



# On the estimation of landslide intensity, hazard and density via data-driven models

Mariano Di Napoli<sup>1</sup> · Hakan Tanyas<sup>2</sup> · Daniela Castro-Camilo<sup>3</sup> · Domenico Calcaterra<sup>4</sup> · Andrea Cevasco<sup>1</sup> · Diego Di Martire<sup>4</sup> · Giacomo Pepe<sup>1</sup> · Pierluigi Brandolini<sup>1</sup> · Luigi Lombardo<sup>2</sup> 

Received: 26 May 2023 / Accepted: 22 August 2023 / Published online: 12 September 2023  
© The Author(s) 2023

## Abstract

Maps that attempt to predict landslide occurrences have essentially stayed the same since 1972. In fact, most of the geo-scientific efforts have been dedicated to improve the landslide prediction ability with models that have largely increased their complexity but still have addressed the same binary classification task. In other words, even though the tools have certainly changed and improved in 50 years, the geomorphological community addressed and still mostly addresses landslide prediction via data-driven solutions by estimating whether a given slope is potentially stable or unstable. This concept corresponds to the landslide susceptibility, a paradigm that neglects how many landslides may trigger within a given slope, how large these landslides may be and what proportion of the given slope they may disrupt. The landslide intensity concept summarized how threatening a landslide or a population of landslide in a study area may be. Recently, landslide intensity has been spatially modeled as a function of how many landslides may occur per mapping unit, something, which has later been shown to closely correlate to the planimetric extent of landslides per mapping unit. In this work, we take this observation a step further, as we use the relation between landslide count and planimetric extent to generate maps that predict the aggregated size of landslides per slope, and the proportion of the slope they may affect. Our findings suggest that it may be time for the geoscientific community as a whole, to expand the research efforts beyond the use of susceptibility assessment, in favor of more informative analytical schemes. In fact, our results show that landslide susceptibility can be also reliably estimated (AUC of 0.92 and 0.91 for the goodness-of-fit and prediction skill, respectively) as part of a Log-Gaussian Cox Process model, from which the intensity expressed as count per unit (Pearson correlation coefficient of 0.91 and 0.90 for the goodness-of-fit and prediction skill, respectively) can also be derived and then converted into how large a landslide or several coalescing ones may become, once they trigger and propagate downhill. This chain of landslide intensity, hazard and density may lead to substantially improve decision-making processes related to landslide risk.

**Keywords** Landslide susceptibility · Landslide intensity · Landslide hazard · Landslide density · Integrated nested laplace approximation

## 1 Introduction

The international guidelines on managing the landslide threat in mountainous areas (Fell et al. 2008) have long stressed the need to report comprehensive information of landslide hazard. Specifically, in addition to the expectation of landslide occurrences across a given geographic space, the size of the landslides are reported to be equally important (see also, Bellugi et al. 2021). This information is commonly included in physically-based models (van den Bout et al. 2021a), although the requirement of geotechnical data often limits the scope of the analyses at the catchment scale (Van den Bout et al. 2021b).

In a complementary manner, statistical models have the ability to cover large geographic areas, ranging from catchment to regional scales (Catani et al. 2013; Goetz et al. 2015; Di Napoli et al. 2020; Novellino et al. 2021). Such feat can be achieved because statistical models do not strictly require geotechnical parameters to be built. Instead, statistical models traditionally feature proxies of such geotechnical properties, which nowadays can be obtained through remote sensing techniques (van Westen et al. 2008). However, the current literature almost unanimously presents spatial models that evaluate whether a given mapping unit is expected to be stable or unstable (Reichenbach et al. 2018). Therefore, these models inevitably neglect the potential number of landslides within a given mapping unit, as well as the expected planimetric area or volume associated with landslides triggered within the same unit.

An indication of the landslide size is separately computed and it refers to the landslide event magnitude (mL; Malamud et al. 2004), where mL corresponds to a lumped measure which depends on the total number of mapped landslides and their overall planimetric extent. As a result, mL is not spatially distributed but it is rather a single value associated to specific landslide events (Lombardo et al. 2021).

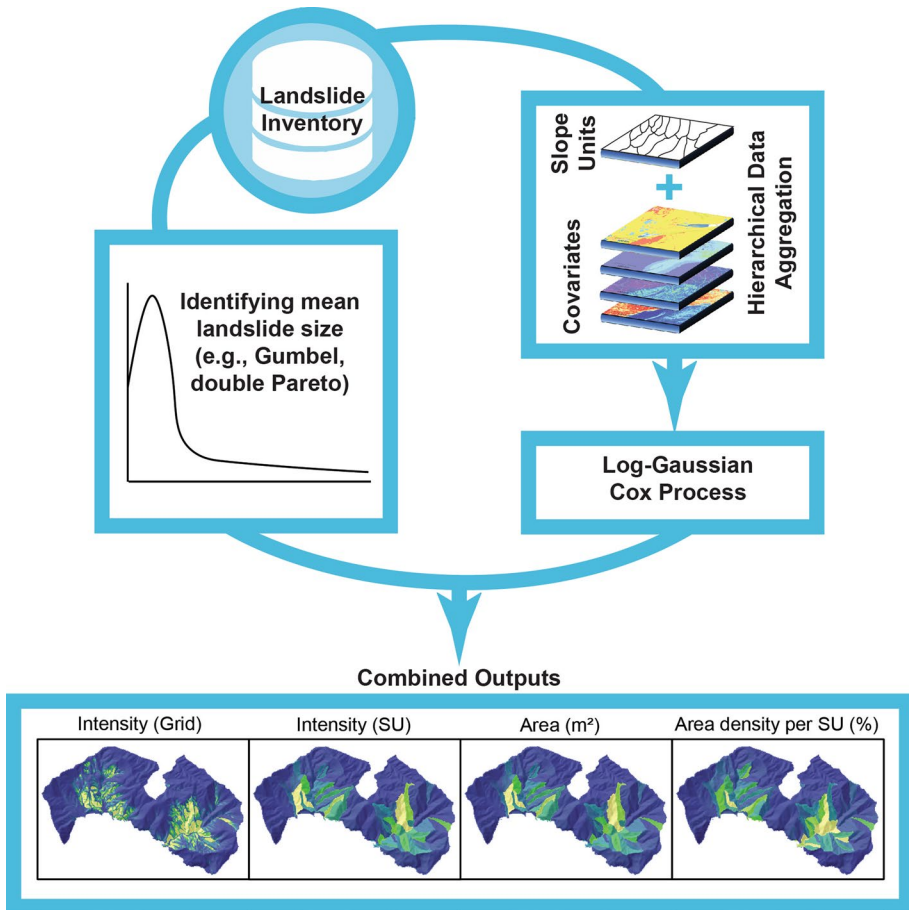
Few attempts have been made to extend the stable/unstable framework in a spatially-explicit context. For instance, Luti et al. (2020) have tried to model the percentage of a mapping unit affected by landslides. Another alternative has been recently proposed by Lombardo et al. (2018), modeling the rate of landslide per unit area or mapping unit. This intensity measure was also shown to closely correlate to the cumulated landslide extent per mapping unit (see Fig. 13 in Lombardo et al. 2020). The latter contribution inspired the work we present here, by further exploiting the relation between intensity and landslide extent. Specifically, we propose a protocol to estimate the intensity first and later convert it to spatially predicted metrics linked to landslide size statistics per mapping unit.

## 2 Materials and methods

This section will briefly introduce the study area and the landslide inventory we used. Subsequently, we will describe the spatial partition and the covariate set, together with the model we selected, referring to the articles where an extensive mathematical formulation is provided. A graphical summary of each analytical step is shown in Fig. 1.

### 2.1 Study area and landslide inventory

The study area where we tested our modeling strategy occupies 18.23 km<sup>2</sup> of the *Cinque Terre* National Park, Italy. Such a name derives from the five epochal villages of



**Fig. 1** Graphical summary of the workflow implemented in this manuscript

*Monterosso al Mare, Vernazza, Corniglia, Manarola and Riomaggiore.* These hamlets are placed along a coastal stretch, which represents a worldwide known tourist attraction. Because of its environmental, cultural, and historical heritage, the *Cinque Terre* territory was listed as a World Heritage Site by UNESCO in 1997 and was declared National Park in 1999. Being most of the *Liguria* region characterized by a rugged morphology and by the absence of large flat areas adapted for cultivation, slope soil covers, over the century, were intensively reworked to build terraces sustained by dry-stone walls (Brandolini 2017). Terraced slopes cover a wide portion of the study area (about 15 km<sup>2</sup>); even if, about 67% of the terraced areas were abandoned (Raso et al. 2021). Agricultural terraces, if correctly preserved, assume a positive role on soil conservation by reducing runoff velocity and soil erosion (Moreno-de-las-Heras et al. 2019). Conversely, when the terraces are abandoned, land degradation issues (i.e., gully erosion, terrace failure, mass movement, piping, hydrological connectivity) arise (Tarolli et al. 2014; Arnáez et al. 2015; Di Napoli et al. 2021). In addition, when intense rainfall occurs, abandoned terraced are particularly prone to shallow mass movements (Brandolini et al. 2018; Cevasco et al. 2015). Geologically, the area is characterized by the presence of five overlapping tectonics units belonging to the Tuscan,

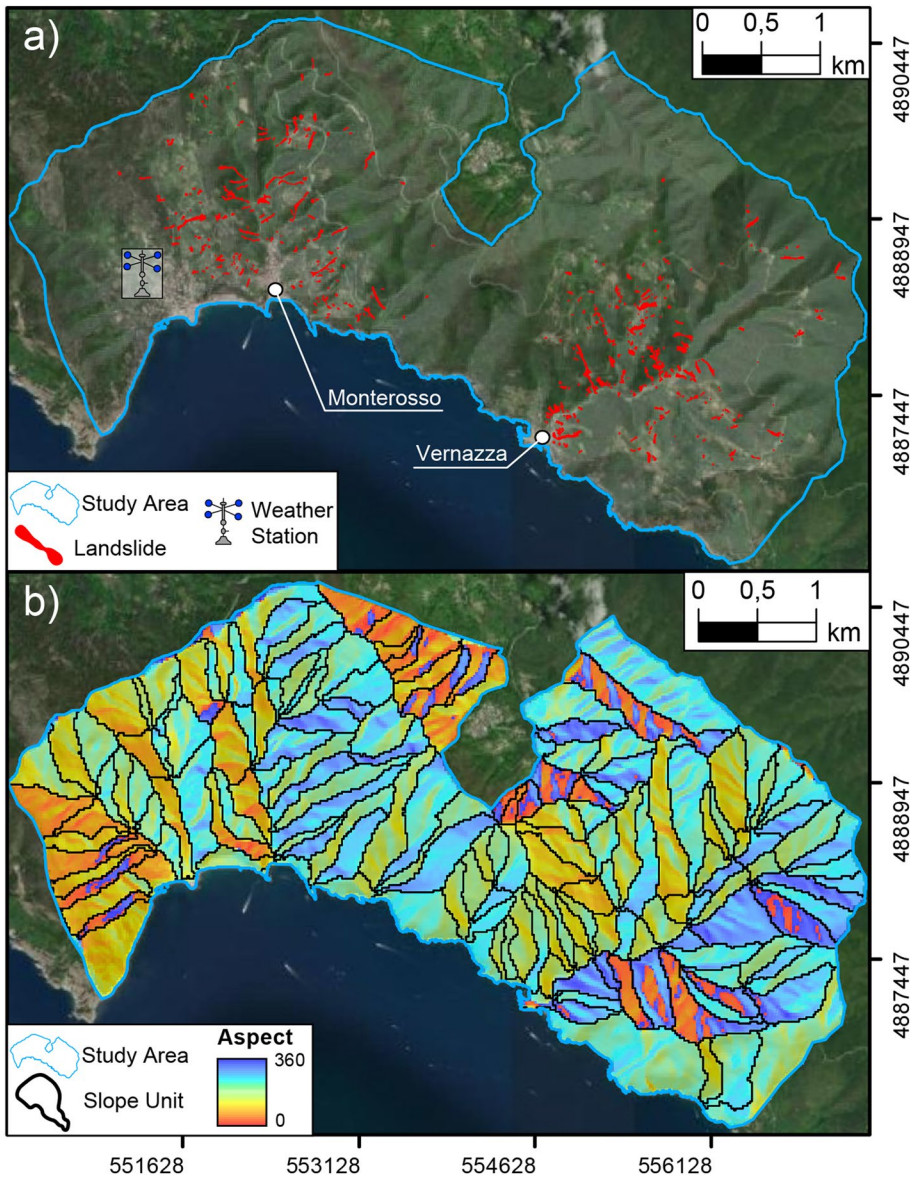
Sub-Ligurian and Ligurian domains (Terranova et al. 2006). From top to bottom, the five overlapping tectonics units include: Gottero Unit, Ottone Unit, Marra Unit, Canetolo Unit and Tuscan Nappe (Abbate et al. 2005). These units belong to an NW–SE oriented segment of the chain, which during the Tertiary orogenic phases, were arranged in a large SW-verging anti-form fold (Gianmarino and Giglia 1990). Most of the territory of the Cinque Terre is occupied by the Tuscan Nappe, which primarily includes thick sandstone-claystone turbidites, diffusely cropping out both along the coast and inland and locally known as Macigno Formation, and secondly by limestones, only visible in the proximity of the easternmost border of the park (Raso et al. 2021). Such complex territory is subjected to high-intensity precipitation capable of mobilizing large amounts of materials having a considerable impact on the solid discharge, flow and energy of streams. These dynamics often cause flow-like movements, debris floods and flash floods, which are increasingly affecting, in particular, the coastal settlements (Zingaro et al. 2019; Brandolini et al. 2012).

These phenomena represent a serious threat to human settlements, inhabitants and trail users, as dramatically the area experienced after the intense rainstorm that hit the *Monterosso* and *Vernazza* areas on 25 October 2011 (Cevasco et al. 2015; Rinaldi et al. 2016), which triggered hundreds of shallow landslides as well as destructive debris floods (Cevasco et al. 2014). Figure 2 shows an overview of the study area and the landslides triggered by the convective storm occurred on October 25 2011. During that day, up to 382 mm of rain were discharged in few hours, as recorded at the weather station of *Monterosso* (Cevasco et al. 2015).

Landslide inventory was redacted from detailed field surveys and analysis of high-resolution aerial images (ground resolution from 3 to 50 cm, according to the altitude) taken a few days after the catastrophic event. The photointerpretation analysis, executed from georeferenced orthophotos provided by *Liguria* Regional Administration, was validated by an intensive field surveys carried out from November 2011 to March 2012 (Cevasco et al. 2013). The mapped mass movements were classified according to Cruden and Varnes (1996) based on the prevalent type of movement and material. The main landslide types recognized can be associated with debris flow, debris avalanche and debris slide, involving colluvial and anthropically reworked deposits overlaying the fractured bedrock. The average landslide density was about 65 landslides/km<sup>2</sup> and landslide areal extent ranged between a few tens up to a few thousands of square meters. Lastly, the event inventory featured 695 mass movements.

## 2.2 Mapping units

To model landslide intensity, we chose a hierarchical structure. The high-resolution mapping unit corresponds to Grid-Cells (GCs, Malamud et al. 2004). These are hierarchically combined with the coarser Slope Units (SUs, Carrara et al. 1995), at which level we computed the Latent Spatial Effect and we aggregated the intensity estimates (see Lombardo et al. 2019a). Specifically, we selected a 20 m resolution GC partition, whereas we computed the SUs by using the *r.slopeunits* software (Alvioli et al. 2016). We parameterized *r.slopeunits* with a circular variance of 0.4, a minimum SU area of 12,500 m<sup>2</sup> and a flow accumulation threshold of 100,000 m<sup>2</sup> (see, Titti et al. 2022). This operation returned 171 SUs. Circular variance is the parameter that *r.slopeunits* uses to control the rigid or flexible aspect criterion for the SU generation. For instance, a value close to zero indicates a very limited aspect variation allowed inside a slope unit. Taking this figuratively to the extreme, the most rigid choice would end up creating one slope unit polygon for each pixel in the



**Fig. 2** **a** Overview of the landslide inventory and boundary of the study area; **b** Slope Unit partition superimposed to the slope aspect to represent clearer the Slope Unit representativeness. The rain gauge symbol corresponds to the location of the Monterosso weather station

aspect raster. On the contrary, a circular variance close to one would make *r.slopeunits* too generous in the SU delineation. Taking this figuratively to the extreme again, the most flexible choice would produce a single SU polygon for the whole study area because it would allow for large aspect variations. The following two parameters are dependent of the concept that catchments and half-catchments are fractal physiographic entities (La Barbera and

Rosso 1989), which means that one can apply the same concept to extract them at different geographic scales (or hydrological order; Blöschl and Sivapalan 1995). The minimum SU area is in fact the parameter that controls the convergence of the polygonal partition to a specific reference planimetric area. As for the accumulation threshold, this parameter controls the starting extent of the half-catchment delineation.

### 2.3 Covariate set

The morphometric covariates we chose to build our intensity model were derived from a 5 m Digital Elevation Model (DEM) accessed from the geo-portal of the Ligurian region. This DEM has been later resampled at 20 m resolution to match the squared lattice we defined. We computed the Euclidean distance from each GC to the nearest road or trail.

We also used the thematic properties described in (Di Napoli et al. 2021). As a result, our covariate set featured: (1) Elevation; (2) Slope Steepness; (3) Eastness; (4) Northness; (5) Planar and (6) Profile Curvatures; (7) Relative Slope Position; (8) Topographic Wetness Index; (9) Distance to road or trail; (10) Land Use; (11) Terraced slope status; (12) Geology.

### 2.4 Landslide intensity modeling

By counting the distribution of slope failures per mapping unit, we can model the resulting data as a Point Process. This is possible under the assumption of spatial continuity, which is something a raster structure naturally offers. As a result, we can define a Poisson Point Process as:

$$N(A) \sim \text{Poisson} \int_A \lambda(s) ds \quad (1)$$

where  $N(A)$  is the number of expected landslides within the study area  $A$ , the selected sector of the *Cinque Terre* National Park in this case,  $\lambda$  is the intensity assumed to be  $\geq 0$ , and  $s$  is each of the GC within the target area. Notably, here  $\lambda$  is assumed to behave according to a Poisson probability distribution. This framework can be further extended conveniently expressing the intensity in logarithmic scale. This procedure then gives rise to what in statistics is referred to as a Log-Gaussian Cox Process (LGCP). A LGCP is particularly convenient because a Log-Gaussian structure allows one to incorporate any type of linear and nonlinear covariate effects. In our case, this leads to denoting our model as follows:

$$\log \{ \lambda(s) \} \sim \text{Gaussian Process} = \beta_0 + \sum_{j=1}^J \beta_j x_j(s) + f_{\text{Geology}} + f_{\text{Land Use}} + f_{\text{Terraces}} + f_{\text{LSE}} \quad (2)$$

where  $\beta_0$  is the global intercept,  $\beta_j$  are the fixed effects used to model continuous covariates and  $f_{\text{Geology}}$ ,  $f_{\text{Land Use}}$  and  $f_{\text{Terraces}}$  are the random effects for categorical properties, whereas  $f_{\text{LSE}}$  is the random effect for the Latent Spatial Effect (LSE). We recall here that the terms fixed and random effects are synonyms of linear and nonlinear models in the context of Bayesian statistics. In other words, a single fixed effect to be exemplified here can be the Elevation, whose use as a linear covariate implies that the landslide intensity is assumed to increase or decrease as a function of the elevation with a fixed rate that cannot change at different levels of altitude. As for the random effects, these are more complex models to integrate how certain covariates contribute to the intensity estimates. Taking the Geology

as an example, this is a discrete covariate whose classes contribute to the LGCP each one independently from the other. Another example of random effect can be instead visualized in the slope steepness. Here, we classified it into 20 classes in a pre-processing step, leading its original continuous information to become discrete as well. However, differently from the Geology case, each class retains some ordinal structure where all Grid-Cells contained in a given bin are always greater than all the Grid-Cells contained in the bin before and smaller than those in the subsequent bin. This is the reason why here we select a different approach for the slope steepness by introducing its effect onto the model with a random walk of the first order (see, Lombardo et al. 2018), for this structure allows for adjacent class dependence to be accounted for. Ultimately, the use of a Latent Spatial Effect is a very different tool from those explained above. In the landslide literature, Grid-Cells, catchments or any other mapping units are often modeled equally in space. In other words, mapping units that are close to each other are treated in the same way as those that are far apart. The only elements that allow the model estimates to change in space are the covariate values. However, in statistics there are solutions to inform the model about the spatial structure in the data. Specifically, adjacency matrix (see Fig. 3 of Opitz et al. 2022) can control this information which is then passed to the model as a latent covariate. The relation above corresponds to a Generalized Additive Mixed Models (GAMM, Steger et al. 2021), which we implement here in its Bayesian form via INLA (Bakka et al. 2018). We recall now two important properties of the landslide intensity. The intensity can always be converted into the most common susceptibility being the latter binary case a simpler realization of the count framework (Lombardo et al. 2019b). This can be achieved as follows:

$$\text{Susceptibility} = 1 - e^{-\lambda_A} \quad (3)$$

In addition, handling the intensity information over space is more convenient than doing the same in the susceptibility case. In fact, the susceptibility is mapping-unit dependent whereas the intensity benefits from the Poisson aggregation property across any spatial units. In this work, we use this property to aggregate  $\lambda$  values estimated for each GC contained in a given SU (see Fig. 5 in Lombardo et al. 2019a).

## 2.5 From landslide intensity to hazard and density

The landslide intensity has been shown to correlate with the total planimetric extent of landslides for each mapping unit (Lombardo et al. 2020). This contribution states that a model able to estimate landslide counts indirectly satisfies the current definition of hazard. Following the definition proposed by Corominas et al. (2014), hazard assessment aims to determine the spatial and temporal probability of slope failures occurrence, together with their mode of propagation, size and intensity. However, Lombardo and co-authors missed an important implication. In fact, if intensity and landslide sizes can be expressed one as the function of the other, this also means that one can convert landslide intensity maps into expected landslide size-related maps.

In this work, this possibility by estimating the intensity per SU and then estimating the landslide extent for each SU by multiplying the intensity for the mean landslide area was explored. We then also take a step further by dividing the estimated landslide areas for the corresponding SU size, thus returning the landslide density.

## 2.6 Performance assessment and model validation

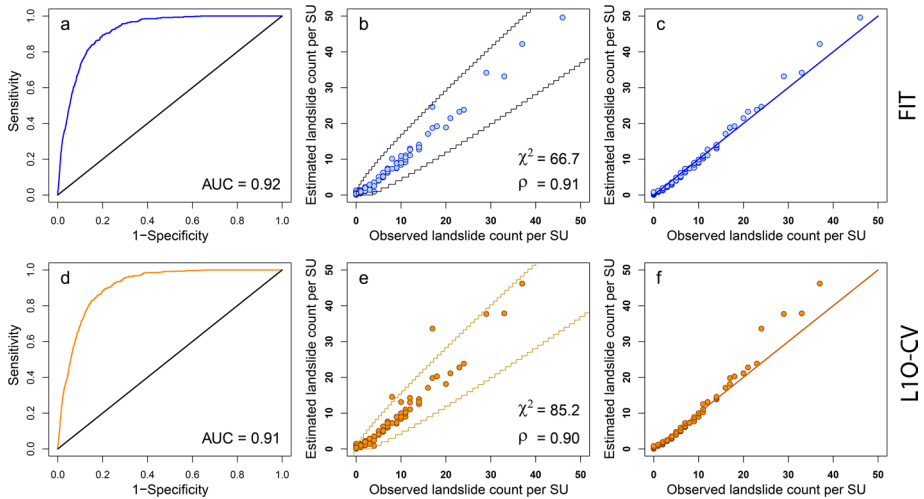
Being our GAMM hierarchical in nature, we separately evaluate the performance at the level of the two mapping units. At the GC scale, where the data is almost binary in nature, Receiver Operating Characteristic (ROC) curves (Hosmer et al. 2013) was employed and their integral or AUC (Rahmati et al. 2019). At the SU level, the agreement between observed landslide counts and aggregated intensities via  $\chi^2$  test and the Spearman correlation coefficient ( $\rho$ ) were checked. In addition, the predictive performance of a fitted model was evaluated using the Leave-One-Out spatial Cross Validation (L1O-CV) method (Tanyaş et al. 2019; Lombardo et al. 2020). The idea of the cross-validation is to perform several splitting of the data into training sample used for fitting the model, and into the validation sample (remaining data) employed for evaluating the predictive accuracy. In L1O procedure, each data object (in this case SU) is left out from the sample and used for validation. This means that different models are fitted (namely, 171), each one calibrates on 170 SU and alternatively predicted on the remaining one. It is important to note that this CV scheme was chosen to perturb the spatial dependence contained in the landslides distribution. In fact, if single GC were removed at random, the spatial scale at which these units act, would have not weakened the spatial structure captured via the latent spatial effect. Conversely, removing all the GC contained in each of the 171 SU partitioning the study area would have caused any residual spatial dependence to be weakened enough to test the model as a predictive tool rather than a simpler exploratory tool. A similar operation in the context of landslide modeling is extensively described in Steger et al. (2016). In our case, we extracted all the GCs contained in each SU. Thus, being the SU different in size, a different number of GCs is extracted for each spatial cross-validation run.

## 3 Results

At the scale of a fine GC, the data are usually near-binary in nature. In other words, no large landslide counts are contained. For this reason, at the GC level, ROC and their AUC were computed; while, at the scale of the SU, the agreement between observed and estimates landslides counts were checked. Furthermore, the L1O-CV was carried out in order to test the model a predictive tool, understanding if the model can correctly estimate unknown landslide counts. The goodness-of-fit and prediction-skill results are, respectively, shown in Fig. 3, where the whole modeling procedure appears to suitably perform, irrespective of the considered mapping unit. Figure 3 summarizes the performance obtained for both the fit and the L1O. Specifically, the AUC for the fit is equal to 0.92 (Fig. 3a), while the AUC obtained from the L1O is 0.91 (Fig. 3d). Considering the classification proposed by Swets (1988) where values lower than 0.5 are considered random, values between 0.5 and 0.7 are accepted as poor, fair in the range 0.7–0.9 and, lastly, excellent for values greater than 0.9; the obtained AUC is diagnostic of outstanding performance.

As for the count framework reported in the second and third columns of Fig. 3, the  $\rho$  values confirm the close match between observed and modeled data both for the fit (Fig. 3b) and the L1O-CV (Fig. 3e). We recall here that  $\rho$  is the Pearson correlation coefficient, and values close to one, are diagnostic of two matching datasets. On a side note, we also opted to report the  $\chi^2$  values. Notably, this parameter is not used





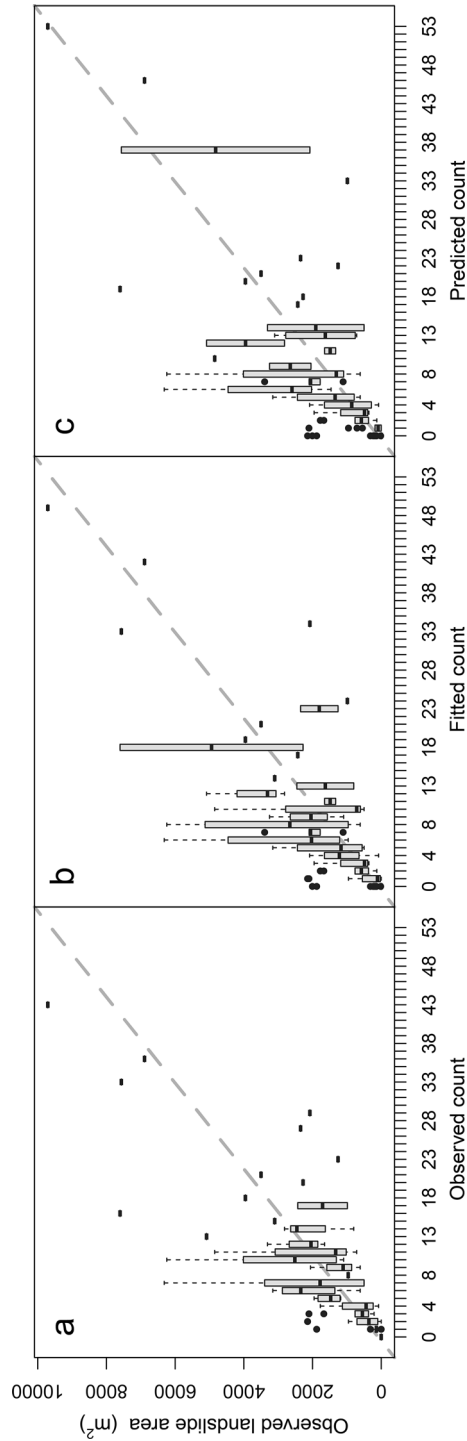
**Fig. 3** Performance assessment overview: the first row shows the goodness-of-fit whereas the second row reports the L10-CV results. In the first column, we show the ROC curves and associated AUCs; the second column summarizes the match between observed and modeled counts, together with their  $\chi^2$  tests; the third column illustrates the QQ-plots again between observed and modeled counts

to indicate performance in absolute value but rather in a comparative manner. Specifically, lower  $\chi^2$  values are diagnostic of a better modeling result and, in this case, they can be used to numerically determine that the results from the fit are slightly better than the ones from the L10-CV. The same considerations can also be made by examining the QQ-plots. This exploratory graphs are built by plotting specific percentiles of two separate distributions under the assumptions that, if the two sets of data have similar characteristics, pairs of quantiles will align along a 45 degree line. In Fig. 3c and f they do match well, with very few cases diverging from the bisector after approximately 25 landslides per SU.

The performance overview shown before is just one component of what one should present in the context of data-driven modeling and particularly anytime the modeling choice falls on statistical tools. In fact, differently from most machine learning approaches, commonly referred to as “black boxes” because of their inherited inability to interpret their prediction, statistical models provide a much more transparent output, including each covariate effect. These are shown in the supplementary material where both linear and nonlinear effects are graphically presented and interpreted.

As introduced in Sect. 1, Fig. 13 in Lombardo et al. (2020) showed that the intensity is closely related to landslide areas per SU. Therefore, we recreated the same plot to test whether this observation holds even in our study area. This is shown in Fig. 4, where the above-mentioned relation appears to be valid also for the shallow landslides mapped within the studied sector of the *Cinque Terre* National Park. Also, this relation does not get lost in the fitting and predicting phases.

We used the estimated landslide intensities to determine the expected landslide area aggregated per SU. This can be achieved by taking the product of the intensity times the mean landslide area per SU. However, the empirical mean of the landslide area distribution may be site specific, therefore, we tested whether we could generalize this information by

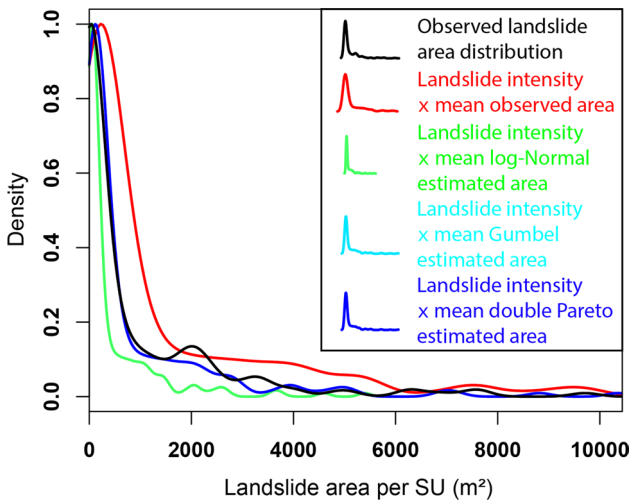


**Fig. 4** Association between landslide counts and areas aggregated at the SU level, for the observation (a), the fit (b) and the LJO-CV (c)

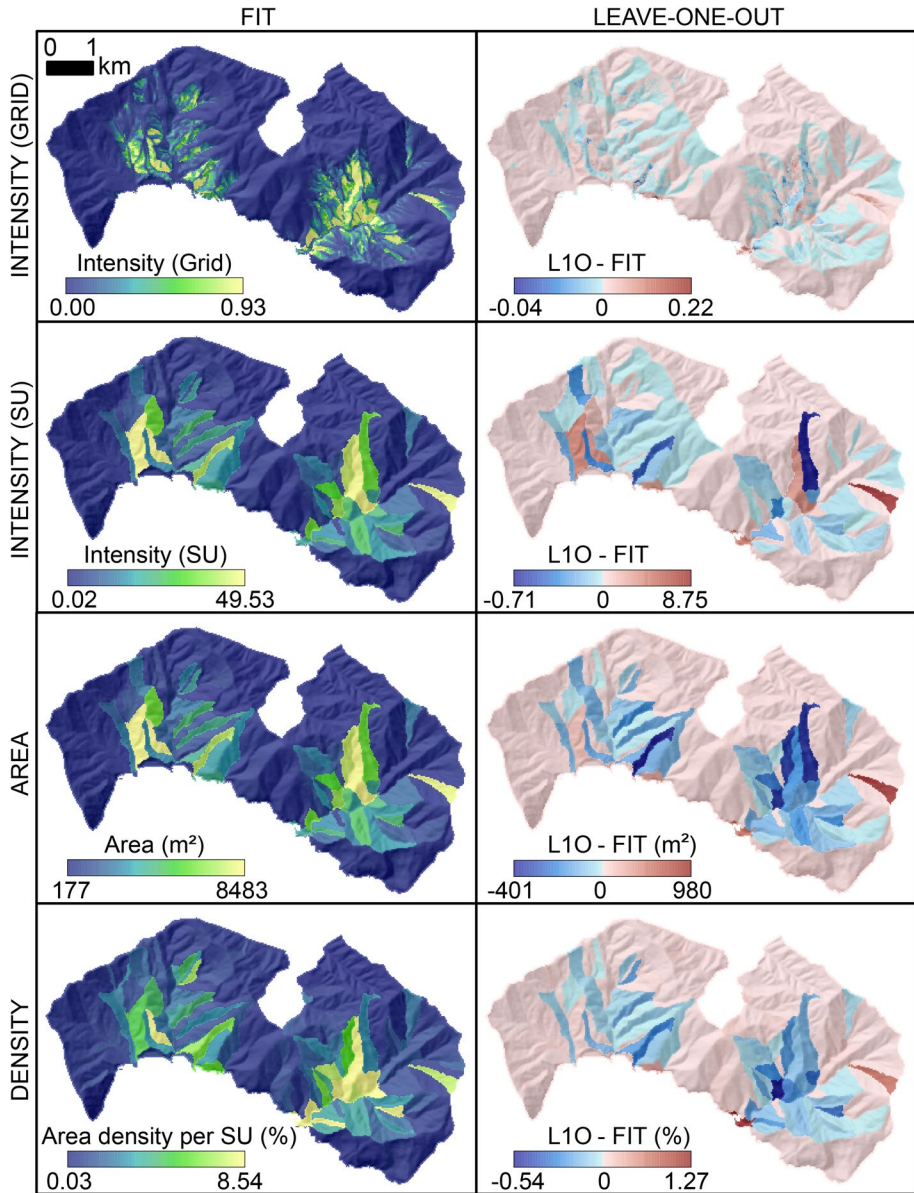
estimating the theoretical mean. This operation follows the assumptions stated in Malamud et al. (2004), although here we extend the same idea to the spatial context.

Specifically, we fitted a series of statistical distributions (Gumbel, inverse-Gamma, double Pareto, log-Gaussian) to get an estimate of the population mean accounting for the heavy tail of the landslide area distribution. We found that both the Gumbel and double Pareto distribution provide a consistent estimate of the mean. We use this estimate to construct a plug-in estimator of the density of the landslide sizes distribution by multiplying it with the estimated intensity. The resulting landslide area distributions are shown in Fig. 5, where the conversion from the Gumbel (or double Pareto) appears to closely match the actual observational data. This result is the foundation of the first mapping procedure in the geo-scientific literature where landslide areas as well as landslide densities are estimated in map form through data-driven model.

Figure 6 graphically summarizes the aforementioned maps. The left column, making use of the fitted intensities, shows a pattern that closely matches the landslide spatial distribution shown in Fig. 2. But, much more information is provided, with the expected number of landslides both at the GC and SU levels, together with the converted landslide area and density. The second column reports the deviation from the fit of the equivalent information. This is computed as the difference of the fitted results being subtracted from the LIO-predicted ones. The figure graphically stresses something already mentioned above, this being the stability of our landslide intensity framework. In fact, very narrow residuals are generally returned across the whole study area. Moreover, the largest ones correspond to single slope units, where likely much localized landscape characteristics affect the distribution of the original landslide counts.



**Fig. 5** Landslide area distributions generated by multiplying the LIO-CV intensity to the empirical mean and the populations means obtained by through a Gumbel, double Pareto (DP), log-Gaussian fits. The means estimated via the Gumbel and DP fits are equivalent and perfectly overlapping. Notably, we tried to fit an inverse-Gamma as well but opted to not report it because the estimated mean tends to infinity



**Fig. 6** Landslide intensity, area and density maps. The landslide area is obtained by multiplying the Gumbel landslide area population mean to the landslide intensity values. As for the landslide density this is equivalent to the landslide area divided by the slope unit extent. The first row is expressed at the grid level, whereas the second, third and fourth one can only be shown at the Slope Unit level, something possible, thanks to the hierarchical nature of the LGCP. The first column shows the results of the fitting procedure whereas the second column shows the residuals between the fitted and cross-validated results

## 4 Discussion

The workflow we propose has some unique features meant to address the landslide hazard definition (Guzzetti et al. 1999). The binary classification context typical of landslide susceptibility studies is left behind in favor of testing a count-oriented model, which is further exploited to derive the expected landslide area and density per SU. The strength of this procedure resides in the advantages it brings with respect to the landslide susceptibility counterpart. In fact, whenever we apply a dichotomous classification to a given study site, we neglect the number of landslides that certain regions may exhibit. Therefore, we may heavily underestimate the threat that any urban settlement may be exposed to. If a SU contains tens of debris flows and another SU contains just one, a binary classifier will treat the two mapping units in the very same way. Conversely, the landslide intensity framework proposed by Lombardo et al. (2018) respect the spatial information carried by the number of landslides per mapping unit. However, even the intensity framework has some weaknesses. For instance, the number of events may be difficult to interpret in terms of hazard because of amalgamation issues (Tanyaş et al. 2019). Conversely, as also clearly stated in the most accepted definition of landslide hazard (Guzzetti et al. 1999) and in the international guidelines (Fell et al. 2008), a much more informative parameter is the landslide area (Turcotte et al. 2002). It is also worth mentioning that an even better intensity parameter is the landslide velocity or kinematic energy (Corominas et al. 2014). However, this parameter can only be obtained via physically-based models and currently no large database exist to support its estimated via data-driven models. Therefore, the landslide area is the most viable solution to estimate landslide hazard in the context of spatially-explicit models (alternatively one could use volumes obtained through empirical conversions, with all the uncertainties they would introduce though (Larsen et al. 2010). Few examples exist on this topic, the first one corresponding to Lombardo et al. (2021). There, the authors modeled the aggregated landslide area per SU via a log-Gaussian GAM. However, even this case has its own limitations. The use of a log-Gaussian likelihood implies that the landslide area is expressed at the logarithmic scale, thus making the interpretation difficult. Moreover, a Log-Gaussian model works well for the bulk of a distribution but not for the tails. Thus, when transforming back from the logarithmic to the actual metric scale, the very small and very large landslides exhibit the largest errors. Moreover, the large ones are also the most threatening ones, thus an error in the tail would result in a large underestimation of the landslide hazard. In addition to this issue, the model introduced by Lombardo et al. (2021) uniquely targets the landslide area without accounting for the proneness to fail of a given slope. In other words, if the model estimates that a slope has the right characteristics to potentially release a large landslide, but the susceptibility is very low, then the hazard would also be very low. Therefore, our contribution fits in the context previously described by combining all the required information. The intensity intrinsically returns an estimate of which slopes are unstable, and through the actual number of expected landslides, we derive the expected landslide size and density per mapping unit.

Our model satisfies most of the requirements of the landslide hazard definition. However, because we used an event-based landslide inventory (Guzzetti et al. 2012), our model lacks the temporal characteristic typical of the hazard context. To extend our model from the purely spatial to the spatio-temporal framework, the inventory must reflect multi-temporal occurrences. As a result, we could implement a space–time LGCP model whose intensities can be converted into expected landslide areas and density per mapping unit according to the user preferences. Also, our model relies on the assumption that as the

number of landslides increases, the landslide area per mapping unit should also proportionally increase. This assumption may be valid but it may also be very site-dependent. In fact, certain slopes may give rise to single and large landslides whose planimetric area maybe much larger than many small landslides combined. In such situations, our assumption may not hold and therefore our model may not be applicable. In many sites, the underestimation brought by very large landslides may affect few if not single slopes. Thus, our approach could still be extremely valuable to assess the hazard across the whole study area. However, in structurally controlled landscapes where landslides tend to be generally large (e.g., Tanyaş et al. 2022), our approach may not be applicable. Also, this assumption has been mainly tested so far for translational landslides and debris flows. More tests are required to validate this assumption in various geographic contexts and different type of failure mechanisms.

Ultimately, the real advantage of the approach we propose has to do with available landslide inventories. The current tendency is for scientists to map landslides as polygonal features. This is clearly the most appropriate approach to mapping. However, the community has not standardized this procedure and a large number of point-based inventories are continuously released, even through semi-automated mapping protocols (Bhuyan et al. 2023). But, even if starting from tomorrow, all landslides would be perfectly mapped and shared via polygonal inventories, this does not change the fact that five decades of geomorphological mapping has produces enormous point-based information. To estimate landslide intensity, one only needs number of landslides per mapping unit, an information easily estimated even with point data. This would by-pass the strict need for planimetric information and allow one to estimate the expected landslide area per mapping unit by converting the intensity. The only requirement would be to have access to a mean landslide area, likely connected to the landslide type and general tectonic and climatic setting, something largely demonstrated in a number of papers (Amato et al. 2021; Malamud et al. 2004; Taylor et al. 2018). As a result, one could mine a large amount of unused information and potentially convert five decades of traditional susceptibility maps into hazard ones.

## 5 Concluding remarks

The modeling protocol we propose represent an attempt to combine the statistics typical of spatial point patterns and of extreme value theories applied to the landslide context. The results we produced are an interesting example of the information one can obtain by considering other modeling frameworks, different from the ones employed in traditional susceptibility assessment. Nevertheless, even this framework can be largely improved. Future development should in fact involve joint probability models. In this work, we used an LGCP and a few light and heavy tailed distributions in a separate manner. However, more advanced statistical solutions allow these two elements to be connected. This modeling framework is commonly referred to as Marked Point Process, something that should be definitely explored in the future. For now though, this work serves as a starting point for a unified landslide hazard assessment through data-driven tools.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11069-023-06153-0>.

**Acknowledgements** The research presented in this article is partially supported by King Abdullah University of Science and Technology (KAUST) in Thuwal, Saudi Arabia, Grant URF/1/4338-01-01.

**Funding** The authors have not disclosed any funding.

## Declarations

**Conflict of interest** The authors declare no competing interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Abbate E, Fanucci F, Benvenuti M, Bruni P, Cipriani N, Falorni P, Fazzuoli M, Morelli D, Pandeli E, Papini M et al. (2005) Carta Geologica D'Italia Alla Scala 1:50.000-Foglio 248-La Spezia Con Note Illustrative; APAT—Agenzia per la Protezione dell'Ambiente e per i Servizi Tecnici. Rome, Italy
- Alvioli M, Marchesini I, Reichenbach P, Rossi M, Ardizzone F, Fiorucci F, Guzzetti F (2016) Automatic delineation of geomorphological slope units with `r.slopeunits v1.0` and their optimization for landslide susceptibility modeling. *Geosci Model Dev* 9:3975–3991. <https://doi.org/10.5194/gmd-9-3975-2016>
- Amato G, Palombi L, Raimondi V (2021) Data-driven classification of landslide types at a national scale by using Artificial Neural Networks. *Int J Appl Earth Obs Geoinf* 104:102549. <https://doi.org/10.1016/j.jag.2021.102549>
- Arnáez J, Lana-Renault N, Lasanta T, Ruiz-Flaño P, Castroviejo J (2015) Effects of farming terraces on hydrological and geomorphological processes. A review. *Catena* 128:122–134
- Bakka H, Rue H, Fuglstad G-A, Riebler A, Bolin D, Illian J, Krainski E, Simpson D, Lindgren F (2018) Spatial modeling with R-INLA: a review. *WIREs Comput Stat* 10:e1443. <https://doi.org/10.1002/wics.1443>
- Bellugi DG, Milledge DG, Cuffey KM, Dietrich WE, Larsen LG (2021) Controls on the size distributions of shallow landslides. *PNAS*. <https://doi.org/10.1073/pnas.2021855118>
- Bhuyan K, Meena SR, Nava L, van Westen C, Floris M, Catani F (2023) Mapping landslides through a temporal lens: an insight toward multi-temporal landslide mapping using the u-net deep learning model. *Gisci Remote Sens* 60(1):2182057
- Blöschl G, Sivapalan M (1995) Scale issues in hydrological modelling: a review. *Hydrol Process* 9(3–4):251–290
- Brandolini P (2017) The outstanding terraced landscape of the Cinque Terre coastal slopes (eastern Liguria). In: *Landscapes and landforms of Italy*, pp 235–244
- Brandolini P, Cevasco A, Firpo M, Robbiano A, Sacchini A (2012) Geo-hydrological risk management for civil protection purposes in the urban area of Genoa (Liguria, NW Italy). *Nat Hazard* 12(4):943–959
- Brandolini P, Cevasco A, Capolongo D, Pepe G, Lovergine F, Del Monte M (2018) Response of terraced slopes to a very intense rainfall event and relationships with land abandonment: a case study from Cinque Terre (Italy). *Land Degrad Dev* 29(3):630–642
- Carrara A, Cardinali M, Guzzetti F, Reichenbach P (1995) GIS technology in mapping landslide hazard. In: Carrara A, Guzzetti F (eds) *Geographical information systems in assessing natural hazards, advances in natural and technological hazards research*. Springer, Dordrecht, pp 135–175. [https://doi.org/10.1007/978-94-015-8404-3\\_8](https://doi.org/10.1007/978-94-015-8404-3_8)

- Catani F, Lagomarsino D, Segoni S, Tofani V (2013) Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues. *Nat Hazard* 13:2815–2831. <https://doi.org/10.5194/nhess-13-2815-2013>
- Cevasco A, Brandolini P, Scopesi C, Rellini I (2013) Relationships between geo-hydrological processes induced by heavy rainfall and land-use: the case of 25 October 2011 in the Vernazza catchment (Cinque Terre, NW Italy). *J Maps* 9(2):289–298
- Cevasco A, Pepe G, Brandolini P (2014) The influences of geological and land use settings on shallow landslides triggered by an intense rainfall event in a coastal terraced environment. *Bull Eng Geol Environ* 73:859–875
- Cevasco A, Diodato N, Revellino P, Fiorillo F, Grelle G, Guadagno FM (2015) Storminess and geo-hydrological events affecting small coastal basins in a terraced Mediterranean environment. *Sci Total Environ* 532:208–219. <https://doi.org/10.1016/j.scitotenv.2015.06.017>
- Corominas J, van Westen C, Frattini P, Cascini L, Malet J-P, Fotopoulou S, Catani F, Van Den Eckhaut M, Mavrouli O, Agliardi F, Pitilakis K, Winter MG, Pastor M, Ferlisi S, Tofani V, Hervás J, Smith JT (2014) Recommendations for the quantitative analysis of landslide risk. *Bull Eng Geol Environ* 73:209–263. <https://doi.org/10.1007/s10064-013-0538-8>
- Cruden DM, Varnes DJ (1996) Landslides: investigation and mitigation. Chapter 3—landslides types and processes. Transportation research board special report, 247
- Di Napoli M, Carotenuto F, Cevasco A, Confuorto P, Di Martire D, Firpo M, Pepe G, Raso E, Calcaterra D (2020) Machine learning ensemble modelling as a tool to improve landslide susceptibility mapping reliability. *Landslides*. <https://doi.org/10.1007/s10346-020-01392-9>
- Di Napoli M, Di Martire D, Bausilio G, Calcaterra D, Confuorto P, Firpo M, Pepe G, Cevasco A (2021) Rainfall-induced shallow landslide detachment, transit and runoff susceptibility mapping by integrating machine learning techniques and GIS-based approaches. *Water* 13:488. <https://doi.org/10.3390/w13040488>
- Fell R, Corominas J, Bonnard C, Cascini L, Leroi E, Savage WZ (2008) Guidelines for landslide susceptibility, hazard and risk zoning for land-use planning. *Eng Geol* 102:99–111. <https://doi.org/10.1016/j.enggeo.2008.03.014>
- Gianmarino S, Giglia G (1990) Gli elementi strutturali della piega di La Spezia nel contesto geodinamico dell'Appennino settentrionale. *Boll della Soc Geol Ital* 109(4):683–692
- Goetz JN, Brenning A, Petschko H, Leopold P (2015) Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. *Comput Geosci* 81:1–11. <https://doi.org/10.1016/j.cageo.2015.04.007>
- Guzzetti F, Carrara A, Cardinali M, Reichenbach P (1999) Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. *Geomorphology* 31:181–216. [https://doi.org/10.1016/S0169-555X\(99\)00078-1](https://doi.org/10.1016/S0169-555X(99)00078-1)
- Guzzetti F, Mondini AC, Cardinali M, Fiorucci F, Santangelo M, Chang K-T (2012) Landslide inventory maps: new tools for an old problem. *Earth Sci Rev* 112:42–66. <https://doi.org/10.1016/j.earscirev.2012.02.001>
- Hosmer DWJ, Lemeshow S, Sturdivant RX (2013) Applied Logistic Regression. John Wiley & Sons
- La Barbera P, Rosso R (1989) On the fractal dimension of stream networks. *Water Resour Res* 25(4):735–741
- Larsen IJ, Montgomery DR, Korup O (2010) Landslide erosion controlled by hillslope material. *Nature Geosci* 3:247–251. <https://doi.org/10.1038/ngeo776>
- Lombardo L, Opitz T, Huser R (2018) Point process-based modeling of multiple debris flow landslides using INLA: an application to the 2009 Messina disaster. *Stoch Environ Res Risk Assess* 32:2179–2198. <https://doi.org/10.1007/s00477-018-1518-0>
- Lombardo L, Bakka H, Tanyas H, van Westen C, Mai PM, Huser R (2019a) Geostatistical modeling to capture seismic-shaking patterns from earthquake-induced landslides. *J Geophys Res Earth Surf* 124:1958–1980. <https://doi.org/10.1029/2019JF005056>
- Lombardo L, Opitz T, Huser R (2019) Numerical recipes for landslide spatial prediction using R-INLA: a step-by-step tutorial. In: Pourghasemi HR, Gokceoglu C (eds) *Spatial modeling in GIS and R for earth and environmental sciences*. Elsevier, Amsterdam, pp 55–83. <https://doi.org/10.1016/B978-0-12-815226-3.00003-X>
- Lombardo L, Opitz T, Ardizzone F, Guzzetti F, Huser R (2020) Space-time landslide predictive modelling. *Earth Sci Rev* 209:103318. <https://doi.org/10.1016/j.earscirev.2020.103318>
- Lombardo L, Tanyas H, Huser R, Guzzetti F, Castro-Camilo D (2021) Landslide size matters: a new data-driven, spatial prototype. *Eng Geol* 293:106288. <https://doi.org/10.1016/j.enggeo.2021.106288>
- Luti T, Segoni S, Catani F, Munafò M, Casagli N (2020) Integration of remotely sensed soil sealing data in landslide susceptibility mapping. *Remote Sens* 12(9):1486
- Malamud BD, Turcotte DL, Guzzetti F, Reichenbach P (2004) Landslide inventories and their statistical properties. *Earth Surf Proc Land* 29:687–711. <https://doi.org/10.1002/esp.1064>



- Moreno-de-las-Heras M, Lindenberger F, Latron J, Lana-Renault N, Llorens P, Arnáez J et al (2019) Hydrogeomorphological consequences of the abandonment of agricultural terraces in the Mediterranean region: key controlling factors and landscape stability patterns. *Geomorphology* 333:73–91
- Novellino A, Cesarano M, Cappelletti P, Di Martire D, Di Napoli M, Ramondini M, Sotter A, Calcaterra D (2021) Slow-moving landslide risk assessment combining machine learning and InSAR techniques. *CATENA* 203:105317. <https://doi.org/10.1016/j.catena.2021.105317>
- Opitz T, Bakka H, Huser R, Lombardo L (2022) High-resolution Bayesian mapping of landslide hazard with unobserved trigger event. *Ann Appl Stat* 16(3):1653–1675
- Rahmati O, Kornejady A, Samadi M, Deo RC, Conoscenti C, Lombardo L, Dayal K, Taghizadeh-Mehrjardi R, Pourghasemi HR, Kumar S, Bui DT (2019) PMT: new analytical framework for automated evaluation of geo-environmental modelling approaches. *Sci Total Environ* 664:296–311. <https://doi.org/10.1016/j.scitotenv.2019.02.017>
- Raso E, Mandarino A, Pepe G, Calcaterra D, Cevasco A, Confuorto P et al (2021) Geomorphology of Cinque Terre National Park (Italy). *J Maps* 17(3):171–184
- Reichenbach P, Rossi M, Malamud BD, Mihir M, Guzzetti F (2018) A review of statistically-based landslide susceptibility models. *Earth Sci Rev* 180:60–91. <https://doi.org/10.1016/j.earscirev.2018.03.001>
- Rinaldi M, Amponsah W, Benvenuti M, Borga M, Comiti F, Lucia A et al (2016) An integrated approach for investigating geomorphic response to extreme events: methodological framework and application to the October 2011 flood in the Magra River catchment, Italy. *Earth Surf Process Landf* 41(6):835–846
- Steger S, Brenning A, Bell R, Glade T (2016) The propagation of inventory-based positional errors into statistical landslide susceptibility models. *Nat Hazard* 16:2729–2745. <https://doi.org/10.5194/nhess-16-2729-2016>
- Steger S, Mair V, Kofler C, Pittore M, Zebisch M, Schneiderbauer S (2021) Correlation does not imply geomorphic causation in data-driven landslide susceptibility modelling: benefits of exploring landslide data collection effects. *Sci Total Environ* 776:145935. <https://doi.org/10.1016/j.scitotenv.2021.145935>
- Swets JA (1988) Measuring the accuracy of diagnostic systems. *Science* 240(4857):1285–1293
- Tanyaş H, van Westen CJ, Allstadt KE, Jibson RW (2019) Factors controlling landslide frequency–area distributions. *Earth Surf Proc Land* 44:900–917. <https://doi.org/10.1002/esp.4543>
- Tanyaş H, Hill K, Mahoney L, Fadel I, Lombardo L (2022) The world’s second-largest, recorded landslide event: lessons learnt from the landslides triggered during and after the 2018 Mw 7.5 Papua New Guinea earthquake. *Eng Geol* 297:106504. <https://doi.org/10.1016/j.enggeo.2021.106504>
- Tarolli P, Preti F, Romano N (2014) Terraced landscapes: from an old best practice to a potential hazard for soil degradation due to land abandonment. *Anthropocene* 6:10–25
- Taylor FE, Malamud BD, Witt A, Guzzetti F (2018) Landslide shape, ellipticity and length-to-width ratios. *Earth Surf Proc Land* 43:3164–3189. <https://doi.org/10.1002/esp.4479>
- Terranova REMO, Zanzucchi G, Bernini MASSIMO, Brandolini PIERLUIGI, Campobasso SILVIA, Faccini F, Zanzucchi F (2006) Geologia, geomorfologia e vini del parco Nazionale delle Cinque Terre (Liguria, Italia). *Boll della Soc Geol Ital Spec* 6:115–128
- Titti G, Napoli GN, Conoscenti C, Lombardo L (2022) Cloud-based interactive susceptibility modeling of gully erosion in Google Earth Engine. *Int J Appl Earth Obs Geoinf* 115:103089
- Turcotte DL, Malamud BD, Guzzetti F, Reichenbach P (2002) Self-organization, the cascade model, and natural hazards. *Proc Natl Acad Sci* 99:2530–2537. <https://doi.org/10.1073/pnas.012582199>
- Van den Bout B, Lombardo L, Chiyang M, van Westen C, Jetten V (2021) Physically-based catchment-scale prediction of slope failure volume and geometry. *Eng Geol* 284:105942. <https://doi.org/10.1016/j.enggeo.2020.105942>
- van den Bout B, van Asch T, Hu W, Tang CX, Mavrouli O, Jetten VG, van Westen CJ (2021b) Towards a model for structured mass movements: the OpenLISEM hazard model 2.0a. *Geosci Model Dev* 14:1841–1864. <https://doi.org/10.5194/gmd-14-1841-2021>
- van Westen CJ, Castellanos E, Kuriakose SL (2008) Spatial data for landslide susceptibility, hazard, and vulnerability assessment: an overview. *Eng Geol Landslide Susceptibility Hazard Risk Zoning Land Use Plan* 102:112–131. <https://doi.org/10.1016/j.enggeo.2008.03.010>
- Zingaro M, Refice A, Giachetta E, D’Addabbo A, Lovergine F, De Pasquale V et al (2019) Sediment mobility and connectivity in a catchment: a new mapping approach. *Sci Total Environ* 672:763–775

## Authors and Affiliations

**Mariano Di Napoli**<sup>1</sup> · **Hakan Tanyas**<sup>2</sup> · **Daniela Castro-Camilo**<sup>3</sup> · **Domenico Calcaterra**<sup>4</sup> · **Andrea Cevasco**<sup>1</sup> · **Diego Di Martire**<sup>4</sup> · **Giacomo Pepe**<sup>1</sup> · **Pierluigi Brandolini**<sup>1</sup> · **Luigi Lombardo**<sup>2</sup> 

✉ Luigi Lombardo  
l.lombardo@utwente.nl

<sup>1</sup> Department of Earth, Environment and Life Sciences, University of Genoa, Corso Europa, 26, 16132 Genoa, Italy

<sup>2</sup> Department of Earth System Analysis, Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, PO Box 217, AE 7500 Enschede, Netherlands

<sup>3</sup> School of Mathematics and Statistics, University of Glasgow, Glasgow G12 8QQ, UK

<sup>4</sup> Department of Earth, Environment and Resources Sciences, Federico II University of Naples, 80138 Naples, Italy