~~Of reliable l~~Landslide forecasting and factors influencing predictability

Emanuele Intrieri[1]*, Giovanni Gigli[1]
[1] *Department of Earth Sciences, University of Studies of Firenze, via La Pira 4, 50121 Firenze, Italy.*
**Corresponding author*

ABSTRACT

Forecasting a catastrophic collapse is a key element in landslide risk reduction, but also a very difficult task, owing to the scientific difficulties in predicting a complex natural event and also to the severe social repercussions caused by a false or a missed alarm. A prediction is always affected by a certain error, however when this error can imply evacuations or other severe consequences a high reliability in the forecast is, at least, desirable.

In order to increase the confidence of predictions, a new methodology is ~~here~~ presented here. Differently from traditional approaches, it iteratively applies several forecasting methods based on displacement data and, also thanks to an innovative data representation, gives a valuation about the reliability of ~~of how the~~ the prediction ~~is reliable~~. This approach has been employed to back-analyse 15 landslide collapses. By introducing a predictability index, this study also contributes to the understanding of how geology and other factors influence the possibility to forecast a slope failure. The results showed how kinematics, and all the factors influencing it such as geomechanics, rainfall and other external agents, ~~is the key feature~~are the key ~~when~~ concerning landslide predictability.

Keywords: landslides; forecasting; geomechanics; early warning; time of failure; slope failure

INTRODUCTION

Natural disaster forecasting for early warning purposes is a field of study that drew the media attention after events such as the 26th December 2004 tsunami of Sumatra. Predicting landslides, with respect to other natural hazards, is a complex task due to the influence of many factors like geomechanical properties, rainfall, ground saturation, topography, earthquakes and many others. So far, few empirical landslide forecasting methods exist (Azimi et al., 1988; Fukuzono, 1985a; Mufundirwa et al., 2010; Saito, 1969; Voight et al., 1988) and none furnishes a reliability degree about the prediction, making them unsuitable for decision making. In particular, when mentioning geomechanics, the reference is ~~we particularly refer~~ to the study of the behaviour of a landslide concerning its deformation with relation to the applied stress, with special ~~particular~~ reference to its post-rupture conditions.

In ~~our~~ the present paper research ~~we present~~ an approach to perform probabilistic forecasting of landslide~~s~~ collapse is presented. This has been achieved by reiterating several predictions using more forecasting methods at the same time on multiple time series. This approach may have important applications to civil protection purposes as it provides the decision makers with a level of confidence about the prediction. Furthermore, this study, performed on 15 different case studies, shows how the possibility or not to forecast the time of collapse of a landslide is affected by geomechanical or geomorphological features as much as by circumstantial conditions.

The inverse velocity forecasting method

Forecasting activity can be considered the fulcrum of early warning systems (Intrieri et al., 2013), i.e. cost-effective tools for mitigating risks by moving the elements at risk away. For many natural phenomena forecasting is common practice (for example for hurricanes; Willoughby et al., 2007), while for others is, at present, impossible (earthquakes; Jordan et al., 2011). Landslides lie in between. Their prediction can be performed through rainfall thresholds (Baum and Godt, 2010), but a more reliable approach should make use of direct measures of potential instability, such as displacements (Lacasse and Nadim, 2010; Blikra, 2008). A first issue is that only a small percentage of landslides in the world is appropriately monitored, that often monitoring is carried out for short periods not encompassing the final pre-failure stages, or may have been carried out with a too low temporal frequency that does not permit to follow the displacement trend. This also causes an insufficient knowledge of the geomechanical processes leading to failure (here meant as the collapse following a sudden acceleration, either a first movement or a reactivation), which is another responsible for our deficiencies in predicting landslides.

In spite of this, few empirical methods for predicting the time of failure based on movement monitoring data have been developed (Azimi et al., 1988; Fukuzono, 1985a; Mufundirwa et al., 2010; Saito, 1969) and further investigated on a physical basis (Voight et al., 1988). They are all based on the hypothesis that if a landslide follows a peculiar time-dependant geomechanical behaviour (called creep; Dusseault and Fordham, 1994), it will display a hyperbolic acceleration of displacements before failure; by extrapolating this trend from a displacement time series through empirical arguments, it is possible to obtain the predicted time of failure. However such methods do not always produce good results. In fact, other than the limitation of working only with creep behaviour~~s~~, sometimes the tertiary creep can evolve such rapidly that a sufficient lead time for evacuation is simply not possible (IEEIRP, 2015). In other cases natural or instrumental noise can hamper the predictions and require ~~further data treatment~~post-processing to allow for effective warnings (more details on the types and effects of noise can be found in Carlà et al., 2016). Other authors also contributed to methodologies to exploit ~~such~~ and optimize the classic forecasting methods (Crosta and Agliardi, 2003; Dick et al., 2015; Manconi and Giordan, 2015).

One of the most famous methods is Fukuzono's (1985a), which derives from Saito's (1969), from here on simply called F and S method, respectively. It requires that during the acceleration typical of the final stage of the creep (tertiary creep), the inverse of displacement velocity (v^{-1}) decreases with time. The collapse is forecasted to occur when the extrapolated line reaches the abscissa axis (corresponding to a theoretical infinite velocity). Such line may either be convex, straight or concave (Fukuzono, 1985a). When it is straight this phenomenon is sometimes referred to as Saito effect (Petley et al., 2008).

The possibility to find landslides showing the Saito effect has been related to the mechanical properties of the sliding mass. However there is no general consensus on this issue.

According to some authors (Petley, 2004; Petley et al., 2002), in order to display the Saito effect, landslides need to display a brittle behaviour (which indicates a drop from peak strength to residual strength value, deformation which is concentrated along a well defined shear surface, sudden movements and catastrophic failure, usually associated with crack formation in strong rocks); furthermore only brittle, intact rocks evolve in catastrophic landslides and therefore can be predicted; for others (Rose and Hungr, 2007), on the opposite, landslides displaying the Saito effect must have ductile failures in order to be forecasted (i.e. slower, indefinite deformation along a shear zone and under a constant stress, typical of sliding on pre-existing surfaces of soft rocks), as brittleness is characterized by sudden, impossible to anticipate, ruptures.

This complex subject is made even more difficult due to the influence of external factors (rainfall, earthquakes, excavations), structural constraints (joints, faults, contacts with different lithologies) and sometimes unknown elements within the mass (the conditions of the shear surface, the history of the landslide, the presence of rock bridges). Therefore it is often hard to establish the mechanical behaviour and even more to find an exact correlation between the mechanical behaviour of a landslide and the possibility to predict its failure.

The concept of predictability

Before assessing the influence of geomechanics on the predictability of a landslide it is first necessary to address the concept of predictability.

In literature (Azimi et al., 1988; Hutchinson, 2001; Mufundirwa et al., 2010; Rose and Hungr, 2007) there are papers that deal with “predictions” made in retrospect, that is thorough post-event analyses showing the signs of a critical pre-collapse acceleration; however whether such signs would have been unambiguous or would have granted a sufficient lead time is often neglected.

On the other hand in ~~our~~ this research ~~we consider~~ an operational definition of predictability is considered (integrating the one of early warning system; UNISDR, 2009) as the feature possessed by a landslide which allows one to forecast its collapse with reasonable confidence and sufficiently in advance, permitting the dispatch of meaningful warning information to enable individuals, communities and organizations threatened by the hazard to prepare and to act appropriately and in sufficient time to reduce the possibility of harm or loss. Therefore, displaying the Saito effect is not the only prerequisite for an operational prediction, there is also the need for repeated time of failure forecasts fluctuating around a constant time value placed not too close in the future. This has been achieved through the reiterative approach and the graphical representation described in the following paragraph. Finally a semi-quantitative parameter called Prediction Index is defined in order to address the success of the predictions.

METHODS

The usual way to apply landslide forecasting methods based on displacements, is to obtain a single predicted time of failure (t_f) and to update such prediction as soon as new data are gathered (Rose and Hungr, 2007). This is a deterministic approach, since the real time of failure (T_f) is predicted through a single inference. ~~At most~~Even if sometimes more predictions ~~can be~~are made ~~in the future~~together with new data,~~ but~~ usually only one (the most recent) is used. On the other hand, in order to account for the uncertainty of the methods and complexity of the phenomena, predictions should have a certain confidence. Confidence may be quantitatively assessed by using the standard deviation of the forecasts t_f as a proxy. In fact the standard deviation furnishes the dispersion (i.e. the precision) of the predictions, which may be used to calculate a time window within which the collapse is more likely to occur. Therefore the lower the standard deviation of a set of forecasts, the higher would be their reliability and the confidence~~(for example given by the standard deviation of t_f)~~. This is especially important for operative early warning systems. ~~We achieved t~~This probabilistic approach is achieved by reiterating the equations from Saito (1969), Fukuzono (1985a) and Mufundirwa et al. (2010) (the latter method will be called M method from here on) for finding t_f, using continuously new data and enabling the calculation of the standard deviation.

The predictions are plotted versus the time when they have been made (time of prediction, t_p). We call these diagrams prediction plots (Figure 1). A prediction is considered reliable when the inferences oscillate around the same t_f. Figure 1 also shows that since reliable predictions usually display an oscillatory trend, the most updated one is not necessarily the most accurate, contrarily to what is usually believed (Rose and Hungr, 2007) in fact, the length of the dataset is more important, from which T_f can be estimated through simple statistical analyses (like mean and standard deviation).

Since in some cases a single forecasting method can fail to give satisfactory results, in order to improve even more the confidence in the predictions, a multi-model approach is adopted together with the probabilistic approach. In fact, according to the Diversity Prediction Theorem (Page, 2007; Hong and Page, 2008), diversity in predictive models reduces collective error. The highest confidence, of course, is reached when all the employed method independently converge towards the same result.

Formattato: Tipo di carattere:

Formattato: Tipo di carattere: Corsivo, Pedice

On the other hand, confidence it may also be considered as a qualitative increase in the awareness of the decision makers that can estimate the time of failure of a landslide by evaluating a large set of different predictions and their dispersions.

-For this research ~~we confronted~~ the results from S and F methods have been confronted and from the method by Mufundirwa et al. (2010). The equations used for the iteration are obtained from the respective authors and are:

$$t_r = \frac{t_2^2 - (t_1 \cdot t_3)}{2t_2 - (t_1 + t_3)}, \quad (1)$$

for S method, where t_1, t_2, t_3 are times taken so that the displacement occurred between t_1 and t_2 is the same as between t_2 and t_3.

$$t_r = \frac{t_2 \frac{1}{v_1} - t_1 \frac{1}{v_2}}{\frac{1}{v_1} - \frac{1}{v_2}}, \quad (2)$$

for F method, where v_1 and v_2 are the velocities at arbitrary times t_1 and t_2.

$$t \frac{dD}{dt} = t_r \frac{dD}{dt} - B, \quad (3)$$

for M method, where D is the displacement and t_r is the angular coefficient of the line represented in a $t \frac{dD}{dt} = f\left(\frac{dD}{dt}\right)$ space having B as the intercept. For the purposes of this paper t_r expressed in all these equations is equivalent to t_f.

Formattato: Tipo di carattere:

Formattato: Tipo di carattere: Corsivo, Pedice

Formattato: Tipo di carattere:

Formattato: Tipo di carattere: Corsivo, Pedice

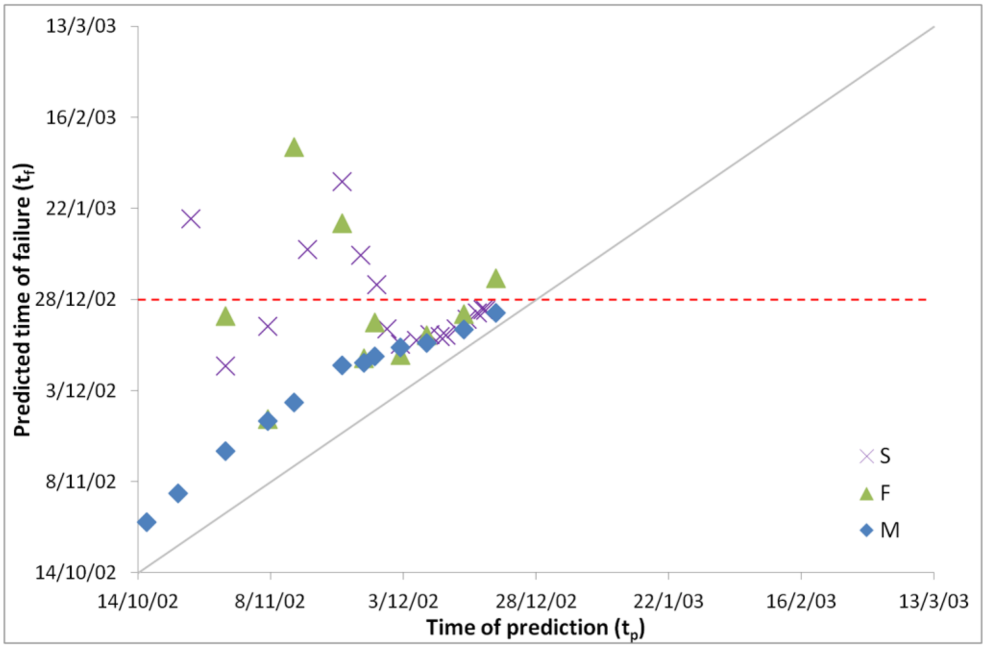


Figure 1. This graph represents probabilistic predictions performed with 3 different forecasting methods (Fukuzono, 1985a; Mufundirwa et al., 2010; Saito, 1969) applied to the MB34-35' displacement time series of Mount Beni landslide (Gigli et al., 2011). The horizontal dashed line indicates the observed time of failure (T_f) and the grey diagonal line the equality between t_f and t_p. Therefore the vertical distance between a point and the dashed line indicates the prediction error. The vertical distance between the diagonal line and a prediction above it is the life expectancy of the landslide at the time of prediction. In this case the predictions obtained through S and F methods give a good estimation of T_f, while the one from Mufundirwa et al. (2010) consistently forecasts the collapse few days ahead.

The proposed procedure consists in iteratively calculating the time of failure t_f by using the aforementioned methods and to repeat the calculation as soon as new monitoring data are available. All the forecasts are recorded together with the time when they are made, in order to create a time series of $t_f = f(t)$. This can be represented in a prediction plot having t_f and t (the time when the prediction is made) as coordinates. Finally, from the distribution of the forecasts with time it is possible to assess the time of failure.

Formattato: Normale
Formattato: Tipo di carattere:
Formattato: Tipo di carattere: Corsivo, Pedice
Formattato: Tipo di carattere:
Formattato: Tipo di carattere: Corsivo, Pedice
Formattato: Tipo di carattere:
Formattato: Tipo di carattere: Corsivo, Pedice
Formattato: Tipo di carattere:

TIME OF FAILURE PREDICTION

In order to find a relation between the predictability of a failure and the geological features of the landslide, S, F and M methods have been applied to a number of different real case studies. Some geological features of interest relative to such cases are reported in TABLE 1, when they were known or applicable. Concerning brittleness, since it was rarely explicitly stated in the referenced articles, it was assessed based on information such as the type of material, the presence of a reactivated landslide, the weathering and the shape of the displacement time series.

Since this lead to approximations, brittleness has been evaluated using broad and qualitative definitions.

Since T_f must be known in order to assess the quality of predictions, all the case studies are from past landslides that have already failed. Therefore the respective time of failures are all a posteriori known.

A few representative examples of prediction plots are showed in Figure 1 and Figure 2. Mount Beni landslide is a 500.000 m^3 topple that evolved as a rockslide (Gigli et al., 2011). It developed on a slope object of quarrying activity. The predictions oscillate quite regularly around the observed time of failure (T_f, dashed line in Figure 2). It is this convergence that permits to correctly forecast the collapse a priori at least since late November, i.e. a month before the failure, whereas a single forecast would not be able to give a confidence of the prediction. The three methods are similar to the point that S and F previsions can be partially overlapped. M previsions overlap as well but only in the final part. The M method alone would not be sufficient for spreading a reliable alarm as the single forecasts do not converge but move forward to a different time of failure as the time passes by.

Similar behaviours can be observed also for the cases of Figure 2 that display landslides with a different array of geological features (as seen in TABLE 1). The best results are obtained when the forecasts oscillate around T_f with sufficient time in advance (as for Vajont and, limited to F method, for Liberty Pit) or when they consistently give the similar t_f (as for the artificial landslide E, where the terms "artificial landslide" indicate a landslide recreated in laboratory with an artificial slope). In other cases (Avran valley and, limited to S and M method, for Liberty Pit) the predictions are too scattered or simply never converge toward a single result, thus making it impossible to foresee a reliable time of failure.

Notably, considering for example only the results of the S method in the case of the Avran valley landslide, since the end of September the forecasts are constantly furnishing a time of failure preceding the actual T_f. Although this may be considered a case of safe predictions (that is an error not producing a false positive and therefore not dangerous for the elements at risk), this also means that, at every forecast that is made, t_f is postponed. Given a series of ever increasing values of t_f, it is impossible to assess which of them (if any) can be assumed as a good estimate of the actual time of failure. However, since the time series of predictions is long enough, past forecasts (before early September) furnish values of t_f that, if considered together with the late ones, centre the value of T_f. Therefore it is clear how a prediction plot may allow decision makers to make more aware evaluations of the time of collapse of a landslide.

The results of the prediction plots can be roughly summarized reporting the mean and standard deviation of the forecasts for each method (Figure 3).

TABLE 1. LANDSLIDE CASE HISTORIES

Name	Material	Type	Brittleness	Volume (m^3)	Predisposing factor	Trigger	History	Basal geometry	Ref. *
Liberty Pit	Weathered quartz monzonite	Rockslide?	Medium/high	6x10^6	N.D.	Blasts, pore water pressure	First time failure	Planar?	1, 2
Landslide in mine	Consolidated alluvial sediments, weathered bedrock	Deep-seated toppling in bedrock	Medium	10^6	Blasts, pore water pressure	N.D.	First time failure?	N.D.	1
Betze-Post	Weathered granodiorite	Rockslide?	Medium/high	2x10^6	N.D.	Rainfall	First time failure?	Wedge intersections?	1

Formattato: Tipo di carattere:
Formattato: Tipo di carattere: Corsivo, Pedice
Formattato: Tipo di carattere:
Formattato: Tipo di carattere: Corsivo, Pedice
Formattato: Tipo di carattere:
Formattato: Tipo di carattere: Corsivo, Pedice
Formattato: Tipo di carattere:
Formattato: Tipo di carattere: Corsivo, Pedice
Formattato: Tipo di carattere:
Formattato: Tipo di carattere: Corsivo, Pedice

Vajont	limestone and clay	Rock slide	High	$2.7x10^{8}$	N.D.	Pore water pressure	Reactivated	Concave	1, 3
Stromboli †	Shoshonitic basalts	Bulging (not a landslide)	Medium/high	N.D.	N.D.	Sill intrusion	First time failure	N.D.	4
Monte Beni	Ophiolitic breccias	Topple/rock slide	High	$5x10^{5}$	Rainfall, structure, basal excavation	N.D.	First time failure	Stepped	5
Cerzeto	Weathered metamorphic rocks on top, cataclastic zone and Pliocene clays	Debris slide-earth flow	Medium/low	$5x10^{6}$	Tectonized area, permeability differences	Prolonged rainfalls	Reactivated ?	Compound (steeper and irregular in the upper zone and gentler in the clays	6
Rock mass failure Japan	Clayey limestone	Rockslide?	High (within limestone)?	$5x10^{2}$	"Structural complexity" (?)	Intense rainfall	First time failure?	Planar?	7
Asamushi	Liparitic tuff, jointed and weathered. Clay in the joints		Medium/low	10^{5}	N.D.	N.D.	N.D.	Concave?	7, 8
Avran valley	Chalk	Rockslide	Medium/low	$8x10^{4}$	N.D.	N.D.	First time failure?	Convex	9
Giau Pass	Morainic material	Complex slide	Medium/low	$5x10^{5}$	N.D.	Pore water pressure	Preexisting shear surface	Composite	10, 11
Artificial landslide A	Loam	Earth slide	Low	N.D.	N.D.	Prolonged rainfall	First time failure	Planar	12
Artificial landslide B	Sand	Earth slide	Low	N.D.	N.D.	Prolonged rainfall	First time failure	Planar	12
Artificial landslide C	Sand	Earth slide	Low	N.D.	N.D.	Prolonged rainfall	First time failure	Convex	12
Artificial landslide D	Sand	Earth slide	Low	N.D.	N.D.	Prolonged rainfall	First time failure	Planar	12

*The references used are numbered as follows: 1: Rose and Hungr, 2007; 2: Zavodni and Broadbent, 1980; 3: Semenza and Melidoro, 1992; 4: Casagli et al., 2009; 5: Gigli et al., 2011; 6: Iovine et al., 2006; 7: Mufundirwa et al., 2010; 8: Saito, 1969; 9: Azimi et al., 1988; 10: Petley et al., 2002; 11: Angeli et al., 1989; 12: Fukuzono, 1985b.

† The case of Stromboli is not relative to a landslide, rather to a volcanic bulging preceding a vent opening that was forecasted in a similar fashion of a landslide and therefore here included.

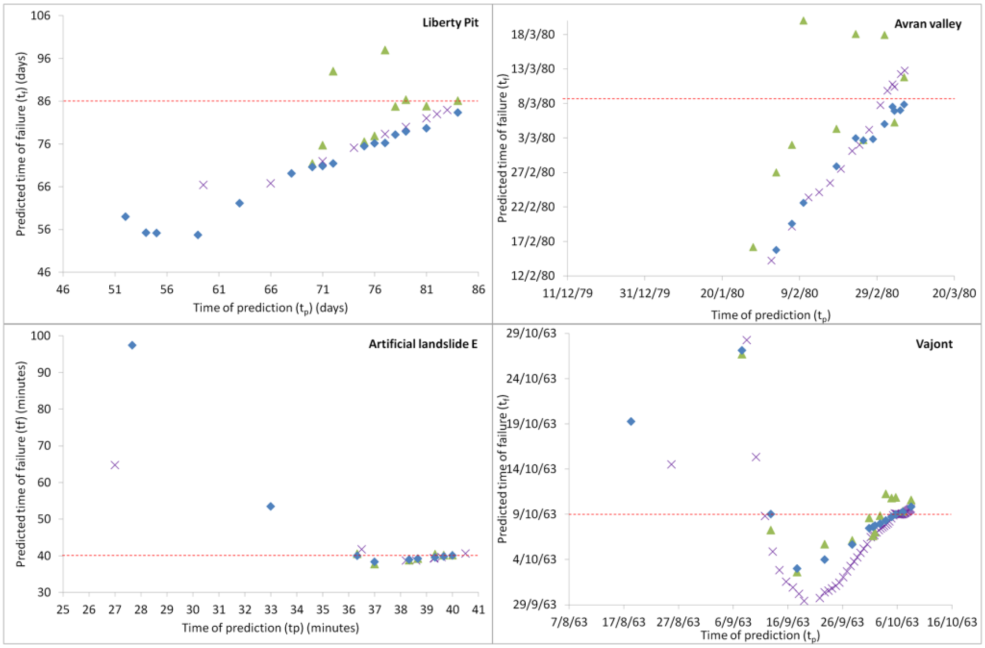
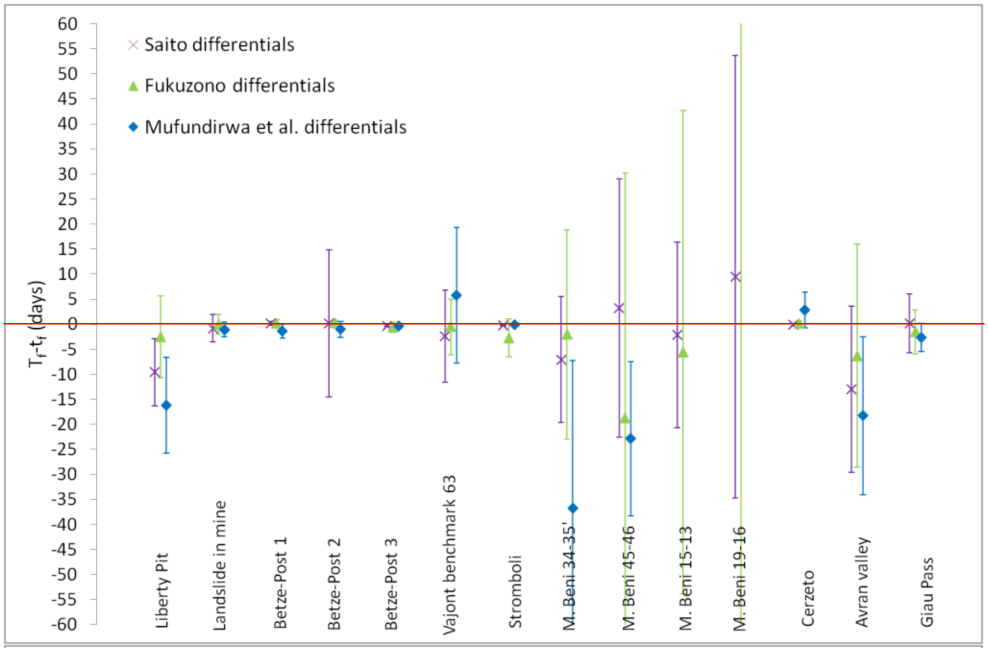


Figure 2. ~~These graphs show how iterating forecasts performed through multiple forecasting methods increases the confidence when estimating the actual time of failure~~Prediction plots of four different case studies. The dashed line indicates ~~(~~T_f~~, dashed line)~~. The crosses represent forecasts performed with S method, the triangles with F method and the diamonds with M method. Note that F forecasts for Avran valley landslide include other less accurate values not showed in the graph as they are out of scale.



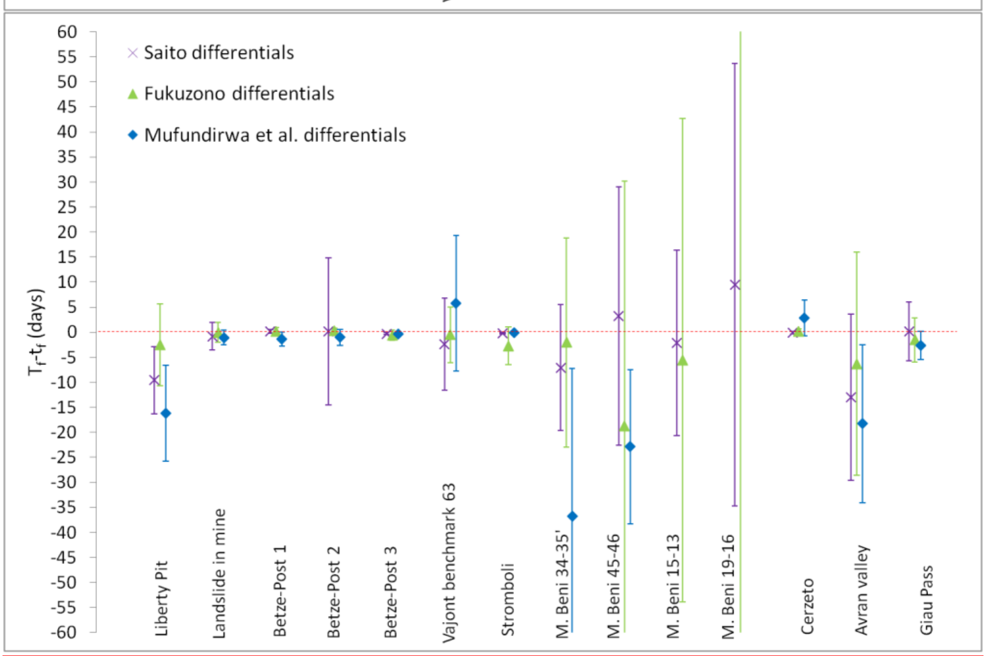


Figure 3. This graph represents for each method the differential between the mean of the forecasts ($\bar{t}_f$) and the actual time of failure (T_f). Negative values are safe predictions as anticipate

the time of failure. The dashed line represents exact predictions (T_f - $\bar{t}_f$ = 0). The standard deviations of the forecasts are represented as error bars. For Betze-Post and Mount Beni landslides, time series from different measuring points are reported. The rock mass failure, Asamushi landslide and the artificial landslides are not shown as were monitored in a different time scale (hours or minutes).

PREDICTABILITY INDEX

In order to evaluate the performance of S, F and M methods and to relate it to the characteristics of the reported examples, an arbitrary scoring system has been implemented and attributed to each prediction plot (considering that every time series has a prediction plot for each forecasting method and that for some case studies more than one time series was available). This permits to quantify the predictability of a collapse based on the prediction plot. A score from 1 to 5 has been assigned according to the following criteria:

- 1 point: the prediction plot never converges on a single t_f (typically t_f increases at every new datum available).
- 2 points: the predictions vary considerably at every new iteration. An average time of failure ($\bar{t}_f$) can be extracted but with high uncertainty.
- 3 points: the predictions oscillate around T_f, although with a certain variance.
- 4 points: the predictions have a low variance although $\bar{t}_f$ is slightly different than T_f. Note that when the variance was low, $\bar{t}_f$ and T_f never differed greatly.
- 5 points: the prediction plot is clearly centred on T_f therefore the reliability of $\bar{t}_f$ is high.

By summing the scores obtained from S, F and M prediction for each time series, what we call the Predictability Index (*PI*) is obtained (TABLE 2). Since PI is a means to evaluate the overall quality of a set of predictions (it requires to observe the time series of t_f and confront it with T_f, it is the predictability index) and also to compare the performance of different forecasting methods with different case studies, naturally it can only be estimated after the collapse.
By using 3 forecasting methods, *PI* ranges from 3 (impossible to predict the time of failure) to 15 (the time of failure can be predicted in advance and with a high reliability). Though a certain degree of subjectivity is unavoidable when assigning the scores, what matters here is the relative difference of *PI* between the case studies. In such a way it is possible to understand in which conditions a landslide is more or less predictable.

TABLE 2. PREDICTABILITY INDEX

Name	S	F	M	*PI*	Inverse velocity trend	Notes
Liberty Pit	1	5	1	7	Asymptotic (linear at the end)	Open pit mine, structural control of 2 intersecting faults
Landslide in mine	5	5	5	15	Linear	Open pit mine
Betze-Post 1	3	3	1	7	Linear	Open pit mine
Betze-Post 2	4	5	4	13	Linear	Open pit mine
Betze-Post 3	5	4	1	10	Linear	Open pit mine
Vajont benchmark 63	5	5	5	15	Linear	Air pressure and cementation caused catastrophic collapse
Stromboli	1	2	2	5	Asymptotic	Volcanic context
Mount Beni 12-9	4	5	1	10	Concave	Back fracture
Mount Beni a'b'	1	3	1	5	Linear	Short time series

Mount Beni 15-13	5	3	1	9	Linear	Internal fracture
Mount Beni 34-35'	5	3	1	9	Linear	Lateral fracture, short time series
Mount Beni 45-47	2	3	1	6	Linear	Back fracture, short time series
Mount Beni 3-2	5	2	1	8	Concave	Back fracture
Mount Beni 4'-6	1	4	1	6	Linear	Back fracture, short time series
Mount Beni 24-23	4	2	1	7	Linear	lateral fracture
Mount Beni 49-24	5	1	1	7	Linear	Lateral fracture, short time series
Mount Beni 35'-36	2	5	1	8	Linear	Lateral fracture, short time series
Mount Beni 33-35'	3	3	1	7	Linear	Lateral fracture, short time series
Mount Beni 36-37	4	3	1	8	Linear	Lateral fracture
Mount Beni 19-16	2	2	1	5	Linear	Lateral fracture
Mount Beni 19-17	1	2	1	4	Linear	Lateral fracture, short time series
Mount Beni 33-34	4	2	1	7	Linear	Internal fracture
Mount Beni 43-44	3	2	1	6	Asymptotic (constant velocity at the end)	Internal fracture, short time series
Mount Beni 40-41	3	2	1	6	Asymptotic (constant velocity at the end)	Internal fracture, short time series
Mount Beni 40-42	3	3	1	7	Linear	Internal fracture, short time series
Mount Beni 45-46	3	2	2	7	Linear	Back fracture, short time series
Mount Beni 1-2	4	2	1	7	Linear	Back fracture
Cerzeto	5	5	1	11	Linear	N.A.
Rock mass failure Japan	2	2	1	5	Convex	Open pit mine, very small landslide
Asamushi	5	3	1	9	Linear	N.A.
Avran valley 5	1	2	1	4	Concave	N.A.
Avran valley 6	1	1	1	3	Asymptotic	N.A.
Avran valley 7	1	2	1	4	Concave	N.A.
Giau Pass	3	3	1	7	Asymptotic /concave	N.A.
Artificial landslide A	5	5	5	15	Convex	40° artificial slope
Artificial landslide B	2	2	3	7	Concave	40° artificial slope
Artificial landslide C	1	2	3	6	Linear (slightly convex)	40° artificial slope
Artificial landslide D	5	5	5	15	Linear	30° artificial slope

DISCUSSION

TABLE 2 shows how the most predictable events (*PI* > 8) can display very different features and are quite irrespective of the shape of the inverse velocity plot, the volume, the brittleness of the material, the history of the landslide and so on (see also TABLE 1).

A comparison between Figure 3 and TABLE 2 illustrates how the mean and standard deviation of the forecasts alone are not enough to represent the quality of predictions and, consequently, the predictability of a landslide. In fact the importance of a single forecast strongly depends on the time when it is made; for example, given the same set of forecasts ($t_{f,i}$), a higher *PI* is obtained if the first predictions done are the farthest from T_f while the final ones tend to converge to it; in this way the prediction plot assumes an oscillatory shape (as for S and F forecasts in Figure 1). Conversely, if the same forecasts are made with a different order so that they get closer and closer to T_f as time passes by (that is $|t_{f,i} - T_f| < |t_{f,i-1} - T_f|$), then there is no $t_{f,i}$ prevailing on the others and it is not possible to define a more probable time of collapse (as for

M forecasts in Figure 1). However the average and standard deviation of t_f are the same for both cases and this explains why these two statistics alone are not as informative as a prediction plot. From TABLE 2 it is also possible to assess which method gives the best results. The sum of the scores for S, F and M is 119, 115 and 63 respectively. Overall S and F perform similarly, but for a specific case study their effectiveness can be very different, therefore their result are independent and not redundant; there is no indisputable clue suggesting when F method is more performing than S and vice versa; nonetheless it appears that S is negatively influenced when the displacement curve is not regularly accelerating (Liberty Pit, Stromboli), whereas for F a few aligned points in the final tract in the inverse velocity plot are sufficient for predicting the failure; however F forecasts are more disturbed when displacement data are noisy, since they use their derivative (velocity) as input. Eventually M forecasts generally perform more poorly and rarely (i.e. artificial landslides B and C) surpass those obtained from S and F methods.

Interestingly, different displacement time series belonging to the same landslide can display different behaviours. This is a strong evidence that, even though the geological features do influence the predictability of a landslide, assuming that they keep the same for the whole landslide, other factors must determine the quality of the predictions. The last column of TABLE 2 shows for each time series what such factors could be, such as lithology (the asymptotic trends of the cases of Avran valley and Giau Pass can be explained as consequences of a lowly brittle material according to Petley’s experiments; Petley, 2004), external forces (excavation in open pit mines, volcanic activity, rainfall), local effects (structural constraints, displacement measured relative to internal or lateral fractures not representing the general instability of the landslide), quality of data (length of the time series, frequency of the observations, level of noise, representativeness of the monitored point) etc.

All these case histories show that the main responsible for the predictability of a landslide, and secondary also for the presence or not of the “Saito effect”, is in a way or another connected to geology, However this relation is not simple nor direct~~ but not simply and directly~~. Instead both the predictability and the “Saito effect” depend on the kinematics of the landslide, since only a landslide accelerating with a certain trend can be forecasted using S, F and M methods. Naturally, the kinematics~~which~~ in turn depend~~s~~ on the geological conditions. In the complex relation between geology and kinematics the aforementioned factors may intervene. Although their interaction may not be known, its effect on displacement data can be easily measured. As a result it has been found that~~and~~ asymptotic trends in the inverse velocity plot have been encountered also for first failure ruptures (as found in some time series of Mount Beni landslide), contrarily to what is described by Petley (2004). This can be explained as an effect of those interactions which may alter in an unknown way the normal relation between geology and kinematics, thus making focusing on kinematics as the key more reliable than relying on geology alone.

~~In other words, e~~In fact, even though geomechanics is unquestionably a key factor, a complete geomechanical characterization is often difficult to accomplish, ~~it is sometimes difficult to have a deep knowledge of the geomechanical features of a landslide,~~ especially ~~in the field and~~ in emergency situations~~, although some safe assumptions can always been done by observation and a broad knowledge of the area~~. Hints of a particular geomechanical behaviour ~~What it may be known about them is in part thanks to what is~~are often derived from displacement data. Like in a black box model, even if the real properties of a phenomenon are not known, ~~we can draw~~ conclusions may be drawn from the output of those properties (i.e. the kinematics). In this case, importance has been done to kinematics because what is generally measured by monitoring are

Formattato: Tipo di carattere: Non Corsivo

Formattato: Tipo di carattere: Non Corsivo

Formattato: Tipo di carattere: Non Corsivo

Formattato: Tipo di carattere: Non Corsivo

Formattato: Tipo di carattere: Non Corsivo

Formattato: Tipo di carattere: Non Corsivo

Formattato: Tipo di carattere: Non Corsivo displacement data, Furthermore,~~ and because~~ many other unknown factors (rainfall, ground saturation, earthquakes, anthropic disturbance, etc.) are included in the black box model together with the geomechanics; this makes it virtually impossible to know in advance what may be the degree of influence of geomechanics alone with respect to other factors, thus leading to focusing on kinematics instead. Moreover, even though geomechanics is a key element in determining landslide predictability (for example because it is responsible for the creep behaviour), ~~we~~ the results of the present study showed that landslide prediction can be carried out with a variety of different geomechanical settings, as can also be observed by comparing TABLE 1 (which furnishes evaluations concerning the geomechanical properties of the case studies) with TABLE 2 (which states whether a collapse was predictable or not).

~~Finally, t~~The prediction plots clearly show that, contrarily to what is generally believed (Rose and Hungr, 2007), the last forecasts are not necessarily the most accurate and that past ones (starting from the initiation of the tertiary creep) are essential to estimate the correct time of failure. In fact older forecasts can be more accurate and in any case furnish precious information about the general reliability of the final prediction, as explained above. Therefore the present study highlights the importance of considering the whole set of predictions made with time. The integration of more forecasting methods further raises reliability of the predictions, which is of great importance for early warning systems, in particular when evacuations are envisaged.

Limitations of the proposed approach are those related to the intrinsic limitations of the forecasting methods that have been integrated. In fact, since S, F and M methods are all based on the creep theory, the occurrence of a tertiary creep phase slow enough to allow to monitor and take action is necessary. Voight (1988) also assumes that there must be no external force acting on the landslide, but the examples shown in this paper demonstrate that this may not represent a limitation.

Figure 3 shows that the mean of the predictions can be used as a proxy for the time of failure but, as stated above in this paragraph, it is also shown that the obtained accuracy may not be enough as the mean does not exploits all the information provided by a prediction plot. Other statistical indicators have been attempted but none of them appeared to better approximate the value of T_f, mainly due to the difficulty of accounting for the important time factor in the forecasts and also because not every prediction plot displays the characteristic oscillations. Therefore, the interpretation of the prediction plot (and in particular of the dispersion of the forecasts with time) represents the most valuable tool for decision makers, who, in this way, can make aware judgements informed with a large set of quantitative and redundant data and therefore assessing the “weight” of a single prediction by comparing it with many others.

Resuming, the proposed methodology can be summarized as in Figure 4.

Formattato: Tipo di carattere: Non Corsivo

Formattato: Tipo di carattere: Non Corsivo

Formattato: Tipo di carattere: Non Corsivo

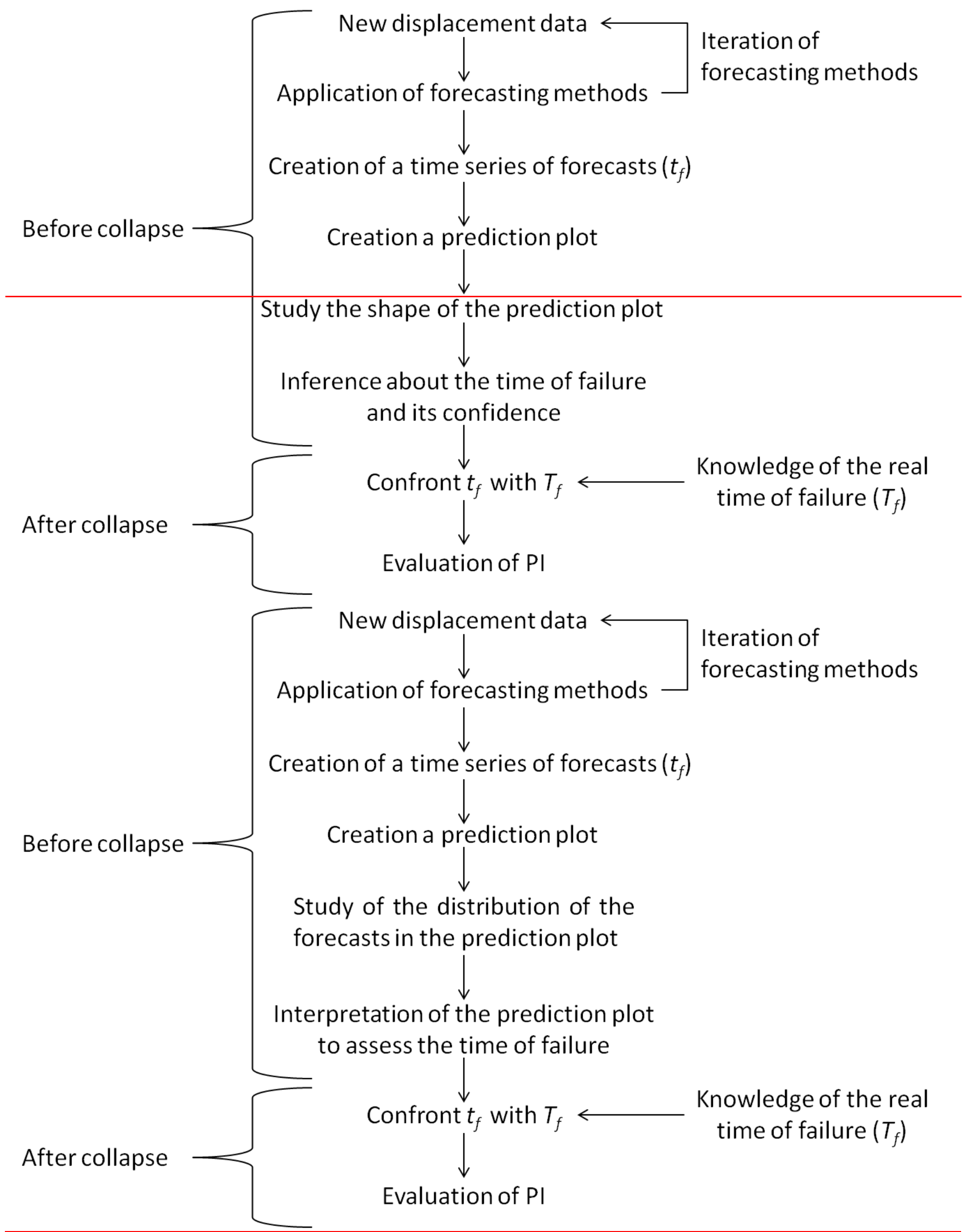
Formattato: Tipo di carattere: Non Corsivo

Formattato: Tipo di carattere: Non Corsivo

Formattato: Tipo di carattere:

Formattato: Tipo di carattere: Corsivo, Pedice

362



363
364 **Figure 4.** Flow-chart that synthesises the proposed procedure.

CONCLUSIONS

In conclusion, the main aspect of the proposed methodology concerns a way to produce and represent forecasting data. Then this methodology is used to assess the influence of different factors in the predictability of a landslide. ~~t~~The main results of ~~the~~ such study are the following:

- Prediction plots are introduced as graphs showing the evolution of collapse forecasts with time. Such plots provide more information than simple average and standard deviation of the forecasts and improve the reliability of the final prediction.
- A predictability index (*PI*) has been introduced as a scoring system based on the description of the prediction plot, in order to evaluate the quality of a set of predictions.
- The predictability of a landslide depends firstly on its kinematics and then on what determines it (geology, external forces, local effects etc.).
- Landslide collapses can be forecasted whether they are in highly or lowly brittle materials, in rock or in earth material, of different types, with different sliding surface geometries, volumes and triggers.
- Contrarily to what is generally assumed (Voight, 1988; Rose and Hungr, 2007), landslides can be forecasted also with external forces acting.
- The asymptotic behaviour of the inverse velocity curve does not imply that the landslide cannot be correctly forecasted, even though it can hinder the prediction.
- The asymptotic behaviour may be induced by external factors, lithology and local effects, rather than only by crack propagation. In fact asymptotic trends have been found in first time failures and in both brittle and lowly brittle materials. The crack propagation explanation is not neglected, but it may not represent the general rule.
- Most recent displacement monitoring data increase the confidence when estimating the time of failure but do not necessary provide more accurate predictions than the older ones (provided that they start from after the initiation of the tertiary creep).
- The developed approach integrates more forecasting methods to further improve the reliability of the prediction.

AUTHOR CONTRIBUTION

E. Intrieri developed the idea and performed the analyses. G. Gigli supervised and improved the manuscript.

ACKNOWLEDGEMENTS

The authors are thankful to Antonio Intrieri for his important technical contribution when computing the calculations needed for this work.
No competing financial interests exist.

REFERENCES

Angeli, M-G., Gasparetto, P., Pasuto, A. and Silvano, S.: Examples of landslide instrumentation (Italy). In: Proceedings of 12th International Conference on Soil Mechanics and Foundation Engineering, Rio de Janeiro, Brazil, 3:1531-1534, 1989.

Azimi, C., Biarez, J., Desvarreux, P. and Keime, F.: Prévision d'éboulement en terrain gypseux. In: Bonnard C, Balkema AA (eds) Proceedings of 5th International Symposium on Landslides, Lausanne, Rotterdam; 1:531–536, 1988. In French.

Baum, R. L. and Godt, J. W.: Early warning of rainfall-induced shallow landslides and debris flows in the USA. Landslides, 7:259-272, 2010.

Blikra, L.H.: The Åknes rockslide: Monitoring, threshold values and early-warning. 10th International Symposium on Landslides and Engineered Slopes, 30th Jun - 4th Jul, Xian, China, 1089-1094, 2008.

Carlà, T., Intrieri, E., Di Traglia, F., Nolesini, T., Gigli, G., Casagli, N.: Guidelines on the use of inverse velocity method as a tool for setting alarm thresholds and forecasting landslides and structure collapses. Landslides, 1-18, 2016. DOI 10.1007/s10346-016-0731-5.

Casagli, N., Tibaldi, A., Merri, A., Del Ventisette, C., Apuani, T., Guerri, L., Tarchi, D., Fortuny-Guasch, J., Leva, D. and Nico, G.: Deformation of Stromboli Volcano (Italy) during the 2007 crisis by radar interferometry, numerical modeling and field structural data, Journal of Volcanology and Geothermal Research, 182:182-200, 2009.

Crosta, G.B. and Agliardi, F.: Failure forecast for large rock slides by surface displacement measurements. Can. Geotech. J., 40:176-91, 2003.

Dick, G.J., Eberhaardt, E., Cabrejo-Liévano, A.G., Stead, D. and Rose, N.: Development of an early-warning time-of-failure analysis methodology for open-pit mine slopes utilizing ground-based slope stability radar monitoring data. Can. Geotech. J., 52:515-29, 2015.

Dusseault, M. B. and Fordham, C. J.: Time-dependent behavior of rocks. Chapter 6, in Hudson, J.A. ed., Comprehensive Rock Engineering 4, Pergamon Press, 119-149, 1994.

Fukuzono, T.: A method to predict the time of slope failure caused by rainfall using the inverse number of velocity of surface displacement, Journal of Japanese Landslide Society, 22:8-13, 1985a.

Fukuzono, T.: A new method for predicting the failure time of a slope failure. In: Proceedings of 4th International Conference and Field Workshop on Landslides, Tokyo, Japan, 145-150, 1985b.

Gigli, G., Fanti, R., Canuti, P. and Casagli, N. Integration of advanced monitoring and numerical modeling techniques for the complete risk scenario analysis of rockslides: The case of Mt. Beni (Florence, Italy). Engineering Geology, 120(1-4):48–59, 2011.

Hong, L. and Page, S. E.: Some microfoundations of collective wisdom. In: Landemore H, Elster J (eds) Collective Wisdom Principles and Mechanisms, Cambridge University Press, 392 p, 2008.

Hutchinson, J. N.: Landslide risk - to know, to foresee, to prevent, Geologia Tecnica e Ambientale, 9:3-24, 2001.

IEEIRP (Independent Expert Engineering Investigation and Review Panel): Report on Mount Polley Tailings Storage Facility Breach. Province of British Columbia. 156 p, 2015. https://www.mountpolleyreviewpanel.ca

Intrieri, E., Gigli, G., Casagli, N. and Nadim, F.: Brief communication: Landslide Early Warning System: Toolbox and General Concepts. Natural Hazards and Earth System Science, 13:85-90, 2013.

Iovine, G., Petrucci, O., Rizzo, V. and Tansi, C.: The March 7th 2005 Cavallerizzo (Cerzeto) landslide in Calabria - Southern Italy. In: Proceedings of 10th IAEG Congress, Nottingham, Great Britain, Geological Society of London, 785:1-12, 2006.

Jordan, T., Chen, Y-T., Gasparini, P., Madariaga, R., Main, I., Marzocchi, W., Papadopoulos, G., Sobolev, G., Yamaoka, K. and Zschau, J.: Operational Earthquake Forecasting: State of Knowledge and Guidelines for Implementation, Annals of Geophysics, 54(4):316-391, 2011.

Lacasse, S. and Nadim, F.: Landslide risk assessment and mitigation strategy. In: Sassa K, Canuti P (eds) Landslides - Disaster Risk Reduction, Springer-Verlag Berlin Heidelberg, 31-61, 2009.

Manconi, A. and Giordan, D.: Landslide early warning based on failure forecast models: the example of the Mt. de La Saxe rockslide, northern Italy. Nat. Hazards Earth Syst. Sci., 15:1639-1644, 2015.

Mufundirwa, A., Fujii, Y. and Kodama, J.: A new practical method for prediction of geomechanical failure-time. In: International Journal of Rock Mechanics & Mining Sciences, 47(7):1079-1090, 2010.

Page, S. E.: The difference: how the power of diversity creates better groups, firms, schools, and societies. Princeton University Press, 424 p, 2007.

Petley, D. N.: The evolution of slope failure: mechanisms of rupture propagation, Natural Hazards and Earth System Sciences, 4:147-152, 2004.

Petley, D. N., Bulmer, M. H. and Murphy, W.: Patterns of movement in rotational and translational landslide. Geology, 30:719-722, 2002.

Petley, D. N., Petley, D. J. and Allison, R. J.: Temporal prediction in landslides - Understanding the Saito effect. In: Proceedings of 10th International Symposium on Landslides and Engineered Slopes, Xian: China, 865-871, 2008.

Rose, N. D. and Hungr, O.: Forecasting potential rock slope failure in open pit mines using the inverse velocity method. International Journal of Rock Mechanics and Mining Science, 44(2):308-320, 2007.

Saito, M.: Forecasting time of slope failure by tertiary creep. In: Proceedings of 7th International Conference on Soil Mechanics and Foundations Engineering, Montreal, Canada, Pergamon Press, Oxford, Great Britain, 667-683, 1969.

Semenza, E. and Melidoro, G.: Proceedings of the meeting on the 1963 Vaiont landslide. In: Semenza, E. and Melidoro., G. (eds.) Proceedings of the meeting on the 1963 Vaiont landslide, 1986; Ferrara, Italy. IAEG Italian Section, University of Ferrara, 1:1-218, 1992.

UNISDR (United Nations International Strategy for Disaster Reduction): Terminology on Disaster Risk Reduction, 13 p, 2007.

Voight, B. A.: Method for prediction of volcanic eruption. Nature, 332(10):125-130, 1988.

Willoughby, H. E., Rappaport, E. N. and Marks, F. D.: Hurricane Forecasting: The State of the Art. Natural Hazards Reviews, 8(3):45-49, 2007.

Zavodni, Z. M. and Broadbent, C. D.: Slope failure kinematics. Canadian Institute of Mining, Metal Petroleum (CIM) Bulletin, 73(16):69-74, 1980.

Major Revision
Editor Decision: Reconsider after major revisions (further review by Editor and Referees) (20 Sep 2016) by Dr. Thom Bogaard
Comments to the Author:
Dear authors

you received two reviews, both agreeing it is an interesting piece of work which can make a significant scientific contribution. However, the description of your novel method and the used definitions can and should be written in a more formal and precise way. Hereto, reviewer #1 gave several clear suggestions and comments. Also the points of 'confidence' and 'probability' is something to clarify. This will increase the impact of your work significantly.
Please address the comments of both reviewers very accurately in the revised version. The authors are also advised to double check their new text for writing style.

I look forward receiving the resubmission

Kind regards
Thom Bogaard

First of all the Authors want to thank the reviewers and the Editor for their dedication. We find that the paper now is much clearer and more formally correct and that the new additions and changes helped to improve the overall quality of this work.

Submitted on 15 Sep 2016
Anonymous Referee #1
Anonymous during peer-review: Yes
Anonymous in acknowledgements of published article: Yes

Recommendation to the Editor
1) Scientific Significance
Does the manuscript represent a substantial contribution to the understanding of natural hazards and their consequences (new concepts, ideas, methods, or data)?
Excellent Good Fair Poor
2) Scientific Quality
Are the scientific and/or technical approaches and the applied methods valid? Are the results discussed in an appropriate and balanced way (clarity of concepts and discussion, consideration of related work, including appropriate references)?
Excellent Good Fair **Poor**
3) Presentation Quality
Are the scientific data, results and conclusions presented in a clear, concise, and well-structured way (number and quality of figures/tables, appropriate use of technical and English language, simplicity of the language)?
Excellent Good **Fair** Poor

For final publication, the manuscript should be accepted as is.

accepted subject to technical corrections.

accepted subject to minor revisions.

reconsidered after major revisions:

I would like to review the revised paper.

I would NOT be willing to review the revised paper.

rejected.

Please note that this rating only refers to this version of the manuscript!

Suggestions for revision or reasons for rejection (will be published if the paper is accepted for final publication)

General comment:

The paper fails to formally propose a methodology to increase the reliability of landslide forecasting based on displacement monitoring. The approach is presented at the end of the manuscript, as part of the discussion, and with no clear explanation of the sequence of analyses and criteria that should accompany a proposed methodology. Moreover, a probabilistic approach is mentioned, however there are no formal probabilistic techniques or reliability methods formulated that leads to quantified reliability. Plots of average times to failure and standard deviation does not fully address a probabilistic approach.

The findings presented by the authors are important. The databases the authors present are very valuable. The analyses presented associated with their interpretation of the prediction tools and how to compare them are also valuable and worth publication. The analyses presented associated with the application of probabilistic techniques to landslide forecasting reliability, present the necessary data, however they are immature for publication and require further work. The writing style of some paragraphs, in particular the new additions, is not technical and sections of the manuscript are far from NHESS standards.

Particular comments. Line numbers correspond to the file with the authors responses, which include the track changes to the original document submitted.

Title - I suggest the title of the manuscript should not start with a preposition.

The title has been changed accordingly.

L54-59 This paragraph raises an issue. How is failure defined in the paper? Is it first movement? Rupture? I understand the authors refer to the onset of sudden acceleration and collapse, and this should be clearly stated.

This has been now specified.

L67-72 sufficient lead time for what?

For evacuation. Now it has been specified.

What is noise? These require clarification. Do you mean measurement fluctuations around a trend, with a natural origin or caused by monitoring instruments?

This is already specified: “natural or instrumental noise” mean exactly “natural origin or caused by monitoring instruments”. Furthermore the citation at the end of the sentence is referred to a paper that deals in detail with these kinds of noise. This reference has now been evidenced.

Data treatment means post-processing?

Yes, now it is specified.

“exploit such methods” which methods? Data treatment?

No, it refers to forecasting methods. Since this was not clear it has been specified.

L121-127 This paper addresses the variability of predictions through the predictive models adopted but do not address a real “prediction rate”, or prediction-realization success. The method further assumes implicitly that fluctuations in the geomechanical behaviour of landslides can be captured by fluctuations in the predicted time of failure. These should be clearly stated at the start so the reader is aware of them.

The prediction-realization success is quantified by the parameter PI. Now we have added a sentence in the introduction to make the reader aware of it since the beginning of the paper. We do not think that the fluctuations in the predictions reflect the fluctuations in the geomechanical behaviour. There may be a lot of reasons why predictions are not always accurate, and other factors than geomechanics can hamper this accuracy, as deeply commented on in the paper.

L197-201 This figure shows the evolution of the predicted time of failure, however does not directly or clearly show how iterating forecasts increase confidence. This is explained in the text and should be removed from the caption of the figure.

This part has been added to clarify our point.

“Notably, if we consider, for example, only the results of the S method in the case of the Avran valley landslide, we see that since the end of September the forecasts are constantly furnishing a time of failure preceding the actual T_f. Although this may be considered a case of safe predictions (that is an error not producing a false positive and therefore not dangerous for the elements at risk), this also means that, at every forecast that is made, t_f is postponed. Given a series of ever increasing values of t_f, it is impossible to assess which of them (if any) is closer to the actual time of failure. However, if the time series of predictions is long enough, past forecasts (before early September) furnish values of t_f that, if averaged with the late ones, centre the value of T_f. Therefore it is clear how a prediction plot may allow decision makers to make more aware evaluations of the time of collapse of a landslide.”

And this sentence has been modified as follows:

“It is this convergence that permits to correctly forecast the collapse a priori at least since late November, i.e. a month before the failure, whereas a single forecast would not be able to give a confidence of the prediction.”

Furthermore the caption has been modified as suggested.

L281-294 This paragraph is unclear and our of place. Furthermore, the writing style is poor and far from NHESS standards. Authors are encouraged to read the manuscript and ensure a technical style of writing.

This paragraph has been heavily rewritten to improve the style and meet NHESS standards:

“In fact, even though geomechanics is unquestionably a key factor, a complete geomechanical characterization is often difficult to accomplish, especially in emergency situations. The clearer hints of a particular geomechanical behaviour are often derived from displacement data. Like in a black box model, even if the real properties of a phenomenon are not known, conclusions may be drawn from the output of those properties (i.e. the kinematics). In this case, importance has been done to kinematics because what is generally measured by monitoring are displacement data. Furthermore, many other

unknown factors (rainfall, ground saturation, earthquakes, anthropic disturbance etc.) are included in the black box model together with the geomechanics; this makes it virtually impossible to know in advance what may be the degree of influence of geomechanics alone with respect to other factors, thus leading to focusing on kinematics instead. Moreover, even though geomechanics is a key element in determining landslide predictability (for example because it is responsible for the creep behaviour), the results of the present study showed that landslide prediction can be carried out with a variety of different geomechanical settings."

The position of this paragraph is due to a comment made by another reviewer in the previous revision step, where the reviewer asked to explain more in detail the concept explained above.

L295-297 Contrary to the author's statement, prediction plots are not clear in this matter. Supplementary plots or adequate highlights within the plot would be required for the authors to derive this statement and the readers to clearly observed the author's observations.

In fact this sentence is not referred to the prediction plots but to the comments made in the discussion session. They can be easily observed by comparing table 1 (which furnishes evaluations concerning the geomechanical properties of the case studies) with table 2 (which states whether a collapse was predictable or not). This has been added to the text to make it easier to understand for the reader.

L309 The proposed methodology is presented at the end of the manuscript and as part of the discussion, when it should have been introduced early on the manuscript and then proved to the reader. In this methodology, the steps of "Study the shape of the prediction plot" and "inference about the time of failure" have no substance. Although the authors do study the plots and infer times of failure during their discussions of the prediction methods and plots, there is no clear sequence of analyses and criteria that should accompany a proposed methodology. I argue that this manuscript, as it is written, does not formally proposes a methodology for predicting landslide time of failure.

As it is now explained in the conclusions, the main aspect of our methodology concerns a way to produce and represent forecasting data. The reviewer here probably asks for a method to interpret such data and in particular to retrieve an estimate of Tf from the time series of tf. This is already furnished in the paper. We showed (figure 3) that the mean of the predictions can be a proxy for the time of failure. However we also have to note that this criterion does not employ all the information derived from a prediction plot and in fact it sometimes furnishes predictions that not accurate enough. Other indicators have been adopted (mode, average between maximum and minimum etc.) but unfortunately it appears that there is no quantitative or univocal method to calculate the a good estimate of Tf, also because not every prediction plot displays the characteristic oscillations. Therefore we have found that the last part of the procedure must be left to expert judgement, as indicated in the figure 4 where we state "inference about the time of failure". This expert judgement is however informed with quantitative and redundant data that are much more reliable than a single forecast, even if this single forecast might be completely derived from a quantitative computation. In fact Fukuzono users typically adopt a quantitative method to extrapolate a single time of failure forecast; however this does not improve the general accuracy of the prediction; instead it gives a false confidence in the user. Instead we give decision makers a tool to critically assess the "weight" of such forecast by comparing it with many others. Here stands the core of our methodology. Moreover this should not

divert the attention from the scope of our paper that is also to use this tool to assess the influence of different factors in the predictability of a landslide.

Nevertheless we recognize that individuating a parameter or an equation that can synthesize the prediction plot in a single number would be an improvement to this methodology. In fact we are currently developing our research in this direction and we look forward into publishing our findings as soon as we achieve the result.

We have explained all these concepts in the discussions as follows:

"Figure 3 shows that the mean of the predictions can be used as a proxy for the time of failure but, as stated above in this paragraph, it is also shown that the obtained accuracy may not be enough as the mean does not exploits all the information provided by a prediction plot. Other statistical indicators have been attempted but none of them appeared to better approximate the value of Tf, mainly due to the difficulty of accounting for the important time factor in the forecasts and also because not every prediction plot displays the characteristic oscillations. Therefore, the interpretation of the prediction plot (and in particular of the dispersion of the forecasts with time) represents the most valuable tool for decision makers, who, in this way, can make aware judgements informed with a large set of quantitative and redundant data and therefore assessing the "weight" of a single prediction by comparing it with many others."

As requested we anticipated the explanation of our procedure in the method paragraph, although figure 4 cannot be moved to an earlier paragraph since it refers to the concept of PI that is introduced only later in the text.

L329-332 This can not be concluded from the data that is presented. You need an adequate presentation of the characteristics of some case studies and their displacement time series that support your conclusion.

This conclusion is drawn from those landslides cited in Table 2 that displayed an asymptotic trend and also moderate values of PI. In particular Liberty Pit gives very good forecasts with F method (figure 2). This is why we conclude that even if the landslide displays an asymptotic trend it can still be forecasted.

Some editorial comments:

L15 should read "is presented here"

L17 "about reliability of prediction"

L22 "are the key"

L23 remove the word "when"

L34-37 needs revisiting. The use of the word "particular" is abused.

L38 Should refer to the paper not to "our research". Landslide instead of landslides

L67 Creep behaviour not behaviours

L119 grammar: "at most more"

L274-280 This paragraph is unclear and requires re-writing.

This paragraph has been heavily rewritten as follows. This new version also helps to explain issues raised in the previous review.

"All these case histories show that the main responsible for the predictability of a landslide, and secondary also for the presence or not of the "Saito effect", is in a way or another connected to geology. However this relation is not simple nor direct. Instead both the predictability and the "Saito effect" depend on the kinematics of the landslide, since only a landslide accelerating with a certain trend can be forecasted using S, F and M methods. Naturally, the kinematics in turn depend on the geological conditions. In the complex

relation between geology and kinematics the aforementioned factors may intervene. Although their interaction may not be known, its effect on displacement data can be easily measured. As a result it has been found that asymptotic trends in the inverse velocity plot have been encountered also for first failure ruptures (as found in some time series of Mount Beni landslide), contrarily to what is described by Petley (2004). This can be explained as an effect of those interactions which may alter in an unknown way the normal relation between geology and kinematics, thus making focusing on kinematics as the key more reliable than relying on geology alone."

All the editorial comments have been addressed.

Report #2

Submitted on 18 Sep 2016

Anonymous Referee #2

Anonymous during peer-review: Yes

Anonymous in acknowledgements of published article: Yes

Recommendation to the Editor

1) Scientific Significance

Does the manuscript represent a substantial contribution to the understanding of natural hazards and their consequences (new concepts, ideas, methods, or data)?

Excellent Good Fair Poor

2) Scientific Quality

Are the scientific and/or technical approaches and the applied methods valid? Are the results discussed in an appropriate and balanced way (clarity of concepts and discussion, consideration of related work, including appropriate references)?

Excellent **Good** Fair Poor

3) Presentation Quality

Are the scientific data, results and conclusions presented in a clear, concise, and well-structured way (number and quality of figures/tables, appropriate use of technical and English language, simplicity of the language)?

Excellent **Good** Fair Poor

For final publication, the manuscript should be

accepted as is.

accepted subject to technical corrections.

accepted subject to minor revisions.

reconsidered after major revisions:

I would like to review the revised paper.

I would NOT be willing to review the revised paper.

rejected.

Please note that this rating only refers to this version of the manuscript!

Suggestions for revision or reasons for rejection (will be published if the paper is accepted for final publication)

Dear Editor,

Please find here below my review of the paper nhess-2016-221 v2:

Of reliable landslide forecasting and factors influencing predictability

By

Emanuele Intrieri, Giovanni Gigli

The new version was greatly improved, the authors followed most the reviewers' comments, and they clarified most of the unclear statements.

The paper is nearly ready for publication, but in my opinion, one problem remains. The confidence is not well defined, if I understand well it is different of PI, and then the confidence can be used for the forecast. It is stated line 118: that "…confidence (for example given by the standard 118 deviation of tf)." This is not really discussed or introduced in the rest of the text except in conclusions and figures. This must be clarified for the final version.

Thank you for raising the problem. This sentence has been added in the method section in order to explain it better:

"Confidence may be quantitatively assessed by using the standard deviation of the forecasts as a proxy. In fact the standard deviation furnishes the dispersion (i.e. the precision) of the predictions, which may be used to calculate a time window within which the collapse is more likely to occur. Therefore the lower the standard deviation of a set of forecasts, the higher would be their reliability and the confidence.

On the other hand, confidence it may also be considered as a qualitative increase in the awareness of the decision makers that can estimate the time of failure of a landslide by evaluating a large set of different predictions and their dispersions."

Last point is the problem of overlap of the graph and text in figure 3. This must be corrected.

It has been corrected.