Brief communication: ~~Rainfall thresholds based on Artificial neural networks can improve landslide early warning~~Introducing rainfall thresholds for landslide triggering based on artificial neural networks

Pierpaolo Distefano[1], David J. Peres[1*], Pietro Scandura[1] and Antonino Cancelliere[1]

[1]Department of Civil Engineering and Architecture, University of Catania, Catania, 95123, Italy

**Correspondence*: David J. Peres (djperes@dica.unict.it)

Abstract. In this communication, we show how the use of artificial neural networks (ANNs) can improve the performance of the rainfall thresholds for landslide early warning. Results for Sicily (Italy), show how performance of a traditional rainfall event duration and depth power law threshold, yielding a true skill statistic (TSS) of 0.50, can be improved by ANNs (TSS = 0.59). Then we show how ANNs allow to easily add other variables, like peak rainfall intensity, with a further performance improvement (TSS = 0.~~64~~66). This may stimulate more research on the use of this powerful tool for deriving landslide early warning thresholds.

Introduction

Landslides triggered by rainfall can cause ~~damage~~ on infrastructures, buildings, and in the worst scenario, even human ~~loss~~ (Froude and Petley, 2018). Commonly, rainfall thresholds indicating the conditions under which landslides can potentially occur, are a key component of ~~-~~warning systems aimed ~~should may be issued to~~at protecting the population from a possible landslide event. In most of the cases, thresholds~~,~~ are determined using empirical methods that link characteristics of precipitation, such as duration D and mean intensity I or cumulated rainfall ~~H~~ $E = I \times D$, to landslide occurrence (Caine, 1980). ~~(Guzzetti et al., 2008; Caine, 1980)~~. Rainfall thresholds are generally determined by assuming a predetermined parametric equation, which in most of the cases is a power law (Guzzetti et al., 2008). Such a constraint can potentially limit the predictive performance of the thresholds, because the informative content of the considered explanatory variables may not be exploited at fullest. This holds true ~~even~~ all the more so when searching for alternative or additional variables with the aim at improving the performances of the thresholds, such as antecedent rainfall conditions (Glade et al., 2000), water storage and soil moisture data (Bogaard and Greco, 2018; Marino et al., 2020). For the case of E-D or I-D thresholds the use of a power law is customary and its rationale has been also verified based on a combined stochastic and physics-based approach (Peres and Cancelliere, 2014). On the contrary, either in the case of a different pair of variables or the analysis of more than two variables, there is no analogous prominent parametric form of the threshold equation. For instance, as reported by Conrad et al., (2021), alternative formulas have been considered for hydrometeorological thresholds – i.e., based on rainfall and soil moisture or catchment storage – , including linear and bilinear functions, interpolated line segments without a mathematical function, cut-off values

for integration of antecedent conditions with traditional rainfall thresholds, and more complex logical operators. The choice of a predetermined threshold equation form can potentially limit the performance of the threshold derivable from the given set of variables, and thus also limit the scientific soundness of the comparison between different approaches for deriving landslide triggering thresholds. Artificial Neural Networks (ANNs), belonging to Artificial intelligence or Machine learning techniques, ~~are a very flexible tool, that~~ allow to potentially remove the mentioned limitation of having to choose a predetermined parametric threshold form~~s~~, as they are ~~capable to reproduce a vast range of non-linear classifiers~~universal approximators, i.e. capable of reproducing any continuous function (Haykin, 1999).

Up to now, a number of studies have used the potentiality of ANNs and of other machine learning techniques in landslide ~~risk~~ analysis. Many studies focused on susceptibility mapping and individual slope instability have exploited the potentialities of ANNs (Reichenbach et al., 2018). ~~. For instance, used ANNs to map landslide susceptibility based on nine predisposing factors, and then combined the results with a simplified model for landslide runout prediction.~~ In other studies, the focus is on the prediction of individual deep seated landslide displacements by machine-learning algorithms using detailed in situ data (e.g., van Natijne et al., 2020) ~~(Cao et al., 2016; Krkač et al., 2017; Miao et al., 2018;. Among these applications, described and listed the available parameters for displacement prediction from radar and remote sensing technology (slope, geology, soil moisture, precipitation/snow melt, land use), and how they may be used within a local early warning system based on machine learning techniques, such as ANNs.~~

Based on ~~the As shown in~~ this ~~short~~ briefly outlined state-of-the-art~~literature review~~, it appears that ANN skills are mainly used to create susceptibility maps and/or in local early warning systems, while application for territorial landslide early warning (Piciullo et al., 2018) has not been investigated so far. In this communication, we present our preliminary investigations showing how ANNs can allow to derive landslide early warning thresholds with higher performances than traditional rainfall intensity-duration~~—depth~~ power law thresholds.

Data and methods

We refer to the case study of Sicily (Fig. 1), one of the 20 regions of Italy ~~(Fig. 1)~~. We have retrieved hourly rainfall from 306 rain gauges distributed within the region, managed by the Regional water observatory (Osservatorio delle Acque, OdA), the SIAS (Sicilian Agro-meteorological Information Service), and by the Regional Civil Protection Department (DRPC). Fig. 1 shows the rain gauge locations for the period 01/–2009-10/2018 (~~red~~ green dots) and those available only for the period 01/2014-10/2018 (black dots).

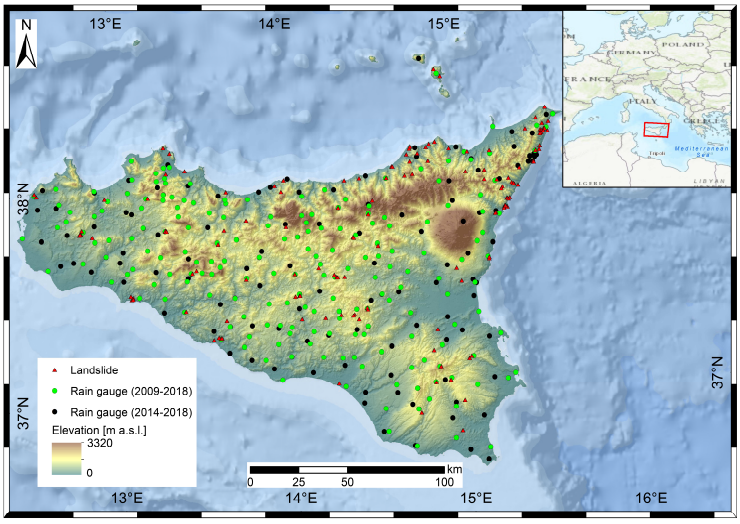

Figure 1: Elevation map showing location of landslides and rain gauges in Sicily ~~used in this study.~~ The rainfall dataset was built by joining dataset managed by different authorities and landslides from the Franeitalia inventory (Calvello and Pecoraro, 2018).

Landslide~~s~~ data ~~comes~~ is retrieved from the FraneItalia database compiled by Calvello and Pecoraro (2018) – see locations on Fig. 1. This database ~~originally contained~~ contains information on landslides occurred in Italy from January 2010 to December ~~2017~~2019~~.~~ ~~The database has been recently updated with landslides occurred in 2018-2019~~and is available online (https://franeitalia.wordpress.com/database/, last accessed on ~~29~~17/~~06~~11/2021). Thus, our analysis is based on the period from January 2010 to October 2018, where both rainfall and landslide information is available. Some landslide events have been removed from the analysis. In particular, this was done based on the ~~The information within this data base concern landslides triggered by rain but also those triggered by anthropogenic causes and earthquakes.~~

-fields included in FraneItalia that characterize the observed landslide events – typology, material and trigger. Only events having "rainfall" o "rainfall and other" trigger have been considered, ~~so~~ to exclude landslides due to earthquakes and anthropogenic activities. Also, events of the "fall" typology combined with "rock" material have been removed from the analysis, as in the case of rockfalls, rainfall may have a triggering role different from the other types of landslides.

~~A flow chart of the applied methodology is shown in Fig. 2a. After collecting the data, some preprocessing has been carried out. In particular, landslides triggered by different precursors than rainfall were removed from our analysis.~~ Rainfall data have been checked ~~so Also, to~~ remove suspicious rainfall data ~~has been removed~~. In particular, where hourly rainfall exceeded 250 mm – corresponding to about one third of mean yearly rainfall for Sicily and to about two times the maximum rainfall ever recorded in 1 hour – the series has been visually inspected, and in the case of an evident error (rain gauge malfunction), the whole rainfall event surrounding the peak has been removed. ~~A~~ flow chart ~~of~~ the applied methodology is shown in Fig. 2a.

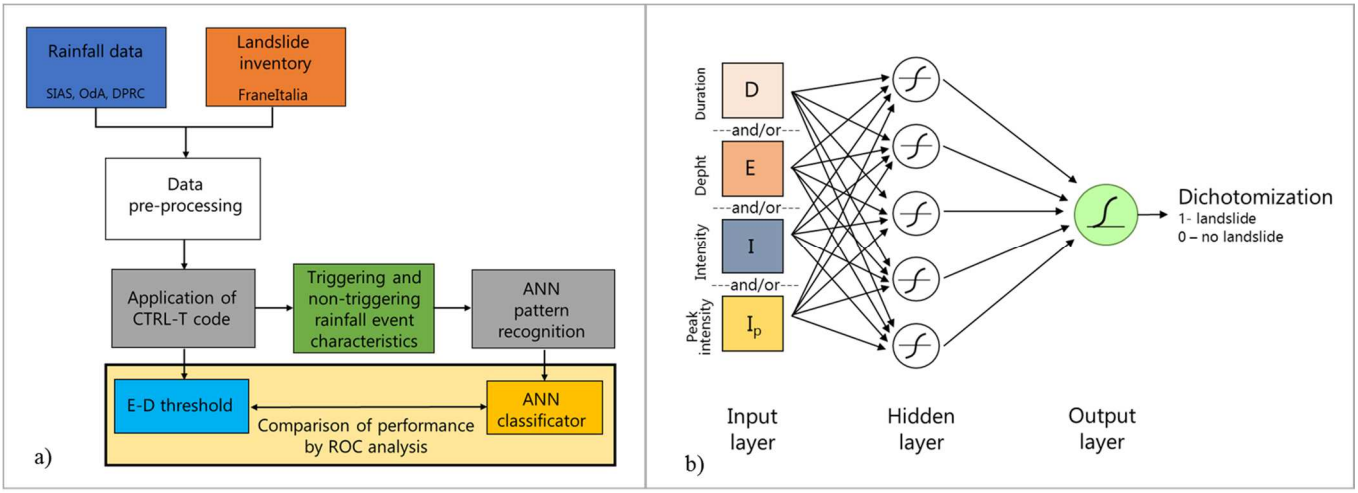

Figure 2: Flow chart illustrating the methodology (a) and the Artificial neural network architecture considered (b).

First, ~~P~~pre-processed precipitation and landslide data were inputted to the CTRL-T (Calculation of Thresholds for Rainfall-induced Landslides-Tool) code (Melillo et al., 2018). The software consists of a code in R language, and allows to reconstruct rainfall events and characterizing them by the following variables: duration D, mean intensity I, total depth ~~H~~ E = D×~~-~~I and peak intensity I_p (defined as the maximum hourly intensity occurring during a rainfall event). The most probable rainfall conditions associated to each landslide (multiple rain gauges available for a given location) event are computed by the software based on distance between rain gauge and the landslide location, and the characteristics of the reconstructed rainfall event. In particular, for a given landslide, all rain gauges within a circle of radius R_b specified by the user are searched and, when more than one rain gauge is located within the circle, the rainfall events from each rain gauge are weighted based on the rain gauge-landslide distance and the rainfall event characteristics (cumulated rainfall and duration). T~~t~~he weight is used to estimate the "probability" associated to each rainfall condition potentially attributable to each landslide event. ~~In particular, the probability,I~~ ~~i~~n case of multiple rainfall conditions, the probability of the individual event is computed by dividing ~~each~~its -weight to ~~their~~the sum of concurrent events' weights. CTRL-T then determines the triggering rainfall conditions of each landslide as those corresponding to the highest probability. When the triggering instant is after the end of the rainfall event, the most probable triggering rainfall conditions are computed considering the whole event, otherwise the event is truncated at the triggering instant. -Finally, the code provides power-law ~~H~~E-D thresholds for different levels of non-exceedance frequency of triggering events. The software allows the user to set different values of the parameters to reconstruct rainfall events in order to take into account seasonality, i.e., different average evapotranspiration rates in different periods of the year. In particular, following the study by Melillo et al. (2016), we assumed that in the warm season C_W (April – October) the minimum dry period separating two rainfall events is of P_{4warm} = 48 hours, while in the cold season a longer period is assumed (P_{4cold} = 96 hours). The rain gauge sensitivity is G_s = 0.1 mm. The rain gauge search radius has been fixed to R_b =16 km. A binary coding has been attributed each rainfall event~~s~~, flagging triggering events as a target with value of 1 and a non-triggering event with null value.

Application of the CTRL-T software allowed to reconstruct the rainfall events associated to the 144 landslide events in the inventory (triggering events) and 47398 non-triggering events. ~~Application of the CTRL-T software yielded 144 triggering rainfall events and 47398 non-triggering events~~. For 103 events, only the day of triggering was known, while for the remainder a more precise indication of the triggering instant was available. In the first case, the triggering instant was attributed to the end of the day, in the second case to the instant of peak rainfall within the time interval when the triggering has occurred. ~~For~~Furthermore, for the 144 landslide events detailed information on the typology was available only in 18 cases, 10 of which were "fall" of "more than one material", 4 "flow" and other 4 "slide". The average distance between rain gauge and landslide for the 144 events is about 5 km, thus seldom the maximum value of R_b= 16 km was reached. ~~-~~

The characteristics of the events were used as input variables to ANNs devised for pattern recognition, as implemented within the Neural Net Pattern Recognition tool in MATLAB. The neural network, characterized by a feed-forward structure (Fig 2b), is composed of three layers: input, hidden and output. Two different activation functions have been considered: a tan-sigmoid function $f(n)$ for the hidden layer, and a log-sigmoid $g(n)$ for the output layer:

$$f(n) = \frac{2}{(1+e^{-2n})} - 1 \qquad (1)$$

$$g(n) = \frac{1}{(1-e^{-n})} \qquad (2)$$

~~For the derivation of thresholds based on ANNs, the following input variable configurations have been investigated: 1) *D*; 2) *E*; 3) *I*; 4) I_p; 5) *D* and *E*; 6) *D* and *I*; 7) *D* and I_p; 8) *E* and I_p; 9) *I* and I_p; 10) *D*, *E* and I_p. The listed input configurations are indeed all possible ones, except those combining both *E* and *I* with duration *D*. This has been done because the two pairs *D*-*I* and *D*-*E* have the same informative content by construction, as confirmed by the fact that the performances of the *D*-*I* and *D*-*E* neural networks do not differ significantly (see later, Tab. 1) – slight differences may occur as ANNs can be sensitive to how a set of variables having together the same information content of another are presented to the network. All the data have been inputted taking their natural logarithms. Different networks have been considered varying the number of hidden neurons from 5 to 20, in order to search for the best value, i.e., the one yielding the highest TSS.~~

~~The entire dataset of rainfall events was divided into a training, a validation, and a test data set, selected randomly from the entire dataset, in the proportions of 70%, 15% and 15%. The training dataset is data used to fit the model; while the validation provides an unbiased evaluation of a model fit on the training dataset while tuning model hyper parameters (, such as the number of training iterations). Finally, the test dataset provides an unbiased evaluation of a final model fit. This subdivision allowed to apply the early stopping criterion to prevent overfitting. According to this criterion, the training of the neural network is stopped when the values of the performance function calculated on the validation dataset start to get worse. It is noteworthy to mention that, in order to ensure representativeness of the data randomly assigned to the training, validation and test datasets, results where the TSS in the test or the validation data set were greater than the TSS in the training data set have been removed from the analysis. Once the network is developed by a procedure considering these three datasets and early stopping, it is "frozen" and ROC metrics (e.g., TSS) can be computed with that network on the entire dataset, and the corresponding performances arecan be considered "generalizable". Thus, Wwhen comparing our proposed approach with the~~

traditional one, we focus on these lastse performances (labeled as "all"). This ,seems the most appropriate way to proceed, as the *I D* power law and its performance is determined with respect to the entire dataset, and thus more representative of the performances that can be expected from the approach we propose.

This subdivision allowed to apply the early stopping criterion to prevent overfitting. According to this criterion, the training of the neural network is stopped when the values of the performance function calculated on the validation dataset start to get worse. The ANNs have been trained through the scaled conjugate gradient backpropagation algorithm, while cross-entropy was assumed as the performance function for training. Denoting the generic ANN output with y_i (assuming values in the open interval between 0 and 1) and the binary target with t_i, i =1,2, …, N, the cross-entropy function F heavily penalizes inaccurate predictions and assigns minimum penalties for correct predictions:

$$F = -\frac{1}{N}\sum_{i=1}^{N}[t_i \log y_i + (1 - t_i)\log(1 - y_i)] \quad (3)$$

The ability to distinguish triggering events from non-triggering events was measured using the confusion matrix, a double-entry table in which it is possible to identify true positive TP (triggering events correctly classified), true negative TNs (non-triggering events correctly classified), the false negative FN (triggering events classified as non-triggering) and FP false positive (non-triggering events classified as triggering). Using Through the confusion matrixmatrix it is possible to determine the True Positive Rate (TPR) and the False Positive Rate (FPR), as well as their difference, known as the True skill statistic (TSS), which is widely used for threshold determination (Peres and Cancelliere, 2021):

$$TPR = \frac{TP}{TP+FN} \quad (4)$$

$$FPR = \frac{FP}{TN+FP} \quad (5)$$

$$TSS = TPR - FPR \quad (6)$$

For our analysis different combination of input data (duration *D*, intensity *I*, total depth *H E* and peak intensity I_p) and different architectures, changing number of hidden neurons were tested. The output of the ANNs is transformed into a binary code, by the application of a a dichotomization threshold. We then identify the threshold maximizing the TSSVarying the threshold, a Receiver Operating Characteristics (ROC) curve is derived (TPR vs FPR), and the threshold maximizing TSS is identified. Maximization of TSS implicitly assumes that all entries of the confusion matrix have the same importance. Quantifying how more important is a false negative respect to a false positive, is a complex task that goes beyond the aim of the present analysis (cf. Sala et al., 2021).

Results from ANNs are compared with rainfall duration-depth power-law thresholds derived through the maximization of TSS – i.e., again, analysing both triggering and non-triggering events.

For our analysis, different combination of input data (duration *D*, intensity *I*, total depth *E* and peak intensity I_p) and different architectures, changing number of hidden neurons were tested. In particular, the following input variable configurations have been investigated: 1) *D*; 2) *E*; 3) *I*; 4) I_p; 5) *D* and *E*; 6) *D* and *I*; 7) *D* and I_p; 8) *E* and I_p; 9) *I* and I_p; 10) *D*, *E* and I_p. The listed input configurations are indeed all possible ones, except those combining both *E* and *I* with duration *D*. This has been done because the two pairs *D-I* and *D-E* have the same informative content by construction, as confirmed by the fact that the

performances of the *D-I* and *D-E* neural networks do not differ significantly (see later, Tab. 1) – slight differences may occur as ANNs can be sensitive to how a set of variables is presented to the network even though the information content is equivalent under a mathematical point of view. All the data have been inputted taking their natural logarithms. Different networks have been considered varying the number of hidden neurons from 5 to 20, in order to search for the best value, i.e., the one yielding the highest TSS.

The entire dataset of rainfall events was divided into a training, a validation, and a test data set, selected randomly from the entire dataset, in the proportions of 70%, 15% and 15%. The training dataset is used to fit the model, whereas the validation provides an unbiased evaluation of a model fit on the training dataset while tuning model hyper-parameters, such as the number of training iterations. Finally, the test dataset provides an unbiased evaluation of a final model fit. This subdivision allowed to apply the early-stopping criterion to prevent overfitting. According to this criterion, the training of the neural network is stopped when the values of the performance function calculated on the validation dataset start to get worse. In order to ensure representativeness of the training, validation and test datasets, when the TSS in the test or the validation data set is ~~were~~~~are~~greater than the TSS in the training data set, ~~and~~ a new training is carried out with a different random data split Once the network is developed considering these three datasets and early stopping, it is "frozen" and ROC metrics (e.g., TSS) can be computed with that network on the entire dataset, and the corresponding performances can be considered generalizable. Thus, when comparing our proposed approach with the traditional one, we focus on these last performances (labeled as "all"). This seems the most appropriate way to proceed, as the *I-D* power law and its performance is determined with respect to the entire dataset.

Results and discussion

Application of the CTRL-T software has allowed to build the dataset of triggering and non-triggering events and to derive the threshold according to the so-called frequentist method (based on triggering events only). Considering a non-exceedance frequency for triggering events equal to 5%, threshold from the software is as follows:

$$\~~HE~~ = 4.9D^{0.26} \quad (7)$$

This threshold is lower than the one obtained for Sicily by Gariano et al. (2015), yet comparable with an updated one ~~-~~derived by Melillo et al. (2016) ~~through an earlier version of the algorithm that was then implemented by CTRL-T software~~. Specifically, thresholds reported on the mentioned two studies are respectively the following (non-exceedance frequency is again 5%):

$$\~~HE~~ = 10.4D^{0.27} \quad (8)$$

$$\~~HE~~ = 5.6\, D^{0.40} \quad (9)$$

It should be mentioned that these thresholds were both derived from rainfall datasets covering the period July 2002-December 2012, which is different from the one we have considered in our analysis. The first threshold has been derived with an earlier

version of the CTRL-T code, which required manual selection of the most representative rain gauge (Melillo et al., 2015), while the second study derives from the updated algorithm, where this selection is made automatically.

These thresholds shown above however are not comparable with those to derive with the proposed ANN approach, because non-triggering events are neglected. We have hence derived the power-law threshold corresponding to the maximum TSS – externally to the CTRL-T software, via MATLAB® global optimization toolbox –, obtaining the following result:

$$HE = 2.40D^{0.68} \quad (10)$$

that has a TSS = TSS_0 = 0.50, obtained from a TPR = 0.76 and a FPR = 0.26. The threshold has a lower intercept but a higher slope, so, after a duration of about 5 hours, it is above that the one given in Eq. 7.

~~For the derivation of thresholds based on ANNs, the following input variable configurations have been investigated: 1) *D*; 2) *HE*; 3) *I*; 4) I_p; 5) *D* and *HE*; 6) *D* and *I*; 7) *D* and I_p; 8) *HE* and I_p; 9) *I* and I_p; 10) *D*, *H E* and I_p. The listed input configurations are indeed all possible ones, except those combining both *H E* and *I* with duration *D*. This has been done because the two pairs *D I* and *D H E* have the same informative content by construction, as confirmed by the fact that the performances of the *D I* and *D H E* neural networks do not differ significantly (see later, Tab. 1). slight differences may occur as ANNs can be sensitive to how a set of variables having together the same information content of another are presented to the network. All the data have been inputted taking their natural logarithms. Different networks have been considered varying the number of hidden neurons from 5 to 20, in order to search for the best value, i.e., the one yielding the highest TSS.~~ Table 1 shows the results obtained from the tested 160 neural network configurations (10 different input layers and 16 different numbers of hidden neurons). In particular, the table shows, for each set of input variables: the optimal number of hidden neurons corresponding to the maximum ~~TSS for the entire data set (third column). The subsequent columns of the table show the~~ TSS for the training, the validation, ~~and~~the test and the entire data set ("all"). For the entire data set also the~~data sets,~~ TPR and the FPR are shown ~~with respect to the reported number of hidden neurons~~. As can be seen, for most of the input configurations, the TSS for the ~~test~~ training and validation data sets is generally quite close. This proves that overfitting has been sufficiently prevented, thanks to the early-stopping criterion – otherwise the performance in the training data set would have been significantly higher than those in the ~~test~~ validation and the test data set.

Table ~~1~~1: Results of tests with ANNs, showing the optimal number of hidden neurons (a number from 5 to 20 has been tested) and the True skill statistics (TSS) for the entire, the training, the validation and the test data sets. Values in the table are compared to TSS_0 = 0.50 which is the maximum value associated to a *D-E* power law threshold.

Input data	Hidden neurons (max TSS)	TSS training	TSS validation	TSS test	TSS all	TPR all	FPR all
D	14	0.35	0.13	0.29	0.30	0.74	0.44
E	18	0.45	0.36	0.41	0.43	0.86	0.43
I	9	0.52	0.41	0.29	0.44	0.74	0.31

I_p	16	0.37	0.35	0.35	0.35	0.81	0.46
D-E	20	0.61	0.59	0.55	0.59	0.82	0.23
D-I	11	0.60	0.60	0.59	0.60	0.80	0.20
D-I_p	7	0.53	0.48	0.47	0.51	0.83	0.32
E-I_p	17	0.46	0.46	0.39	0.44	0.84	0.40
I-I_p	13	0.61	0.54	0.60	0.58	0.82	0.24
D-E-I_p	8	0.68	0.65	0.60	0.66	0.81	0.15

As can be seen from the Table, using only one input variable, the performances are significantly lower than those obtained from the use of the power-law threshold of Eq. 10: however, for the variable with the highest informative content, mean rainfall intensity *I*, the TSS = 0.~~45~~ 44 is quite close to TSS_0 = 0.50. When using input variables in pairs, performances increase significantly. Notably, in the case of the pairs *D-I* and *D-~~H~~ E* – i.e., the same variables used for the power law – the TSS = 0.59 (0.60), which is significantly higher than TSS_0. This is obtained by both an increase of the TPR (~~correct~~true positives ~~prediction~~) and a decrease of the FPR (~~wrong~~false positives ~~predictions~~). The fact that with same input data the neural network provides significantly better performances than the power law, proves that the use of a predetermined parametric form for the threshold equation does not allow to exploit at ~~the~~ fullest the informative content of the input variables, while the flexibility of ANNs allows to achieve a better classification~~.~~. In other words, one of the shortcomings of a power law is that the same equation is usually assumed valid for all the durations, while ANNs ~~is~~are ~~-~~more flexible.

Finally, adding a third variable (network input *D-~~H~~E-*I_p), a further improvement is obtained (TSS = 0.6~~4~~6), mainly due to a decrease of the FPR. This result demonstrates how neural networks can be an aid in searching additional variables that can provide a more reliable dynamic prediction of landslide triggering conditions. In particular, in this case, it has been shown that peak intensity may have an important informative content, an aspect that has not been perhaps sufficiently investigated in the literature~~.~~ even though some researchers have found that the addition of a third variable is a possible way to derive thresholds that better adapt to complex case studies (e.g., Rosi et al., 2021).

~~Conclusions~~Concluding remarks

The identification of rainfall thresholds indicating landslide triggering conditions is a key step for implementing territorial landslide early warning systems. Commonly, thresholds are searched in a limited space, i.e., constrained to a predetermined parametric form, which is generally a power law linking rainfall event, duration *D* and mean intensity *I* (or total depth ~~*H*~~ *E* =*D I*). In this communication we have shown that choosing a predetermined form for the law of the threshold can potentially limit the performance of the empirical model, and how Artificial neural networks are a valuable tool to overcome this limitation. The analysis, referred to the case study of Sicily, has shown that an ~~*H*~~*E*-*D* power-law threshold has a maximum true skill

statistic of TSS = TSS_0 = 0.50. On the other hand, the classifier based on neural networks, using the same pair of input variables, yielded a significantly greater TSS = 0.~~59.~~60 It has also been shown how neural networks allow to easily explore the potential information content of other variables, and hence provide a way to improve predictive performance. For instance, it has been shown that the inclusion of peak rainfall intensity as an additional variable, can lead to an improvement of performance. It is important that when training neural networks, generalization capabilities are ensured, for instance by the early stopping technique. Overfitting is not an issue for the traditional approach based on the power law – or any other parametric equation – as in general the number of free parameters is very low (2 for a power law). This may be a drawback for neural networks, even though it forces one to consider both triggering and non-triggering events, which is fundamental for obtaining thresholds with acceptable statistical characteristics (Peres and Cancelliere, 2021). Another possible disadvantage of neural networks with respect to predetermined-form thresholds is also represented by the fact that it is may ~~generally~~ be difficult ~~not possible~~ to ~~summarize~~ express the neural network classifier as a simple equation. This could ~~hamper~~ limit the practical implementation of triggering thresholds based on neural networks, which could be perceived as impractical by practitioners. However, this limit can ~~potentially~~ for instance be overcome by providing a user-friendly software to the end user.

Data availability. Landslide data from the Franeitalia database (Calvello and Pecoraro, 2018) are available from https://franeitalia.wordpress.com/database/ (last accessed on 29/06/2021), while part of the rainfall data is available from websites of the Servizio Informativo Agreometeorologico Siciliano (SIAS) (http://www.sias.regione.sicilia.it/, last accessed on 05/07/2021) and the Osservatorio delle Acque (http://www.osservatorioacque.it/, last accessed on 05/07/2021).

Author contributions. D.J.P. designed the research, P.D. and D.J.P. conducted the analyses and wrote the paper, A.C. and P.S. and D.J.P. supervised the research and critically revised the manuscript.

Acknowledgements. Pierpaolo Distefano doctoral program's grant is funded by the "Notice 2/2019 for financing the Ph.D. regional grant in Sicily" as part of the Operational Program of European Social Funding 2014–2020 (PO FSE 2014–2020) CUP E65E19000830002. David J. Peres was supported by the post-doctoral grant on "Sviluppo di modelli per la valutazione di strategie innovative di gestione delle risorse idriche in un contesto di cambiamenti climatici" (Development of models for the evaluation of new strategies for water resources management in a changing climate). The research has been partially conducted within the following projects: LIFE 17 CCA/IT/000115 SimetoRES funded by the EASME (now CINEA) of the European Commission, and the Programma Operativo Nazionale Governance e Capacità Istituzionale 2014-2020 - Programma per il supporto al rafforzamento della governance in materia di riduzione del rischio ai fini di protezione civile CUP (Program to support the strengthening of governance in the field of risk reduction for civil protection purposes). APCs funded by "fondi di ateneo 2020-2022, Università di Catania, linea Open Access".

References

Bogaard, T. and Greco, R.: Invited perspectives: Hydrological perspectives on precipitation intensity-duration thresholds for landslide initiation: proposing hydro-meteorological thresholds, Nat. Hazards Earth Syst. Sci., 18(1), 31–39, doi:10.5194/nhess-18-31-2018, 2018.

Caine, N.: The Rainfall Intensity-Duration Control of Shallow Landslides and Debris Flows, Geogr. Ann. Ser. A, Phys. Geogr., 62(1), 23–27, 1980.

Calvello, M. and Pecoraro, G.: FraneItalia: a catalog of recent Italian landslides, Geoenvironmental Disasters, 5(1), doi:10.1186/s40677-018-0105-5, 2018.

Conrad, J. L., Morphew, M. D., Baum, R. L. and Mirus, B. B.: HydroMet: A New Code for Automated Objective Optimization of Hydrometeorological Thresholds for Landslide Initiation, Water 2021, Vol. 13, Page 1752, 13(13), 1752, doi:10.3390/W13131752, 2021.

Froude, M. J. and Petley, D. N.: Global fatal landslide occurrence from 2004 to 2016, Nat. Hazards Earth Syst. Sci., 18(8), 2161–2181, doi:10.5194/nhess-18-2161-2018, 2018.

Gariano, S. L., Brunetti, M. T., Iovine, G., Melillo, M., Peruccacci, S., Terranova, O., Vennari, C. and Guzzetti, F.: Calibration and validation of rainfall thresholds for shallow landslide forecasting in Sicily, southern Italy, Geomorphology, 228, 653–665, doi:10.1016/j.geomorph.2014.10.019, 2015.

Glade, T., Crozier, M. and Smith, P.: Applying probability determination to refine landslide-triggering rainfall thresholds using an empirical "Antecedent Daily Rainfall Model," Pure Appl. Geophys., 157(6–8), 1059–1079, doi:10.1007/s000240050017, 2000.

Guzzetti, F., Peruccacci, S., Rossi, M. and Stark, C. P.: The rainfall intensity-duration control of shallow landslides and debris flows: An update, Landslides, 5(1), 3–17, doi:10.1007/s10346-007-0112-1, 2008.

Haykin, S.: Neural Networks- A Comprehensive Foundation, Second., Prentice Hall., 1999.

Marino, P., Peres, D. J., Cancelliere, A., Greco, R. and Bogaard, T. A.: Soil moisture information can improve shallow landslide forecasting using the hydrometeorological threshold approach, Landslides, 17(9), 2041–2054, doi:10.1007/s10346-020-01420-8, 2020.

Melillo, M., Brunetti, M. T., Peruccacci, S., Gariano, S. L. and Guzzetti, F.: An algorithm for the objective reconstruction of rainfall events responsible for landslides, Landslides, 12(2), 311–320, doi:10.1007/s10346-014-0471-3, 2015.

Melillo, M., Brunetti, M. T., Peruccacci, S., Gariano, S. L. and Guzzetti, F.: Rainfall thresholds for the possible landslide occurrence in Sicily (Southern Italy) based on the automatic reconstruction of rainfall events, Landslides, 13(1), 165–172, doi:10.1007/s10346-015-0630-1, 2016.

Melillo, M., Brunetti, M. T., Peruccacci, S., Gariano, S. L., Roccati, A. and Guzzetti, F.: A tool for the automatic calculation of rainfall thresholds for landslide occurrence, Environ. Model. Softw., 105, 230–243, doi:10.1016/J.ENVSOFT.2018.03.024, 2018.

van Natijne, A. L., Lindenbergh, R. C. and Bogaard, T. A.: Machine learning: New potential for local and regional deep-seated landslide nowcasting, Sensors (Switzerland), 20(5), 1–18, doi:10.3390/s20051425, 2020.

Peres, D. J. and Cancelliere, A.: Derivation and evaluation of landslide-triggering thresholds by a Monte Carlo approach, Hydrol. Earth Syst. Sci., 18(12), 4913–4931, doi:10.5194/hess-18-4913-2014, 2014.

Peres, D. J. and Cancelliere, A.: Comparing methods for determining landslide early warning thresholds: potential use of non-triggering rainfall for locations with scarce landslide data availability, Landslides, 18(9), 3135–3147, doi:10.1007/s10346-021-01704-7, 2021.

Piciullo, L., Calvello, M. and Cepeda, J. M.: Territorial early warning systems for rainfall-induced landslides, Earth-Science Rev., 179, 228–247, doi:10.1016/J.EARSCIREV.2018.02.013, 2018.

Reichenbach, P., Rossi, M., Malamud, B. D., Mihir, M. and Guzzetti, F.: A review of statistically-based landslide susceptibility models, Earth-Science Rev., 180(November 2017), 60–91, doi:10.1016/j.earscirev.2018.03.001, 2018.

Rosi, A., Segoni, S., Canavesi, V., Monni, A., Gallucci, A. and Casagli, N.: Definition of 3D rainfall thresholds to increase operative landslide early warning system performances, Landslides, 18(3), 1045–1057, doi:10.1007/s10346-020-01523-2, 2021.

Sala, G., Lanfranconi, C., Frattini, P., Rusconi, G. and Crosta, G. B.: Cost-sensitive rainfall thresholds for shallow landslides, Landslides, 18(9), 2979–2992, doi:10.1007/s10346-021-01707-4, 2021.