4, 213−228.

Dai, F. C., Lee, C. F. & Ngai, Y. Y. (2002). Landslide risk assessment and management: an overview. *Eng. Geol.* **64**, No. 1, 65−87.

De Gast, T., Hicks, M. A., Van Den Eijnden, A. P. & Vardon, P. J. (2021). On the reliability assessment of a controlled dyke failure. *Géotechnique* **71**, No. 11, 1028–1043.

Dobson, A. J. & Barnett, A. G. (2018). *An Introduction to Generalized Linear Models (4th edn)*. Boca Raton: CRC Press.

El-Ramly, H., Morgenstern, N. R. & Cruden, D. M. (2005). Probabilistic assessment of stability of a cut slope in residual soil. *Géotechnique* **55**, No. 1, 77−84.

Fell, R. (1994). Landslide risk assessment and acceptable risk. *Can. Geotech. J.* **31**, No. 2, 261−272.

Frattini, P., Crosta, G. & Carrara, A. (2010). Techniques for evaluating the performance of landslide susceptibility models. *Eng. Geol.* **111**, No. 1-4, 62−72.

Gao, L., Zhang, L. M. & Cheung, R. W. M. (2018). Relationships between natural terrain landslide magnitudes and triggering rainfall based on a large landslide inventory in Hong Kong. *Landslides* **15**, No. 4, 727−740.

Ghorbanzadeh, O., Blaschke, T., Gholamnia, K., Meena, S. R., Tiede, D. & Aryal, J. (2019). Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection. *Remote Sens.* **11**, No. 2, 196.

Goetz, J. N., Brenning, A., Petschko, H. & Leopold, P. (2015). Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. *Comput. Geosci.* **81**, 1−11.

Gong, P., Liu, H., Zhang, M., et al. (2019). Stable classification with limited sample: transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017. *Sci. Bull.* **64**, No. 6, 370−373.

Guzzetti, F., Gariano, S. L., Peruccacci, S., Brunetti, M. T., Marchesini, I., Rossi, M. & Melillo, M. (2020). Geographical landslide early warning systems. *Earth-Sci. Rev.* **200**, 102973.

Guzzetti, F., Peruccacci, S., Rossi, M. & Stark, C. P. (2007). Rainfall thresholds for the initiation of landslides in central and southern Europe. *Meteorol. Atmos. Phys.* **98**, No. 3-4, 239−267.

Hicks, M. A., Varkey, D., van den Eijnden, A. P., de Gast, T. & Vardon, P. J. (2019). On characteristic values and the reliability-based assessment of dykes. *Georisk* **13**, No. 4, 313−319.

Huang, F., Zhang, J., Zhou, C., Wang, Y., Huang, J. & Zhu, L. (2020). A deep learning algorithm using a fully connected sparse autoencoder neural network for landslide susceptibility prediction. *Landslides* **17**, No. 1, 217−229.

Ko, F. W. Y. & Lo, F. L. C. (2016). Rainfall-based landslide susceptibility analysis for natural terrain in Hong Kong − A direct stock-taking approach. *Eng. Geol.* **215**, 95−107.

Kong, V. W. W., Kwan, J. S. H. & Pun, W. K. (2020). Hong Kong's landslip warning system − 40 years of progress. *Landslides* **17**, No. 6, 1453−1463.

Krkač, M., Špoljarić, D., Bernat, S. & Arbanas, S. M. (2017). Method for prediction of landslide movements based on random forests. *Landslides* **14**, No. 3, 947−960.

Lacasse, S. & Nadim, F. (2009). Landslide risk assessment and mitigation strategy. In *Landslides − Disaster Risk Reduction*, pp. 31−61. Berlin: Springer.

Lam, C. L. H., Lau, J. W. C. & Chan, H. W. (2012). *Factual report on Hong Kong rainfall and landslides in 2008*. GEO Report No. 273. Hong Kong: Geotechnical Engineering Office.

Li, D. Q., Xiao, T., Cao, Z. J., Zhou, C. B. & Zhang, L. M. (2016). Enhancement of random finite element method in reliability analysis and risk assessment of soil slopes using Subset Simulation. *Landslides* **13**, No. 2, 293−303.

Lian, C., Zeng, Z., Yao, W. & Tang, H. (2014). Ensemble of extreme learning machine for landslide displacement prediction based on time series analysis. *Neural Comput. Appl.* **24**, No. 1, 99−107.

Lloret-Cabot, M., Fenton, G. A. & Hicks, M. A. (2014). On the estimation of scale of fluctuation in geostatistics. *Georisk* **8**, No. 2, 129–140.

Lombardo, L., Opitz, T., Ardizzone, F., Guzzetti, F. & Huser, R. (2020). Space-time landslide predictive modelling. *Earth-Sci. Rev.* **209**, 103318.

Merghadi, A., Yunus, A. P., Dou, J., Whiteley, J., ThaiPham, B., Bui, D. T., Avtari, R. & Abderrahmane, B. (2020). Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. *Earth-Sci. Rev.* **207**, 103225.

Piciullo, L., Calvello, M. & Cepeda, J. M. (2018). Territorial early warning systems for rainfall-induced landslides. *Earth-Sci. Rev.* **179**, 228−247.

Qiang, Y. J., Zhang, L. M. & Xiao, T. (2020). Spatial-temporal rain field generation for the Guangdong-Hong Kong-Macau Greater Bay Area considering climate change. *J. Hydrol.* **583**, 124584.

Rahardjo, H., Nistor, M. M., Gofar, N., Satyanaga, A., Qin, X. & Yee, S. I. C. (2020). Spatial distribution, variation and trend of five-day antecedent rainfall in Singapore. *Georisk* **14**, No. 3, 177−191.

Segoni, S., Piciullo, L. & Gariano, S. L. (2018). A review of the recent literature on rainfall thresholds for landslide occurrence. *Landslides* **15**, No. 8, 1483−1501.

Shen, P., Zhang, L., Chen, H. & Fan, R. (2018). EDDA 2.0: integrated simulation of debris flow initiation and dynamics considering two initiation mechanisms. *Geosci. Model Dev.* **11**, No. 7, 2841−2856.

Stumpf, A. & Kerle, N. (2011). Object-oriented mapping of landslides using Random Forests. *Remote Sens. Environ.* **115**, No. 10, 2564−2577.

Su, Z., Chow, J. K., Tan, P. S., Wu, J., Ho, Y. K. & Wang, Y. H. (2021). Deep convolutional neural network–based pixel-wise landslide inventory mapping. *Landslides* **18**, No. 4, 1421−1443.

Van Den Eeckhaut, M., Hervás, J., Jaedicke, C., Malet, J. P., Montanarella, L. & Nadim, F. (2012). Statistical modelling of Europe-wide landslide susceptibility using limited landslide inventory data. *Landslides* **9**, No. 3, 357−369.

Wang, H., Zhang, L., Luo, H., He, J. & Cheung, R. W. M. (2021a). AI-powered landslide susceptibility assessment in Hong Kong. *Eng. Geol.* **288**, 106103.

Wang, H., Zhang, L., Yin, K., Luo, H. & Li, J. (2021b). Landslide identification using machine learning. *Geosci. Front* **12**, No. 1, 351−364.

Wang, Y. & Zhao, T. Y. (2017). Statistical interpretation of soil property profiles from sparse data using Bayesian compressive sampling. *Géotechnique* **67**, No. 6, 523−536.

Wong, A. C. W., Ting, S. M., Shiu, Y. K. & Ho, K. K. S. (2014). Latest developments of Hong Kong's landslip warning system. In *Landslide Science for a Safer Geoenvironment, Vol. 2: Methods of Landslide Studies* (ed. Sassa, K., et al.), pp. 613−618. Switzerland: Springer.

Xiao, T., Li, D. Q., Cao, Z. J., Au, S. K. & Phoon, K. K. (2016). Three-dimensional slope reliability and risk assessment using auxiliary random finite element method. *Comput. Geotech.* **79**, 146−158.

Xiao, T., Li, D. Q., Cao, Z. J. & Zhang, L. M. (2018). CPT-based probabilistic characterization of three-dimensional spatial variability using MLE. *J. Geotech. Geoenviron. Eng.* **144**, No. 5, 04018023.

Xiao, T. & Zhang, L. M. (2020). Evaluation of performance of engineered slopes under extreme rainstorms. In *Geo-Congress 2020: Engineering, Monitoring, and Management of Geotechnical Infrastructure*, GSP No. 316, pp. 737−743. Reston: American Society of Civil Engineers.

Xiao, T., Zou, H. F., Yin, K. S., Du, Y. & Zhang, L. M. (2021). Machine learning-enhanced soil classification by integrating borehole and CPTU data with noise filtering. *Bull. Eng. Geol. Environ.* **80**, No. 12, 9157−9171.

Yang, B., Yin, K., Lacasse, S. & Liu, Z. (2019). Time series analysis and long short-term memory neural network to predict landslide displacement. *Landslides* **16**, No. 4, 677−694.

Yang, Q., Zhang, Y., Dai, W. & Pan, S. J. (2020). *Transfer Learning*. Cambridge: Cambridge University Press.

Yu, Y. F., Lam, J. S., Siu, C. K. & Pun, W. K. (2004). Recent advance in landslip warning system. In *Proceedings of the 1-day Seminar on Recent Advances in Geotechnical Engineering*. pp. 139−147. Hong Kong: Hong Kong Institution of Engineers.

Yuan, J., Papaioannou, I. & Straub, D. (2019). Probabilistic failure analysis of infinite slopes under random rainfall processes and spatially variable soil. *Georisk* **13**, No. 1, 20−33.

Zhou, C., Yin, K., Cao, Y., Intrieri, E., Ahmed, B. & Catani, F. (2018). Displacement prediction of step-like landslide by applying a novel kernel extreme learning machine method. *Landslides* **15**, No. 11, 2211−2225.

Zhou, Z. H. (2012). *Ensemble Methods: Foundations and Algorithms*. Boca Raton: CRC Press.

Table 1. Features extracted from comprehensive databases

Feature category	ID	Variable	Meaning	Type [b]	Example	Note
Man-made feature	1	SF1	Fill slope	C, S	[1, 0]	A retaining wall could be an independent feature or associated with a fill/cut slope, and it is described by wall characteristics.
	2	SF2	Cut slope	C, S	[0, 1]	
	−	SF3 [a]	Retaining wall	C, S	[0, 0]	
Formation time	3	FT	Post-1978	C, S	1	Man-made slope features formed after 1978 correspond to a higher safety standard in
	−	FT2 [a]	Pre-1977	C, S	0	

Feature category	ID	Variable	Meaning	Type [b]	Example	Note
						Hong Kong.
Slope material	4	SM1	Soil	C, S	[1, 0]	[SM1, SM2] = [1, 1] if a slope comprises both soil and rock materials.
	5	SM2	Rock	C, S	[0, 1]	
Slope cover	6	SC1	Percent of vegetation (%)	N, S	60	Slope covers, such as vegetation, help to improve slope stability.
	7	SC2	Percent of chunam, shotcrete or other covers (%)	N, S	30	
	–	SC3 [a]	Percent of bare surface (%)	N, S	10	
Slope characteristics	8	SH	Slope height (m)	N, S	20	All variables of slope characteristics are zeros for an independent retaining wall without any slope.
	9	SA	Slope angle (°)	N, S	30	
	10	SB	Berms in slope	C, S	1	
	11	SW	Weep holes in slope	C, S	1	
	12	SD	Number of drainages in slope	N, S	5	
Wall characteristics	13	WH	Wall height (m)	N, S	10	All variables of wall characteristics are zeros for a man-made slope not protected by a retaining wall.
	14	WA	Wall angle (°)	N, S	90	
	15	WB	Berms in wall	C, S	0	
	16	WW	Weep holes in wall	C, S	1	
	17	WD	Number of drainages in wall	N, S	2	
Terrain	18	TE	Terrain elevation (m)	N, S	400	The terrain information is a
	1	TI	Terrain inclination (°)	N,	30	

Feature category	ID	Variable	Meaning	Type [b]	Example	Note
	9			S		supplement to slope characteristics.
Geological condition	20	GEO1	Granitic rock	C, S	[1, 0, 0]	The geological condition is a supplement to slope materials.
	21	GEO2	Volcanic rock	C, S	[0, 1, 0]	
	22	GEO3	Superficial deposit	C, S	[0, 0, 1]	
	–	GEO4 [a]	Other geological conditions	C, S	[0, 0, 0]	
Land cover	23	LC1	Cropland, shrub or grass	C, S	[1, 0]	The land cover information is a supplement to slope covers.
	24	LC2	Forest	C, S	[0, 1]	
	–	LC3 [a]	Other land covers	C, S	[0, 0]	
Location	25	DF	Distance to faults (0.1 km)	N, S	10	The distance to faults or waterbodies reflects the potential danger of slopes, while the distance to roads or buildings indirectly reflects the protection level of slopes.
	26	DW	Distance to waterbodies (0.1 km)	N, S	5	
	27	DR	Distance to roads (0.1 km)	N, S	1	
	28	DB	Distance to buildings (0.1 km)	N, S	2	
Annual rainfall	29	AR	Annual rainfall amount (mm*12 months)	N, S	150	The annual rainfall amount is related to the average groundwater table.
Rolling rainfall	3	R1h	Maximum rolling 1-h rainfall amount	N,	60	The

Feature category	ID	Variable	Meaning	Type [b]	Example	Note
	0		(mm)	D		maximum rolling rainfall amounts represent the temporal rainfall process in a storm.
	31	R2h	Maximum rolling 2-h rainfall amount (mm)	N, D	90	
	32	R4h	Maximum rolling 4-h rainfall amount (mm)	N, D	140	
	33	R8h	Maximum rolling 8-h rainfall amount (mm)	N, D	200	
	34	R12h	Maximum rolling 12-h rainfall amount (mm)	N, D	300	
	35	R18h	Maximum rolling 18-h rainfall amount (mm)	N, D	330	
	36	R24h	Maximum rolling 24-h rainfall amount (mm)	N, D	350	
	37	R36h	Maximum rolling 36-h rainfall amount (mm)	N, D	400	
	38	R48h	Maximum rolling 48-h rainfall amount (mm)	N, D	400	
Antecedent rainfall	39	A2d	Cumulative 2-day antecedent rainfall amount (mm)	N, D	0	The antecedent rainfall before a storm may raise the groundwater table.
	40	A4d	Cumulative 4-day antecedent rainfall amount (mm)	N, D	50	
	41	A7d	Cumulative 7-day antecedent rainfall amount (mm)	N, D	50	
	42	A10d	Cumulative 10-day antecedent rainfall amount (mm)	N, D	150	
	43	A15d	Cumulative 15-day antecedent rainfall amount (mm)	N, D	250	
Landslide scale	44	LS	No failure	C, D	0	If only two slope stability conditions are considered, LS = 0 and 1 stand for non-landslide and landslide, respectively.
	44	LS	Very minor failure (landslide volume $V < 5$ m^3)	C, D	1	
	44	LS	Minor failure (5 m^3 $\leq V <$ 50 m^3)	C, D	2	
	44	LS	Major failure (50 m^3 $\leq V <$ 500 m^3)	C, D	3	
	44	LS	Very major failure ($V \geq$ 500 m^3)	C, D	4	

Note: [a]hidden feature; [b]N, C, S, and D for numerical, categorical, static, and dynamic variables, respectively.

Accepted manuscript doi:
10.1680/jgeot.21.00160

Table 2. Sources of databases

Database	Figure	Type	Source
Slope	Fig. 1	Polygon	GEO's Slope Information System (https://hkss.cedd.gov.hk/hkss/en/facts-and-figures/slope-information-system/sis/index.html; https://www.geomap.cedd.gov.hk/GEOOpenData/eng/Feature.aspx)
Rainfall	Fig. 2	Point + time series	Rain gauges of GEO (https://www.geomap.cedd.gov.hk/GEOOpenData/eng/Raingauge.aspx) and HKO (https://www.hko.gov.hk/en/cis/climat.htm)
Landslide	Fig. 3	Point	GEO's annual reports on rainfall and landslides (https://www.cedd.gov.hk/eng/publications/geo/geo-reports/index.html) and list of landslide incidents (https://www.geomap.cedd.gov.hk/GEOOpenData/eng/Incident.aspx)
Terrain	Fig. 19(a)	Raster	Survey and Mapping Office (https://www.landsd.gov.hk/en/spatial-data/open-data.html)
Geology	Fig. 19(b)	Polygon	GEO's geological maps (https://www.cedd.gov.hk/eng/about-us/organisation/geo/pub_info/memoirs/geology/index.html; https://www.geomap.cedd.gov.hk/GEOOpenData/eng/GeologicalMap.aspx)
Land cover	Fig. 19(c)	Raster	Planning Department's land utilization map (https://www.pland.gov.hk/pland_en/info_serv/open_data/landu/index.html) or FROM-GLC10 (Gong et al., 2019; used in this study) (http://data.ess.tsinghua.edu.cn)
Infrastructure	Fig. 19(d)	Polygon	Civil Engineering and Development Department (private communication; used in this study) or OpenStreetMap (https://www.openstreetmap.org)
Grid	Fig. 12	−	Survey and Mapping Office (https://www.landsd.gov.hk/doc/en/mapping/paper-map/index/index_1.pdf)

Figure captions

Fig. 1. Registered man-made slopes in Hong Kong
Fig. 2. Distribution of rain gauges and mean annual rainfall isohyets (1990-2010)
Fig. 3. Reported landslide incidents of man-made slope failures (1984-2017)
Fig. 4. Cumulative frequencies: (a) geometry of slope/wall; (b) landslide volume
Fig. 5. The 419 major rainstorm events and landslide incidents in the period of 1984-2017
Fig. 6. An example of landslide incident (modified after Lam et al., 2012)
Fig. 7. Statistical rainfall-landslide correlations
Fig. 8. Machine learning framework for spatio-temporal landslide forecasting
Fig. 9. Data integration: (a) traditional manner; (b) storm-based manner
Fig. 10. Statistics of rainfall features in a storm-based integrated dataset (notation in Table 1)
Fig. 11. Normalised model parameters in multinomial logistic regression (notation in Table 1)
Fig. 12. Spatial distribution of man-made slopes in a grid (grid size: 750 m × 600 m)

Accepted manuscript doi:
10.1680/jgeot.21.00160

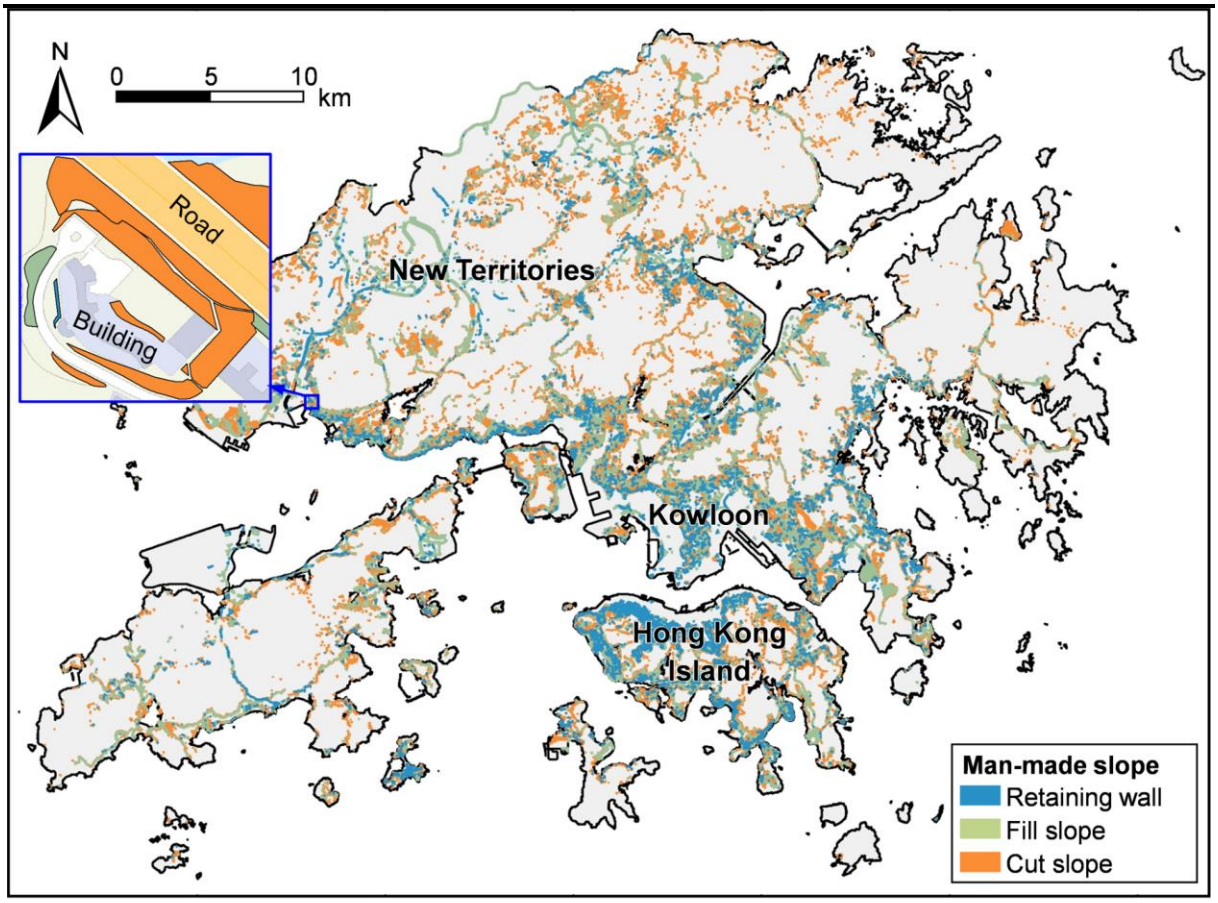

Fig. 1

Accepted manuscript doi:
10.1680/jgeot.21.00160

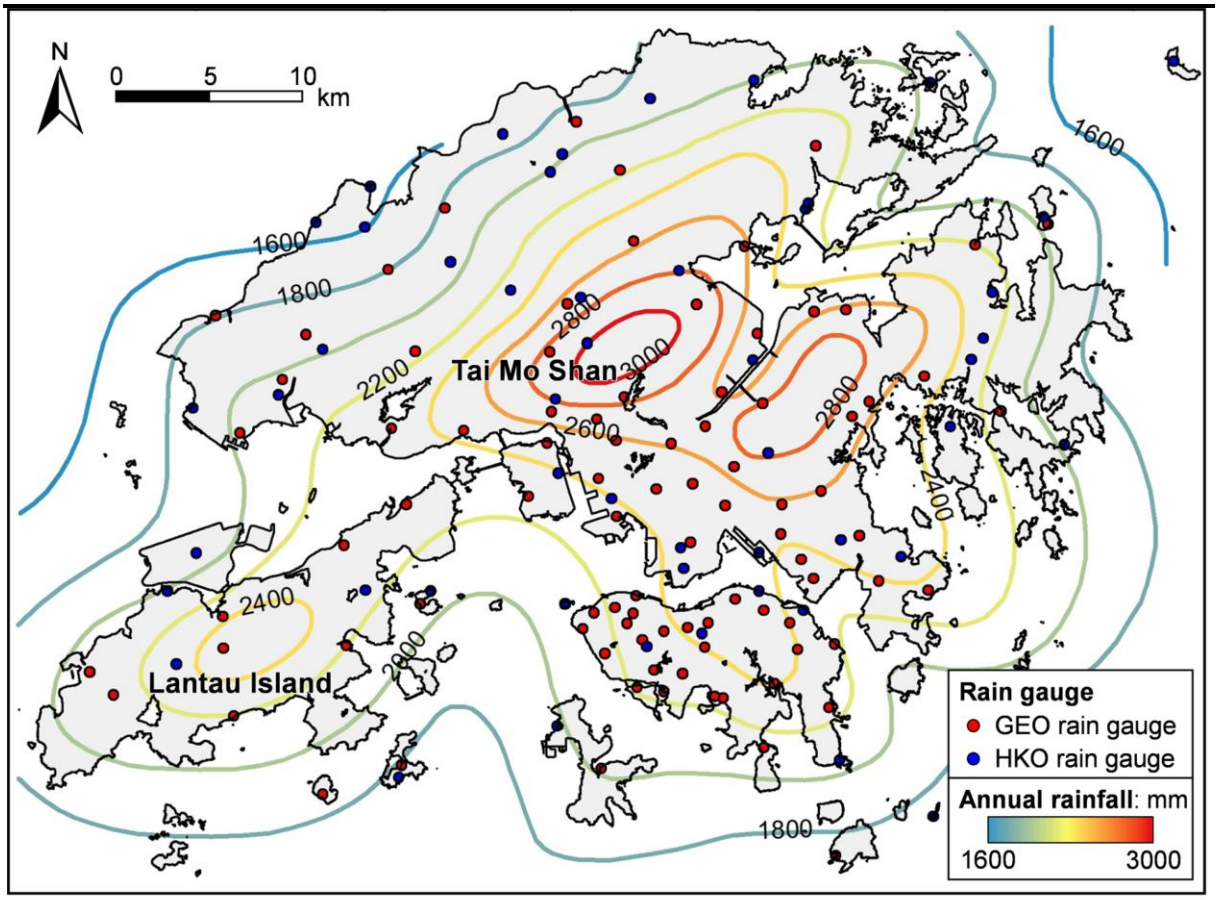

Fig. 2

Accepted manuscript doi:
10.1680/jgeot.21.00160

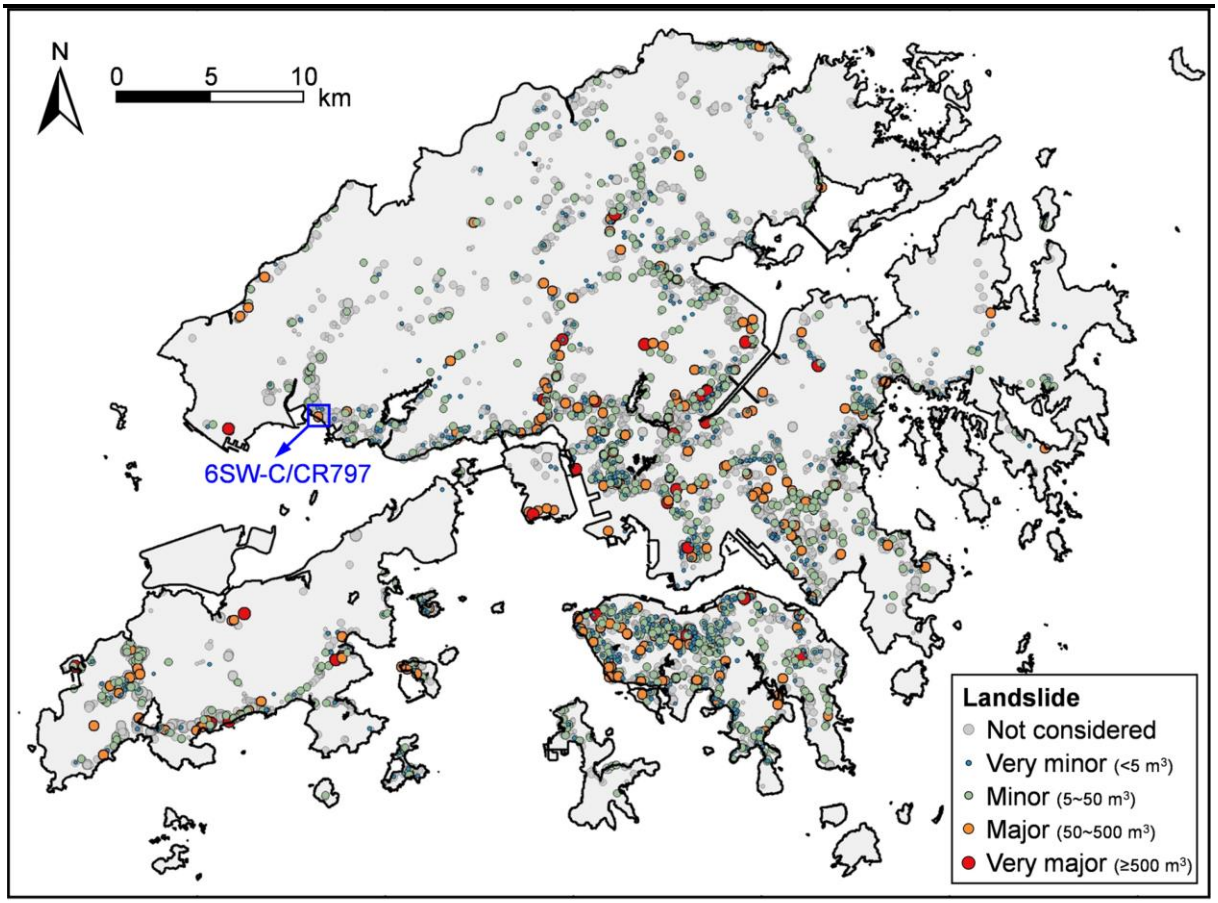

Fig. 3

Accepted manuscript doi:
10.1680/jgeot.21.00160

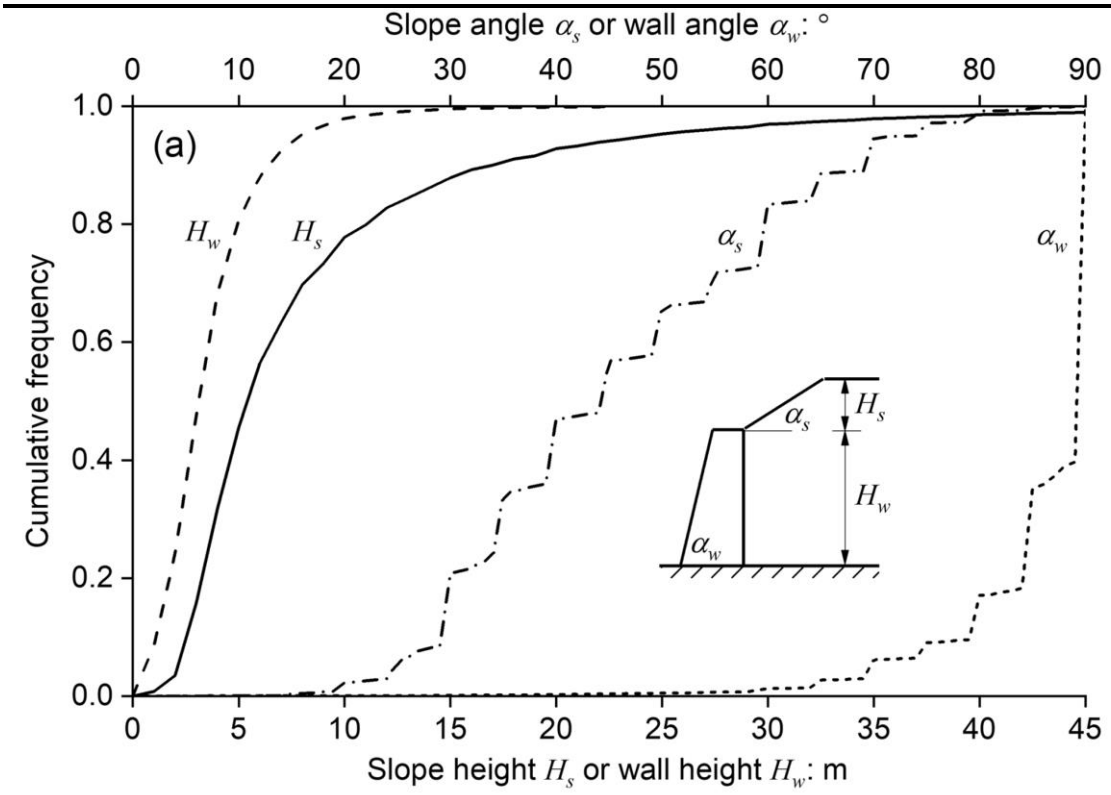

Fig. 4a

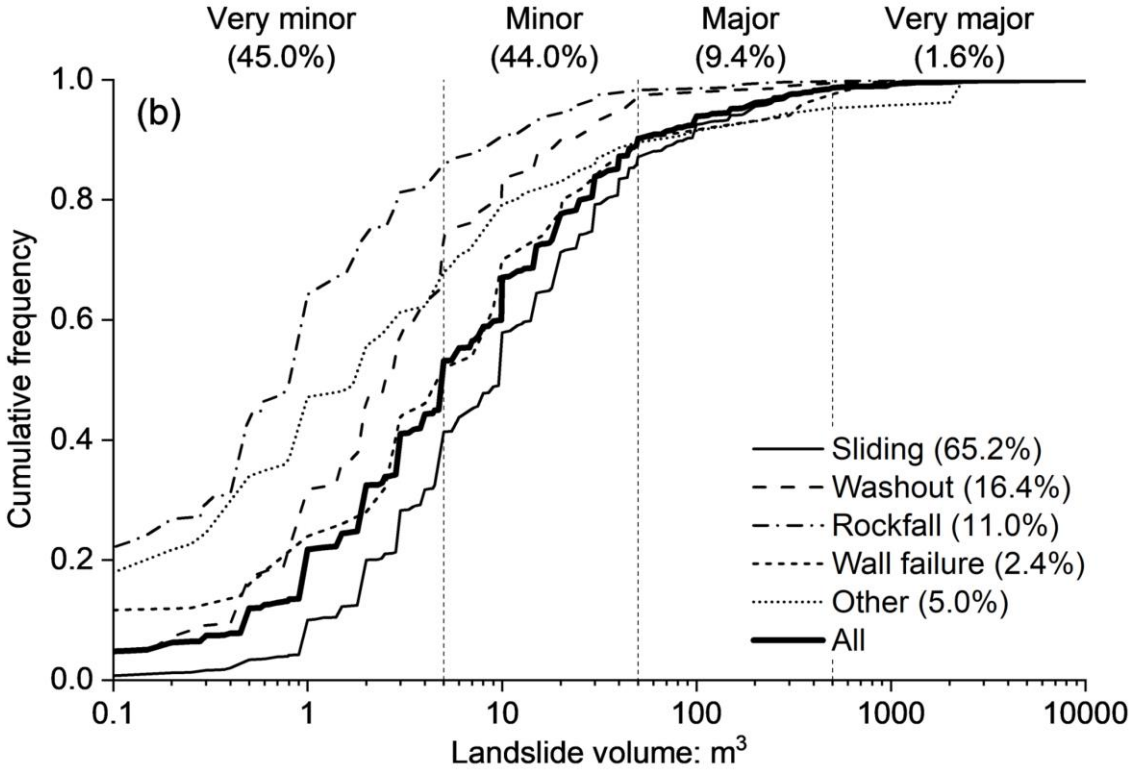

Fig. 4b

Accepted manuscript doi:
10.1680/jgeot.21.00160

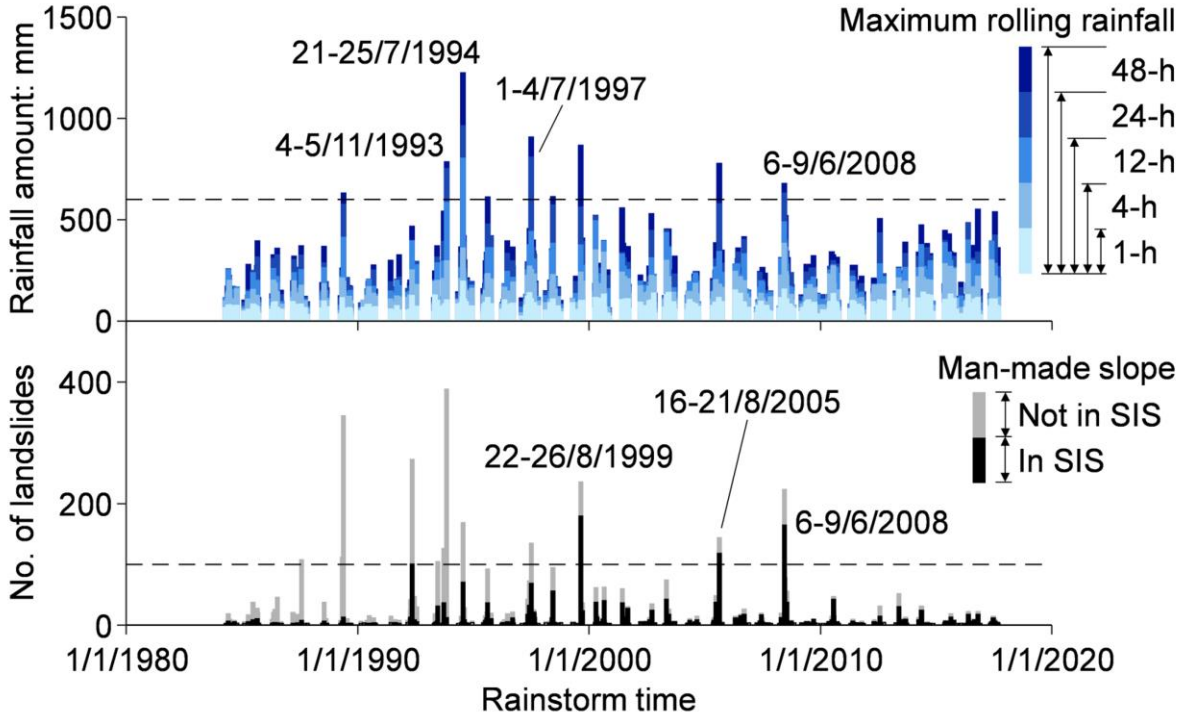

Fig. 5

Accepted manuscript doi:
10.1680/jgeot.21.00160

Incident ID	Location	Feature		Failure			Consequence
		ID	Type	Date	Type	Volume	
2008/06/0129	Hung Hing Store 18.5 milestone, Castle Peak Road, Cafeteria Old Beach, Tuen Mun	6SW-C/CR797	Retaining Wall	7/6/2008	Wall failure	300 m^3	2 fatalities, 12 persons and 4 squatter dwellings temporarily evacuated

Fig. 6-Revised

Accepted manuscript doi:
10.1680/jgeot.21.00160

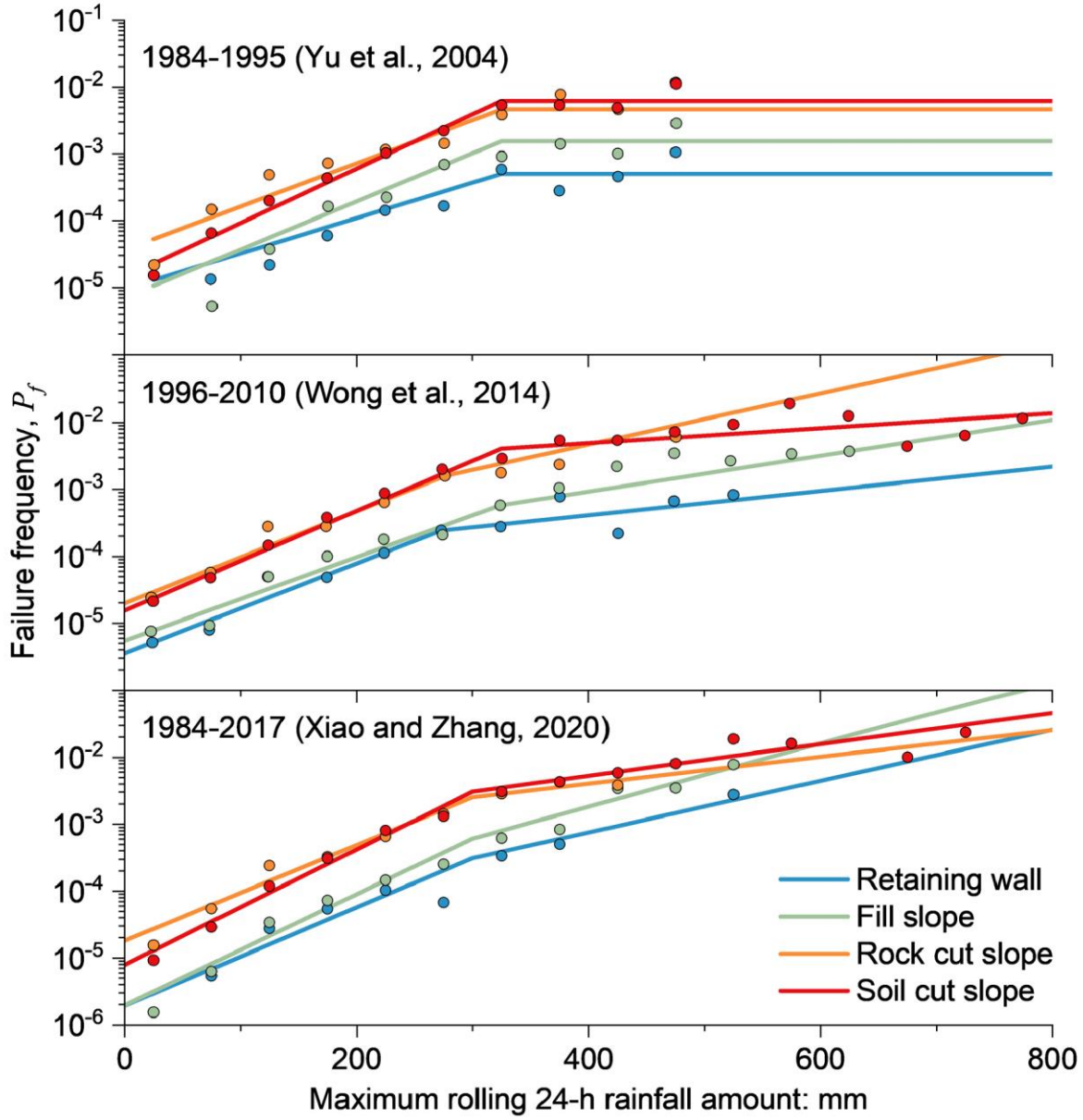

Fig. 7

Accepted manuscript doi:
10.1680/jgeot.21.00160

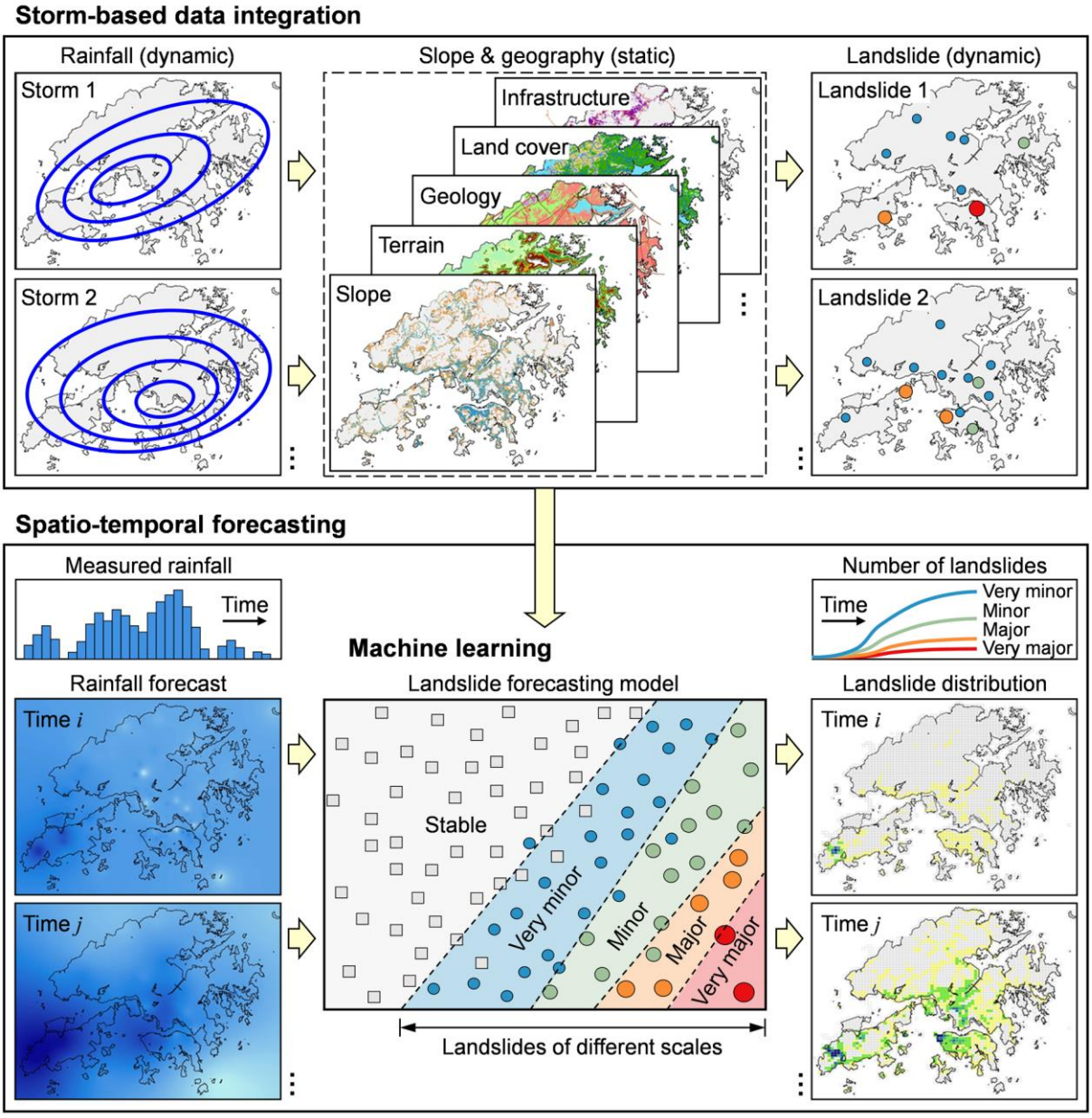

Fig. 8

Accepted manuscript doi:
10.1680/jgeot.21.00160

(a)	Failed (F)	Safe (S)
Record (R) 1	Slope 1, Storm 3	Randomly select the same number of safe examples
Record 2	Slope 3, Storm 2	
Record 3	Slope 3, Storm m	
Record 4	Slope n, Storm m	
…	…	

(b)	Storm 1	Storm 2	Storm 3	Storm 4	…	Storm m
Slope 1	S $_{R1,1}$	S $_{R2,1}$	**F** $_{R3,1}$	S $_{R4,1}$		S $_{Rm,1}$
Slope 2	S $_{R1,2}$	S $_{R2,2}$	S $_{R3,2}$	S $_{R4,2}$		S $_{Rm,2}$
Slope 3	S $_{R1,3}$	**F** $_{R2,3}$	S $_{R3,3}$	S $_{R4,3}$		**F** $_{Rm,3}$
Slope 4	N/A	N/A	S $_{R3,4}$	S $_{R4,4}$		S $_{Rm,4}$
…						
Slope n	N/A	N/A	N/A	S $_{R4,n}$		**F** $_{Rm,n}$

Fig. 9

Accepted manuscript doi:
10.1680/jgeot.21.00160

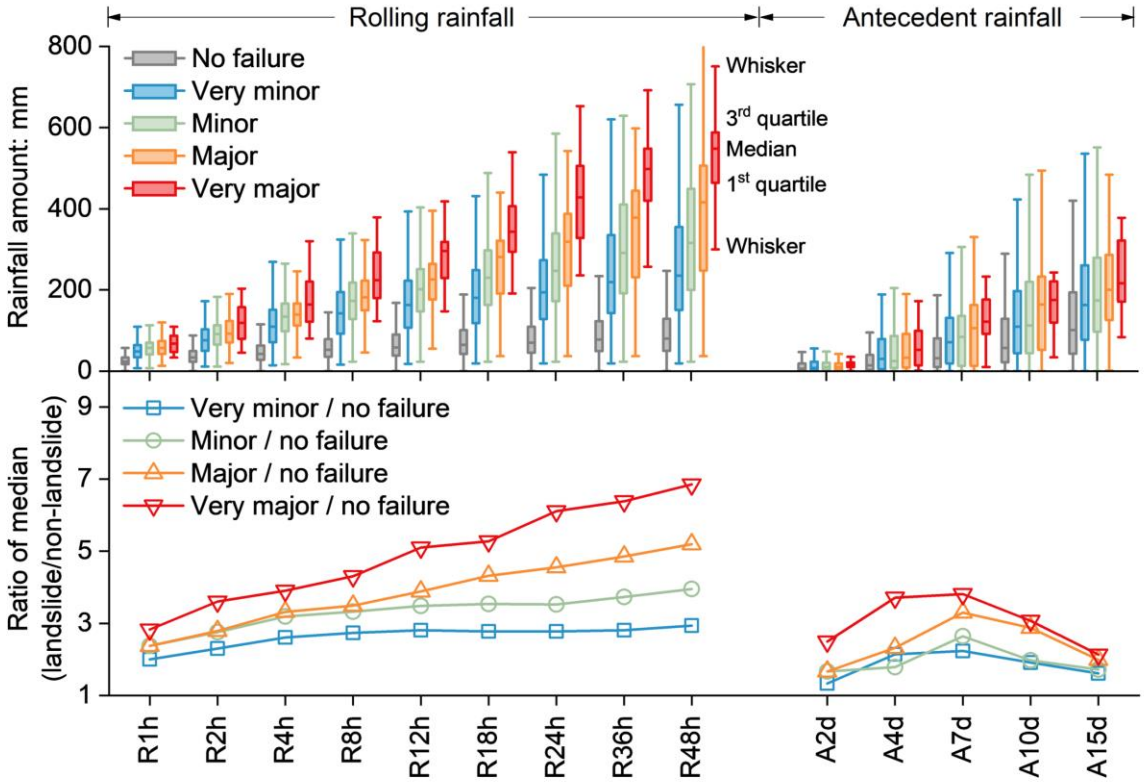

Fig. 10

Accepted manuscript doi:
10.1680/jgeot.21.00160

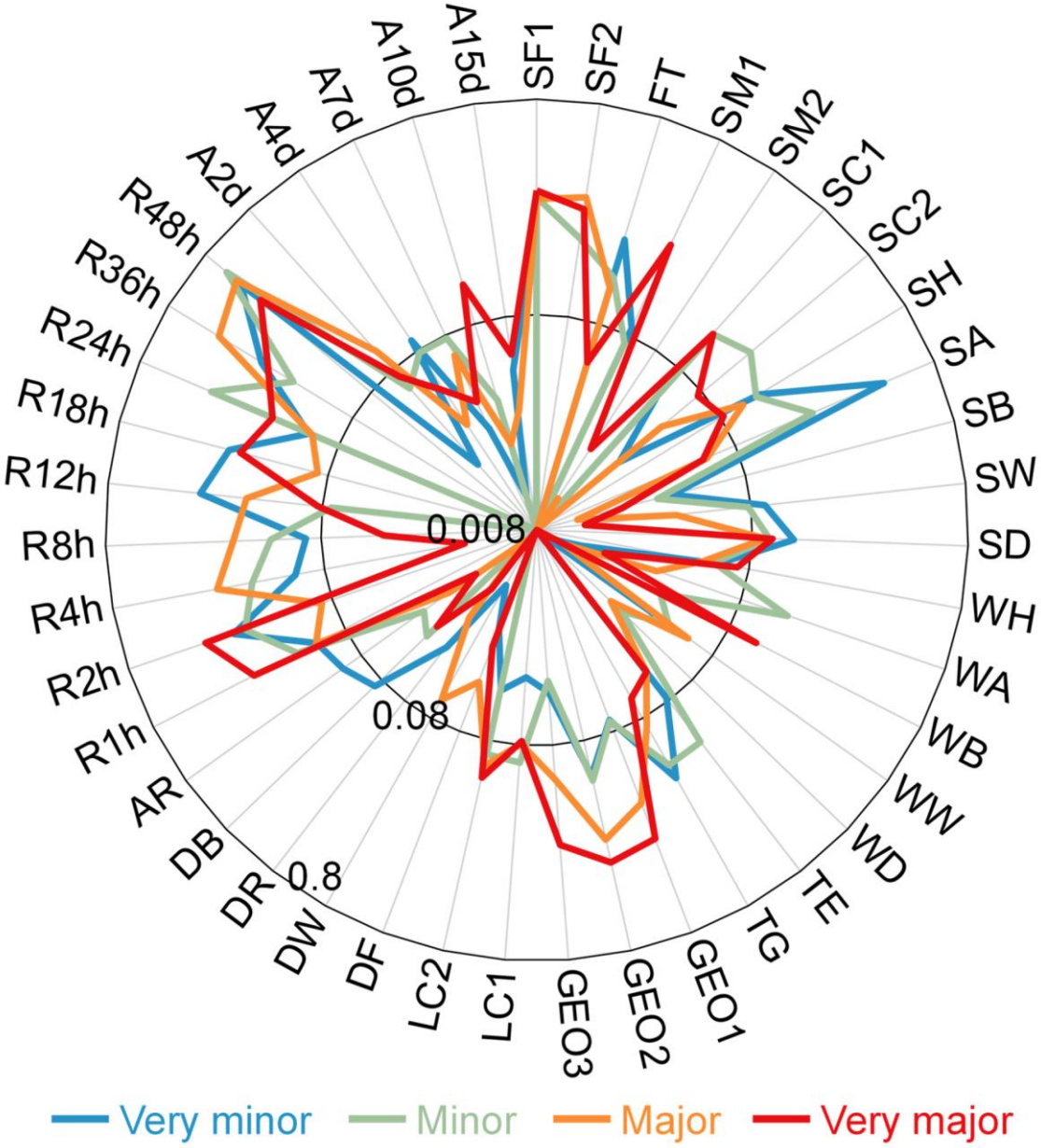

Fig. 11

Accepted manuscript doi:
10.1680/jgeot.21.00160

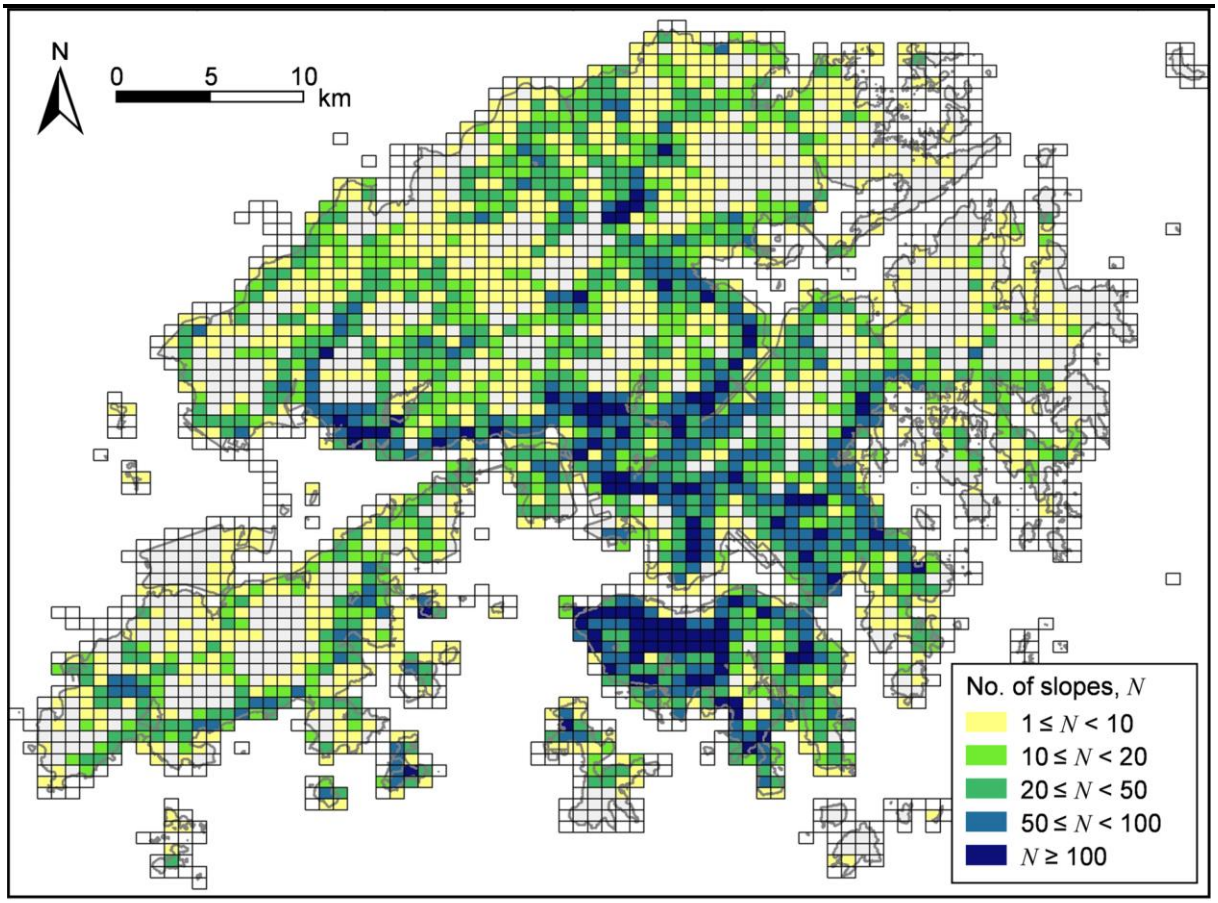

Fig. 12

Accepted manuscript doi:
10.1680/jgeot.21.00160

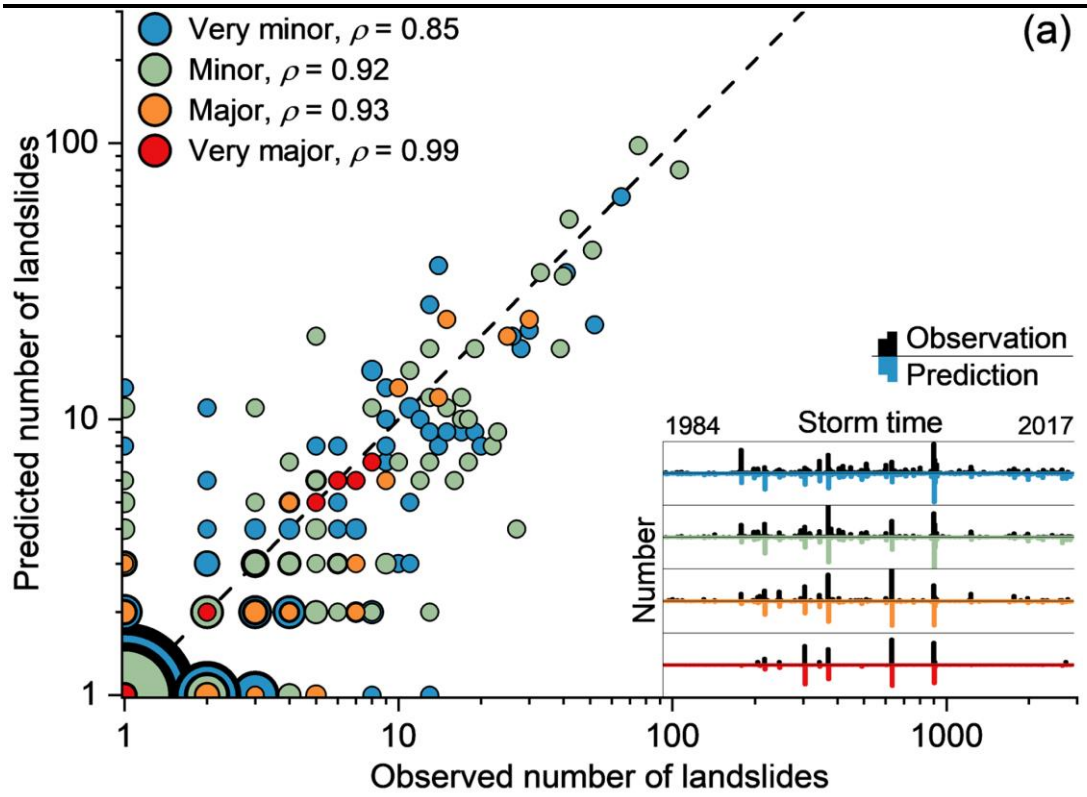

Fig. 13a

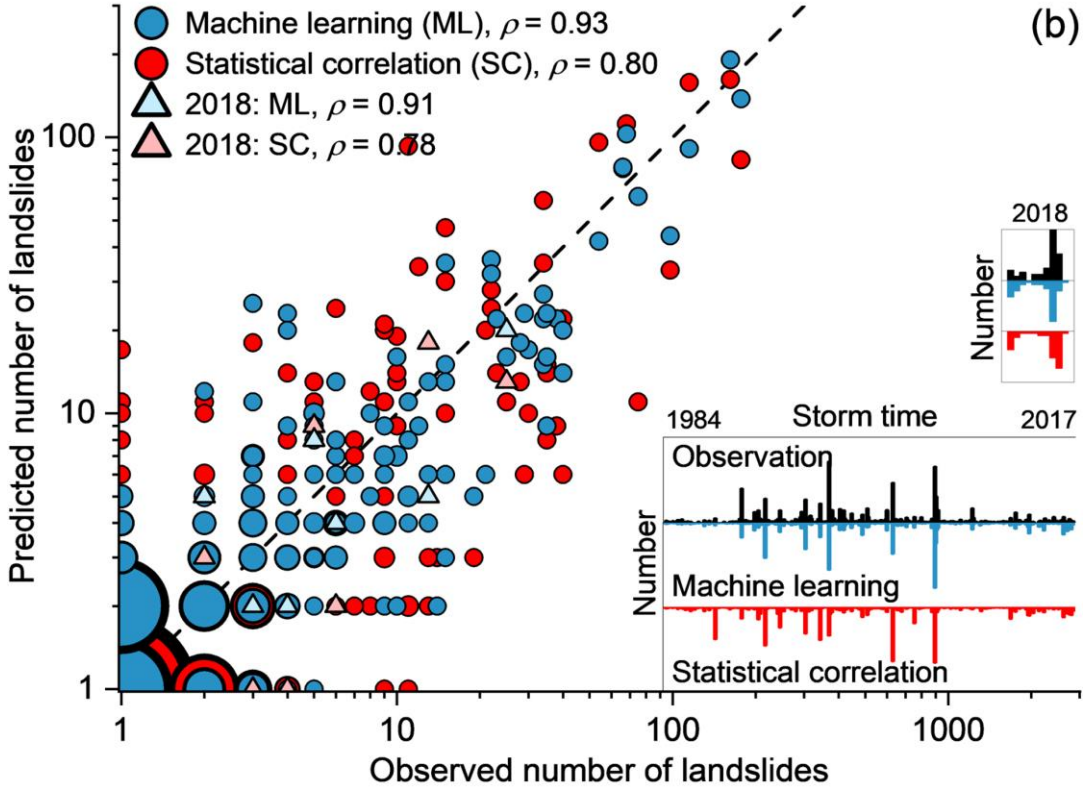

Fig. 13b

Accepted manuscript doi:
10.1680/jgeot.21.00160

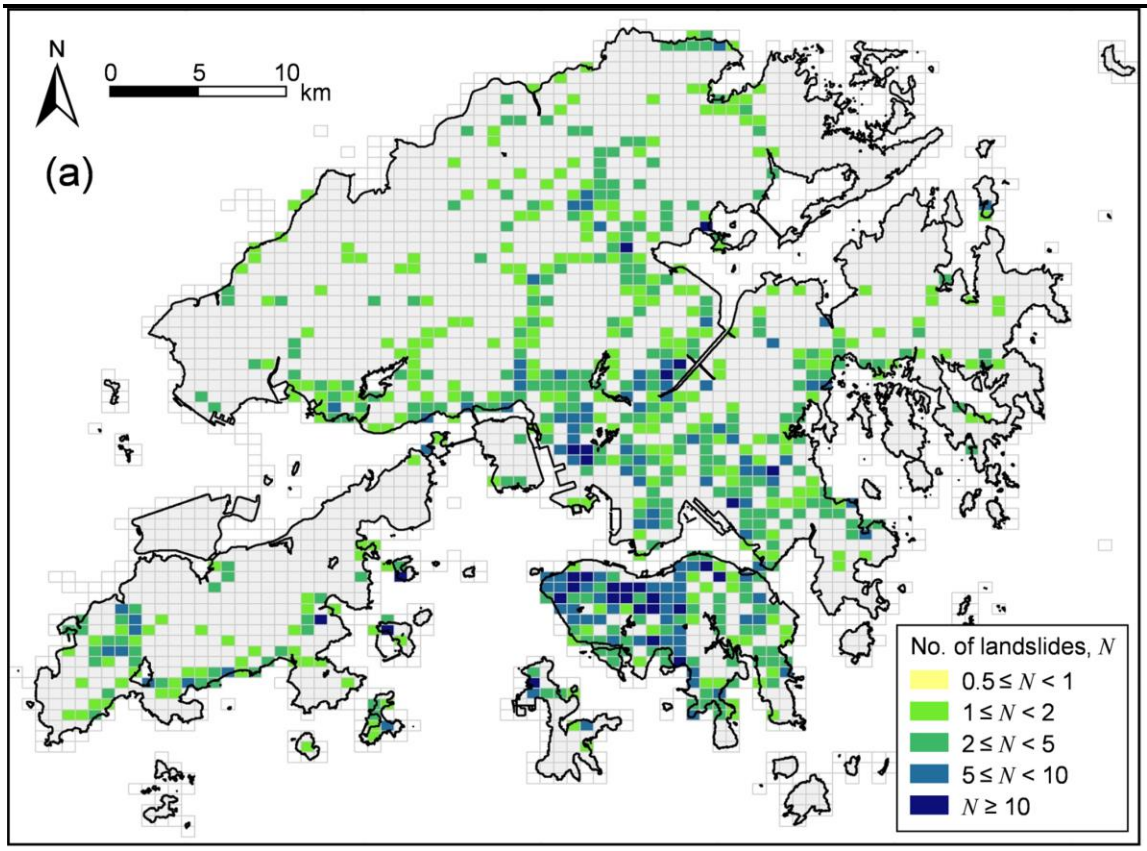

Fig. 14a

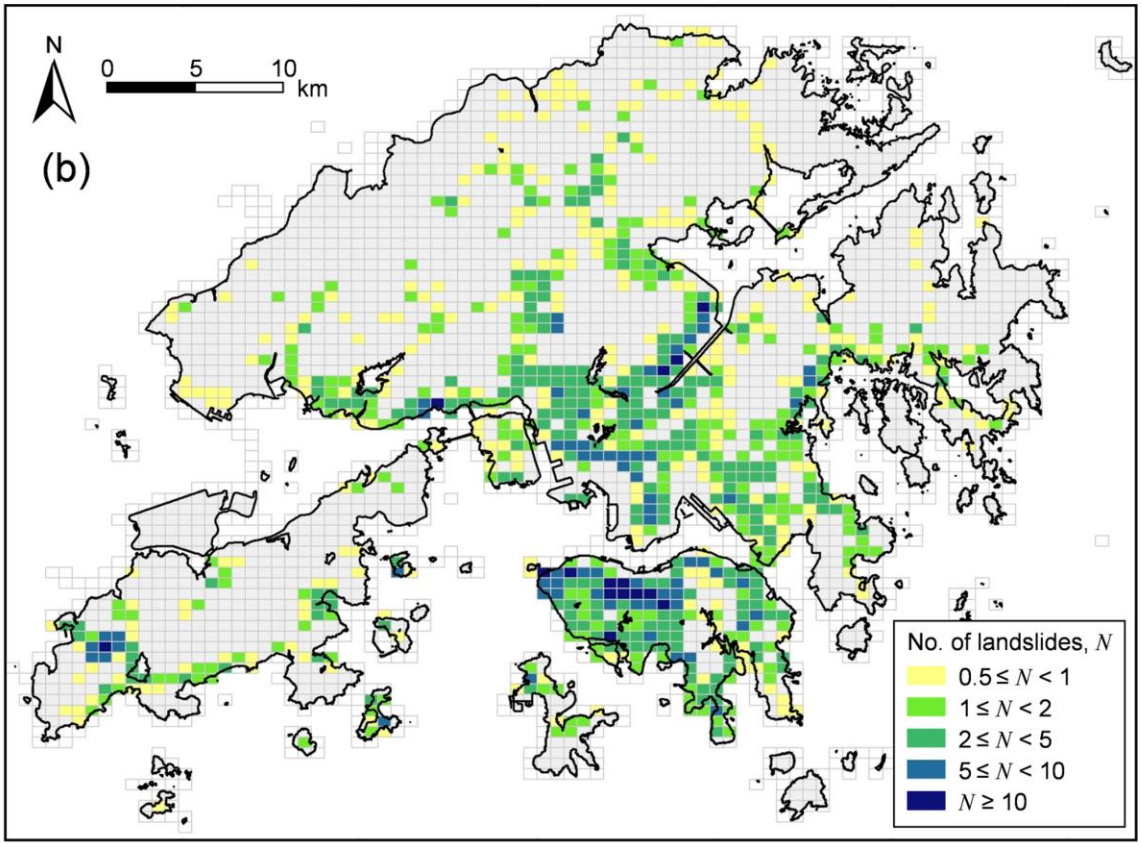

Fig. 14b

Accepted manuscript doi:
10.1680/jgeot.21.00160

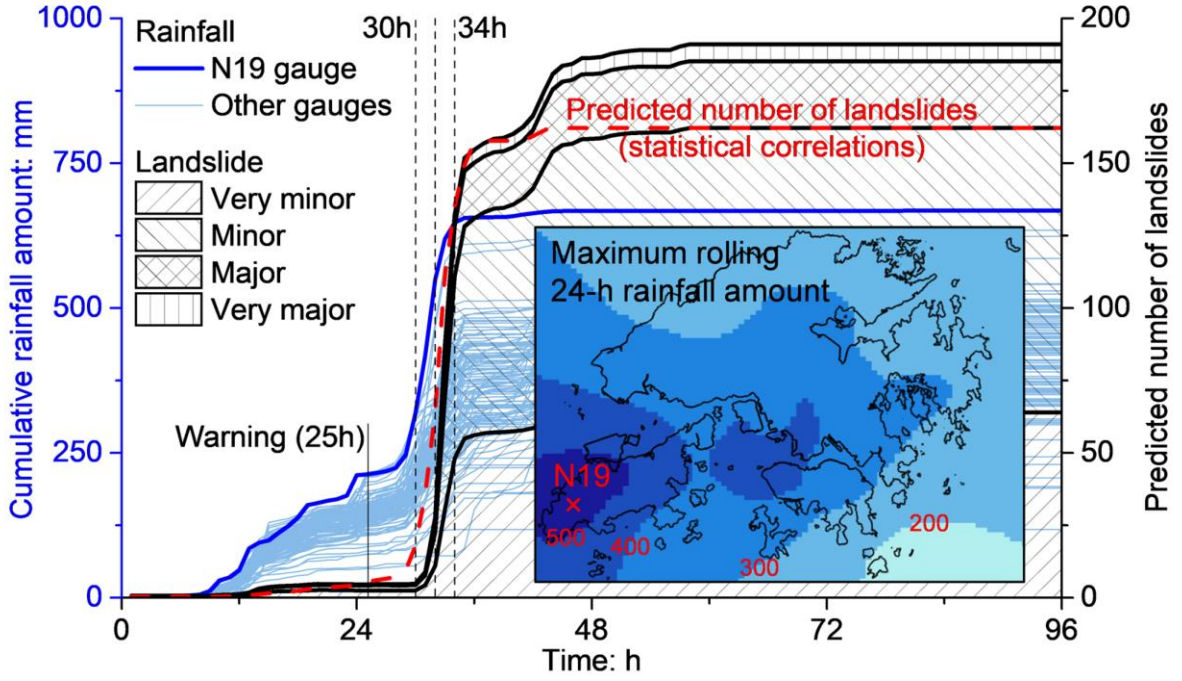

Fig. 15

Accepted manuscript doi:
10.1680/jgeot.21.00160

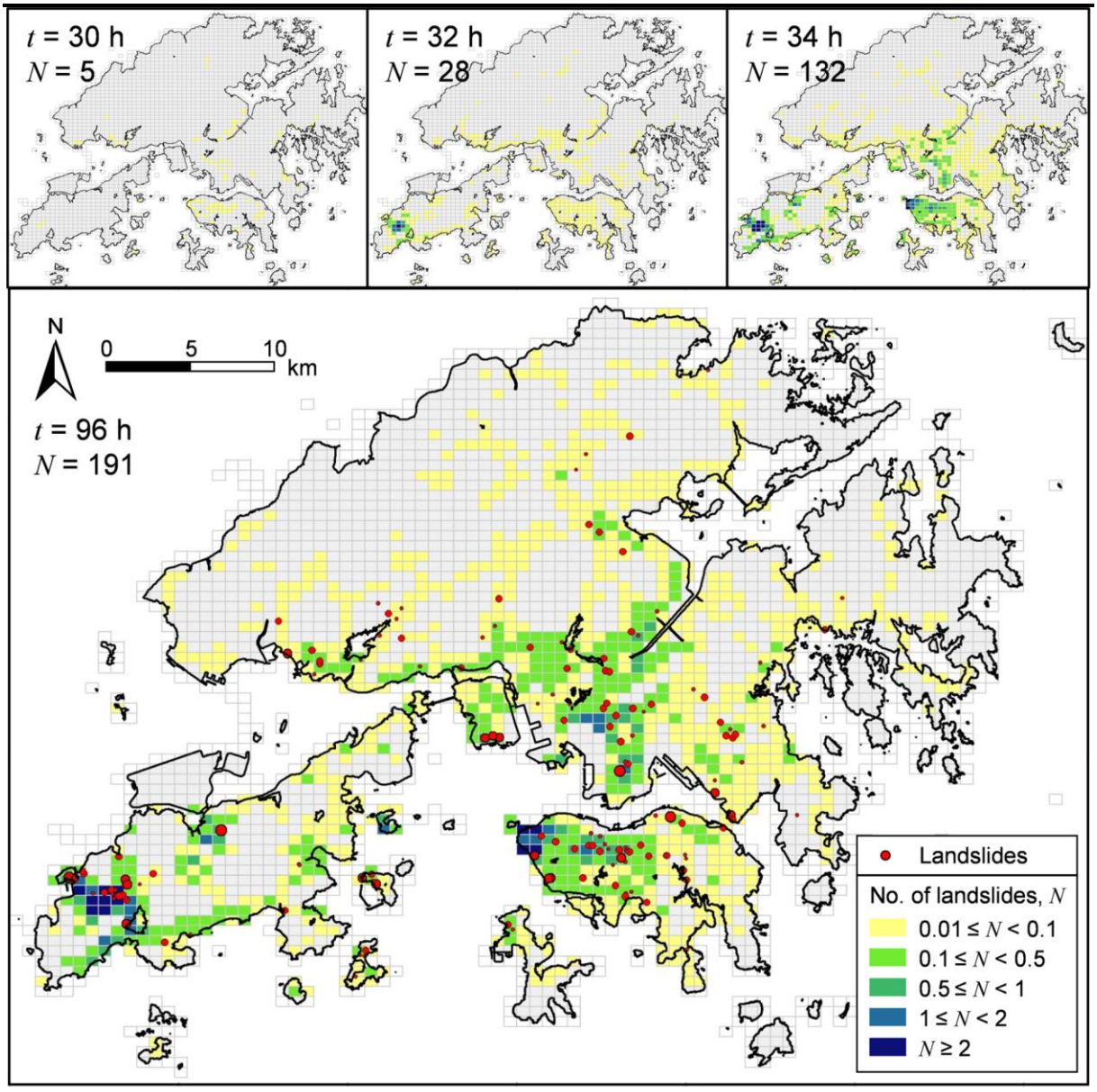

Fig. 16

Accepted manuscript doi:
10.1680/jgeot.21.00160

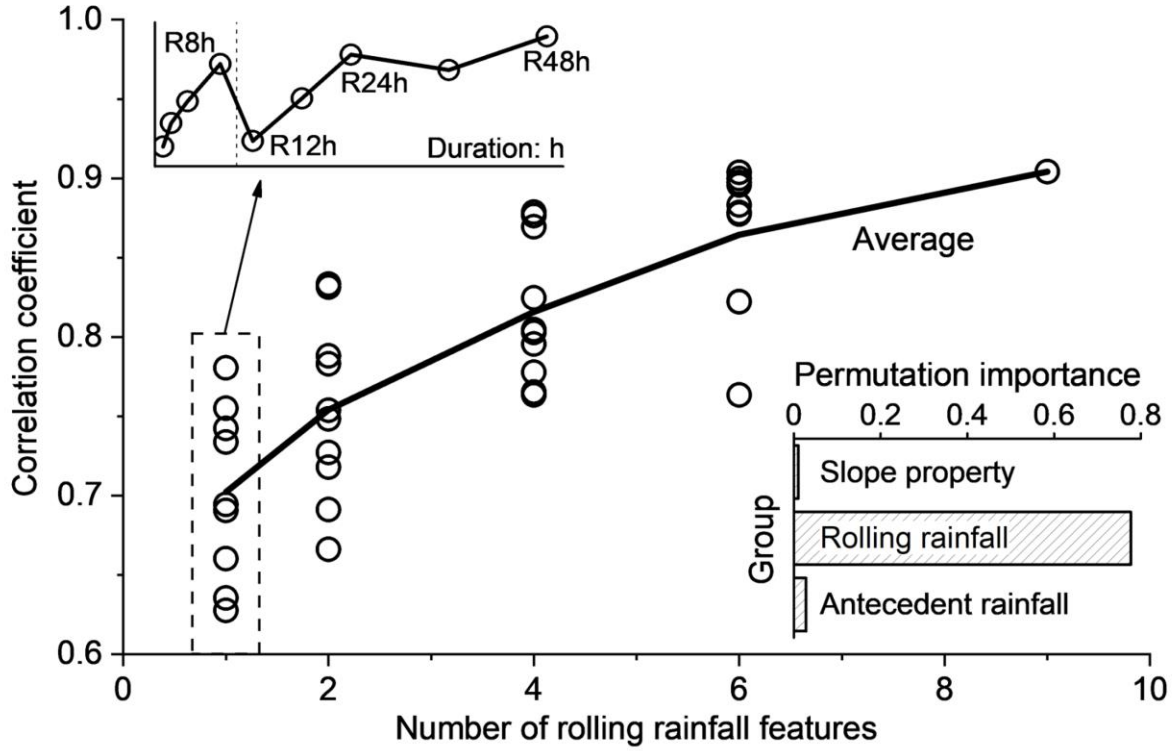

Fig. 17

Accepted manuscript doi:
10.1680/jgeot.21.00160

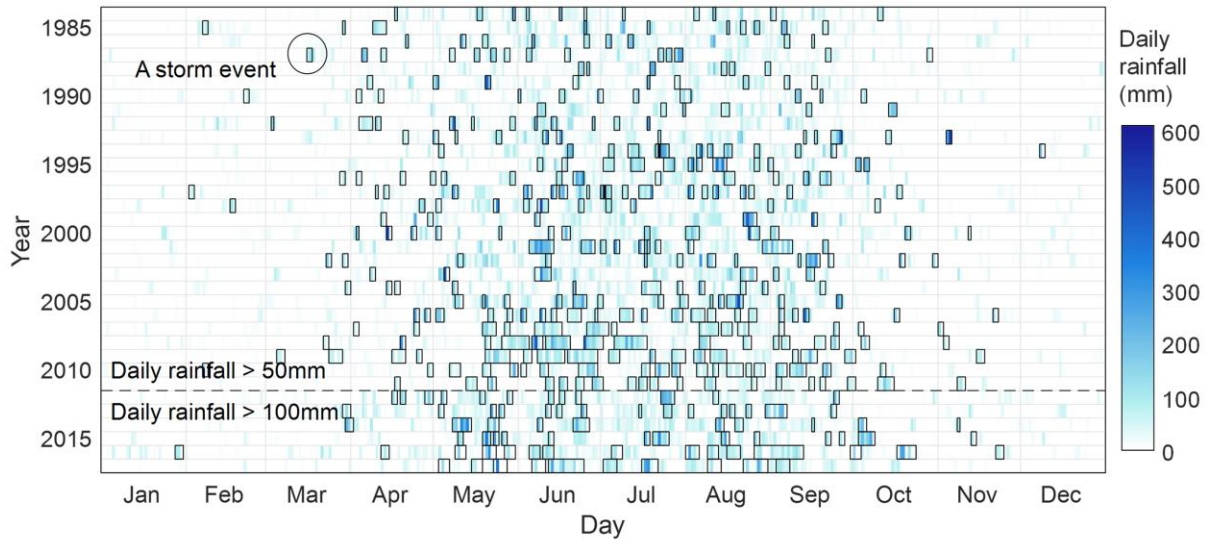

Fig. 18

Accepted manuscript doi:
10.1680/jgeot.21.00160

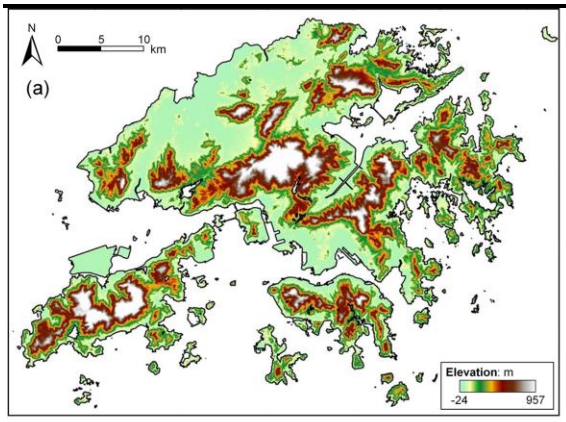

Fig. 19a

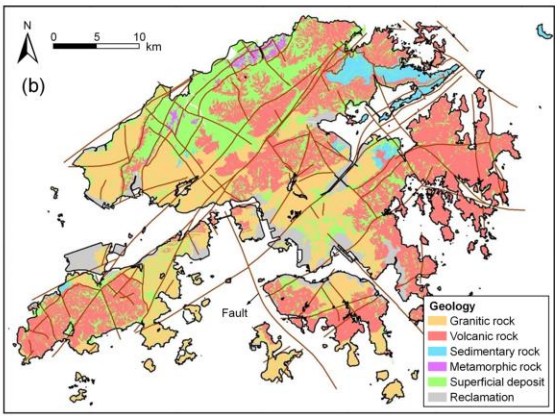

Fig. 19b

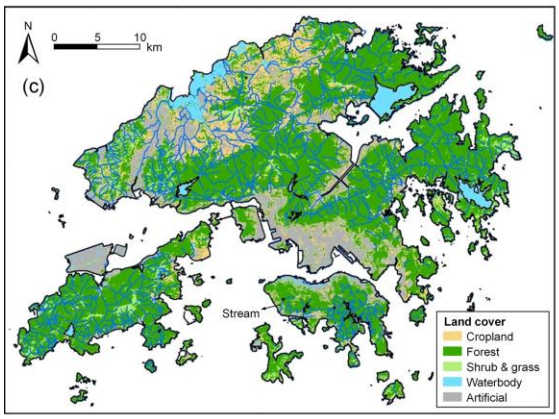

Fig. 19c

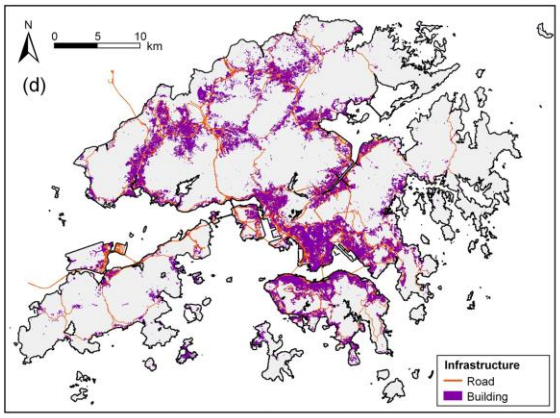

Fig. 19d

Accepted manuscript doi:
10.1680/jgeot.21.00160

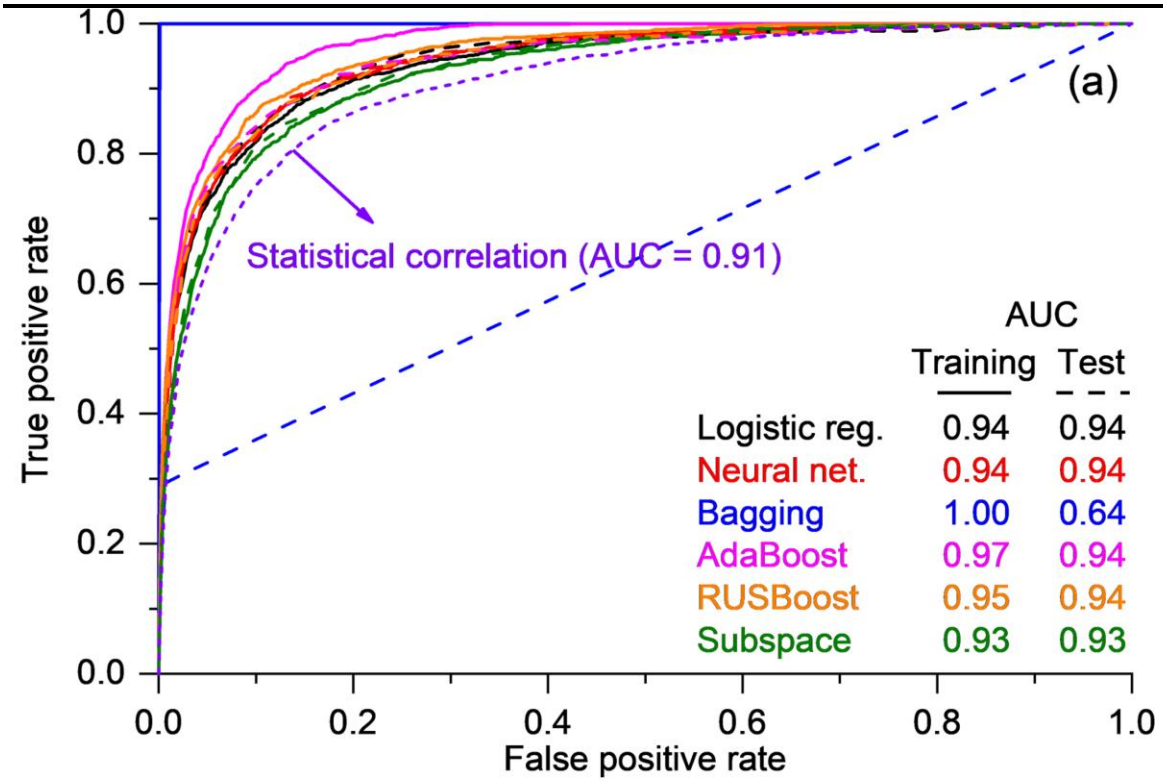

Fig. 20a

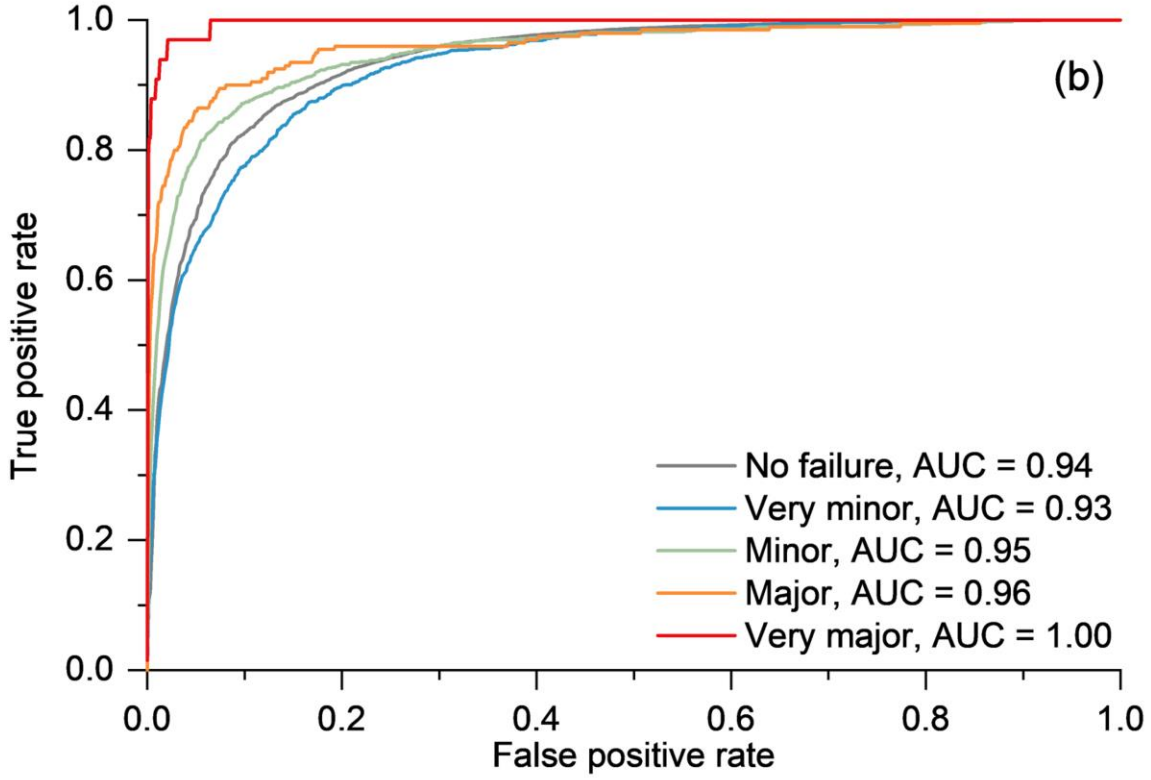

Fig. 20b

Accepted manuscript doi:
10.1680/jgeot.21.00160

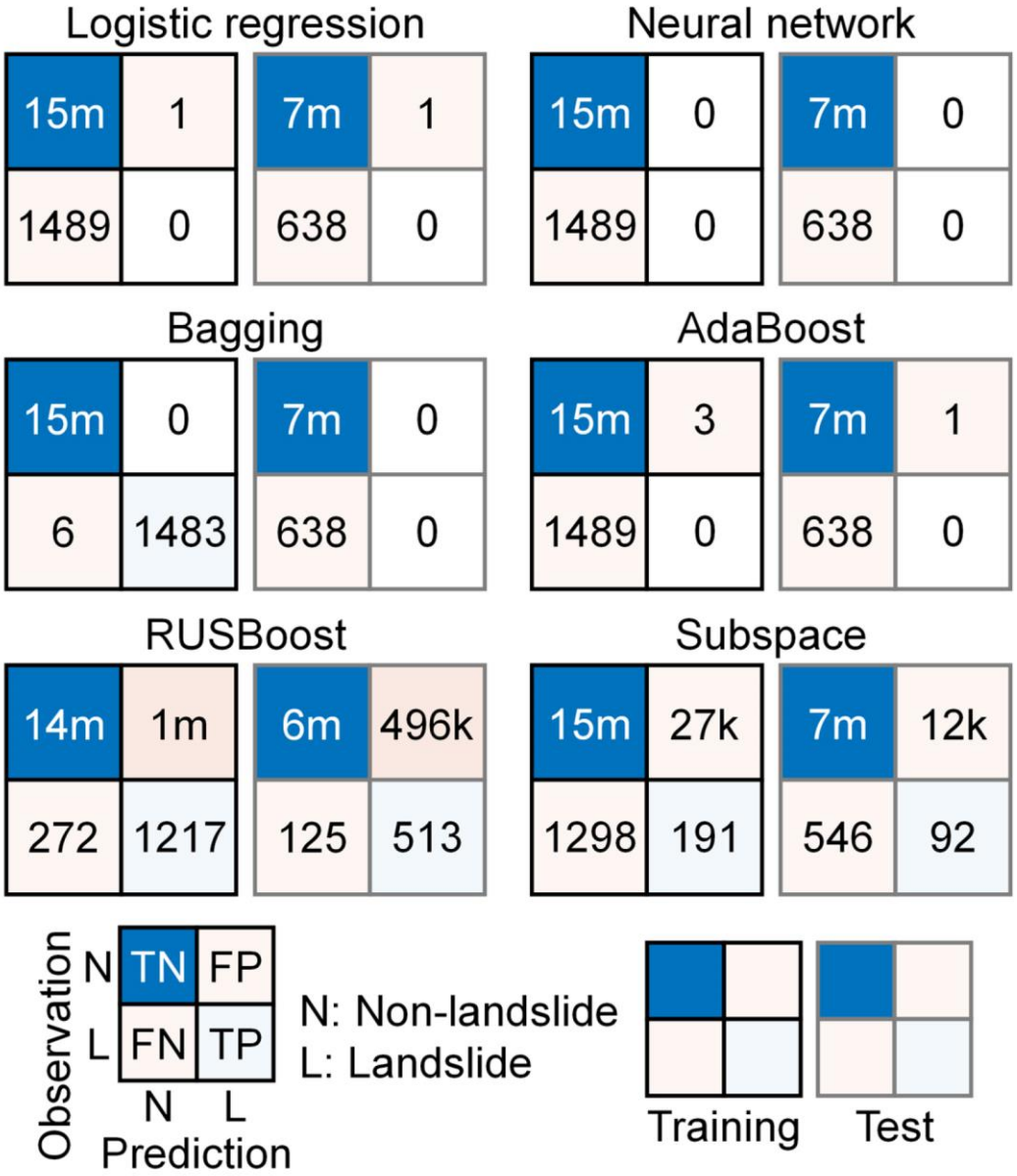

Fig. 21

Accepted manuscript doi:
10.1680/jgeot.21.00160

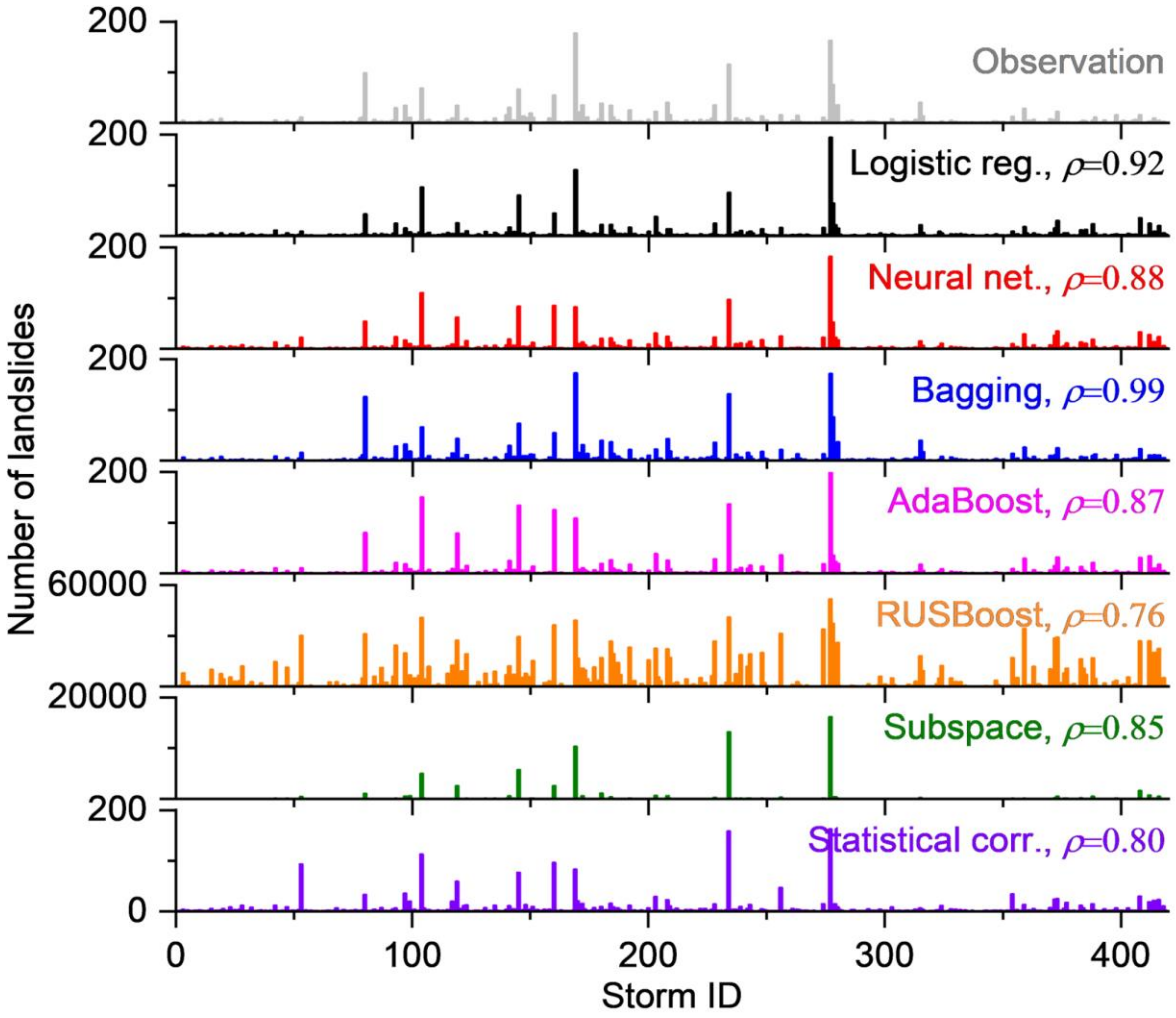

Fig. 22