



# From theory to practice: optimisation of available information for landslide hazard assessment in Rome relying on official, fragmented data sources

**Abstract** The definition of landslide hazard is a step-like procedure that encompasses the quantification of its spatial and temporal attributes, i.e., a reliable definition of landslide susceptibility and a detailed analysis of landslide recurrence. However, available information is often incomplete, fragmented and unsuitable for reliable quantitative analysis. Nevertheless, landslide hazard evaluation has a key role in the implementation of risk mitigation policies and an effort should be done to retrieve information and make it useful for this purpose. In this research, we go through this topic of optimising the information available in catalogues, starting from landslide inventory review and constitution of a boosted training dataset, propaedeutic for susceptibility analysis based on machine learning methods. The temporal recurrence of landslide events has been approached here either through the definitions of large-scale quantitative hazard descriptors or by analysis of historical rainfall (i.e., the main triggering factor for the considered shallow earth slope failures) databases through the definition of rainfall probability curves. Spatial and temporal attributes were integrated, selecting potential landslide source areas ranked in terms of hazard. Data integration was also pursued through persistent scatterer interferometry analysis which pointed out areas of interest within potential landslide source areas featured by ongoing ground movement. The consequential approach led to the definition of the first hazard product of the city of Rome at a local scale functional for advisory purposes or the statutory level, representing a thematic layer able to orient the risk managers and infrastructure stakeholders.

**Keywords** Susceptibility · Machine learning · Rainfall probability · Landslide hazard · Landslide inventories · Interferometry

## Introduction

A shift from the ideal scientific framework to its actual applicability is required when dealing with risk-oriented hazard analyses related to land use planning purposes (Thiery et al. 2020). Such a task can be particularly challenging, as the application of rigorous and reliable scientific approaches can be data demanding, in terms of quality and quantity of information needed to fully exploit their potential (Corominas et al. 2014; Glade 2001). On the contrary, in some cases, the quality and quantity of available data do not match the optimal (and sometimes even fair) requirements to ensure reliable results for regulatory purposes (Cascini 2008). Furthermore, the possibility of acquiring additional or higher resolution data can be limited at times, especially for large-scale prevention and

mitigation activities (i.e., land use planning and the propaedeutic hazard zoning), being related to available financial resources as well as time constraints. Nevertheless, local or national administrations in charge of risk prevention and mitigation could optimise the available information to fit as much as possible scientific standards. For this purpose, data-driven and knowledge-based methods can be implemented to integrate and boost available data sources.

In this perspective, we detailly report and discuss an integrated approach devoted to a reliable assessment of landslide hazards in one of the more challenging contexts exposed to risk: the urban area of Rome. The scenario of gravitational instabilities and related damage resulting from an intense rainfall event occurred in 2014 (Alessi et al. 2014) highlighted how the landslide risk is not negligible in the city of Rome. This statement is justified by the occurrence of recurring, usually small-size landslide events (both as single-slope failures and multiple phenomena over large areas), and the high exposure as regards the number and value of the exposed elements. The landslide conditions are known (Amanti et al. 2013; Del Monte et al. 2016) and taken into due consideration in the official planning tools, such as the Hydro-geological Structure Plan (<https://www.autoritadistrettoac.it/planning/hydrographic-basin-planning/documentation-of-the-tiber-basin-plan>) and the municipal land use plan (<http://www.urbanistica.comune.roma.it/prg-2008-vigente/>). Collaborations between research and government institutions (Amanti et al. 2013; Esposito et al. 2019) show a growing awareness and interest on this issue. However, a proper hazard analysis and the relative zoning of the territory has not been carried out so far. The knowledge of the hazard, from the temporal point of view, is instead a fundamental requirement for a correct and exhaustive definition of the risk and subsequent actions for risk mitigation (Corominas et al. 2014). Although this task is pursued by the municipal administration, at least as Civil Protection plans ([https://www.comune.roma.it/web-resources/cms/documents/Fasc3\\_RischioFrane\\_2021.pdf](https://www.comune.roma.it/web-resources/cms/documents/Fasc3_RischioFrane_2021.pdf)), the fragmentariness of landslide inventories available for Rome, together with the frequent lack of information regarding the date of occurrence of the surveyed landslide (re)activations, severely limit the possibility of evaluating the temporal probability of occurrence. On the other hand, a hazard assessment as detailed as possible can be useful at least for advisory purposes.

In this study, we present a first attempt to quantify the landslide hazard by exploiting the existing databases reporting relevant information for the investigated phenomenon, such as landslide

inventories and other predisposing, preparatory and triggering factors. First, we performed a full review of the landslide databases by means of a geomorphological review of the known and mapped landslides and a check of the available dated landslide events. We then produced an integrated and edited landslide inventory; on this basis, it has been possible to perform a spatial hazard assessment that required additional efforts (i.e., “inventory boosting”) to make such an information fit the requirements needed for a reliable analysis. Furthermore, a statistical analysis of historical rainfall data was performed through the generalised extreme values (GEV; Jenkinson 1955) method for both daily and hourly rainfall. Such an analysis allowed us to infer quantitative rainfall thresholds and assess the return period of rainfall events capable of inducing landslides.

In addition, a preliminary overview of the slopes potentially susceptible to landslides was derived through the persistent scatterer interferometry (PSI) based on the Sentinel-1 SAR images. The presence of movement or changes in the state of activity of already catalogued phenomena were detected by analysing InSAR data, and the results have been integrated with the susceptibility analysis, to add information on the ongoing slope deformations to the intrinsic characteristics of the slope, prone to landslide. For the same level of susceptibility, this integration allowed to rank the actual critical situations by highlighting areas with ongoing deformation, then providing a cognitive dynamic information of practical use for local administrators involved in the management of urban areas.

#### A brief outline of geohazards in Rome

The city of Rome lies in a hilly area where the present geological and morphological setting is mainly related to deposition and erosion in marine (late Pliocene–lower Pleistocene) and continental (middle Pleistocene–Holocene) environments.

The geological context is featured by a succession of Pliocene marine clays (Monte Vaticano Formation), silt and silty sands (Monte Mario Formation) in Lower Pleistocene which turn into littoral, transitional and continental sediments (Ponte Galeria Formation) during lower–middle Pleistocene (Marra and Rosa 1995). These deposits largely outcrop on the hills on the right bank of the Tiber (Funciello and Giordano 2008). Over these terms, a succession of alternating volcanic deposits (resulting from the activity, about 600 ka ago, of the surrounding Colli Albani and Sabatini Volcanic Districts) and continental (mainly alluvial and palustrine) sediments deposited in the area during middle–upper Pleistocene. The present landscape is strongly influenced by valleys and slopes carved by the Tiber River and its tributaries, and only partially filled by alluvial deposits whose thickness can reach tens of metres (Bozzano et al. 2000; Fig. 1).

The hydraulic circulation is governed by the superimposition of medium to high permeability volcanic deposits over the regional aquiclude made up of the blue clays of the Monte Vaticano Fm. and the silty and clayey horizons of the Monte Mario Fm. (La Vigna et al. 2015). Ephemeral and perched water tables develop in the weathered soil covers that transiently over impose as a surficial seepage within unsaturated soils.

The different geological units outcropping on the left and right embankment of the Tiber River valley control the different responses to natural and anthropogenic hazard processes that affect the urban area (Funciello et al. 2008), including subsidence,

sinkholes, floods and landslides (Table 1). In particular, the landslide response is closely related to the outcrop distribution of Plio-Pleistocene sedimentary units, which are frequently involved in shallow and translational landslides (Amanti et al. 2008a, b; Alessi et al. 2014). Ground instabilities are often localised within unsaturated, shallow soil covers resulting from the chemical and physical weathering of the underlying deposits (Schilirò et al. 2019). Ephemeral hydraulic circulation develops along permeability contrasts between volcanic, or debris covers over less permeable sedimentary deposits, can influence the local hydraulic circulation and, thus, the slope stability. Furthermore, falls and topples sometimes involve rocky slopes, mainly made up of volcanic tuffs (Amanti et al. 2008a, b; Alessi et al. 2014).

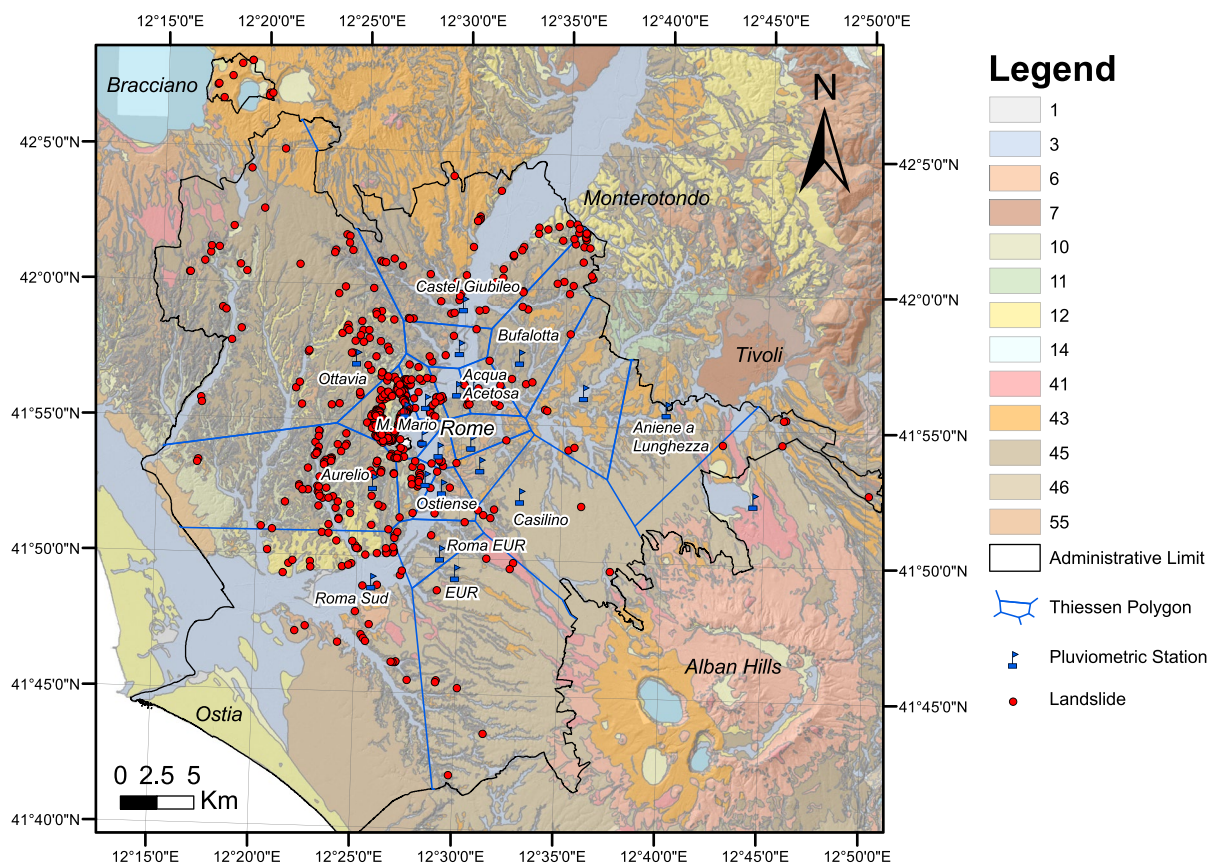
Landslides with volumes of tens up to hundreds of cubic metres (Amanti et al. 2013; Bozzano et al. 2006) as well as natural or anthropogenic sinkholes (Esposito et al. 2021,) are the most frequent and impacting processes in the urban area. Based on the 2014 event (Alessi et al. 2014) and other information collected from a subset of the known landslides (e.g., <http://sgiz.isprambiente.it/frane/roma/>), most of the slope failures have been triggered by intense or prolonged rainfalls. Such events can cause significant damages specially to pipelines, aqueducts and road infrastructure.

#### Materials and methods

Beyond the general goal of providing a first comprehensive landslide hazard analysis in Rome, the main specific objectives of this research can be summarised in the following: i) dataset preparation, ii) susceptibility assessment, iii) in-depth check of multi-temporal information about landslide (re)activations and hydrological analysis with the purpose of evaluating a temporal probability of occurrence and iv) testing the integration of remotely sensed displacement data and susceptibility zoning to provide refined information about critical areas (i.e., those prone to landslide and experiencing an actual deformation).

Basically, whatever the analysis approach chosen among the large variety of those available (Reichenbach et al. 2018), all the methods for susceptibility assessment are based on landslide inventories, that is, the fundamental element of input and/or validation of the analysis (Corominas et al. 2014). Furthermore, the quantification of relationships among causative factors and landslide occurrence requires the acquisition of as many datasets and related information as the number of variables (usually of environmental type, such as geo-thematic and morphometric data) related to predisposing/preparatory and, sometimes, triggering factors (Reichenbach et al. 2018; Fell et al. 2008).

In order to predict the location of potential slope failures (i.e., landslide susceptibility) and to evaluate their intensity and temporal probability of occurrence, the landslide inventories must include information about the exact timing, size and mechanism of the slope movements. The definition of landslide recurrence thus requires at least a raw dating of the events (monthly and yearly precision) to get a reliable statistical assessment of the return period. In case of availability of more precise information (i.e., day and hour), for rainfall-induced landslides, it could be possible to assess the landslide frequency as the recurrence interval of the causative factors which, in turn, can be assessed by hydrological analyses.



**Fig. 1** Lithological units in the study area of the municipality of Rome with the location of the occurred landslides. Thiessen polygons adopted considering the available pluviometric stations are also reported in the map. Key to legend: 1: anthropic deposits; 3: recent and terraced sandy-gravelly alluvial deposits, eluvio-colluvial deposits; 6: silty-sandy alluvial deposits, fluvio-lacustrine deposits; 7: travertines; 10: Plio-Pleistocene clayey and silty deposits; 11: marine Pliocene clays; 12: debris and talus slope deposits, conglomerates and cemented breccias; 14: Marls, Marly limestones and calcarenites; 41: leucititic/trachytic lavas; 43: lithoid tuffs, pomiceous ignimbritic and phreatomagmatic facies; 45: welded tuffs, tuftites; 46: pozzolanic sequence; 55: alternance of loose and welded ignimbrites

Since the precise and complete timing of a landslide in a database is the most valuable but often lacking attribute, the full hazard definition represents the most challenging task for both scientific community and local administrators (Corominas et al. 2014).

### Inventory review and dataset preparation

Our study started from the acquisition and review of the available data sources, including landslide inventory, as well as rainfall time series available in the urban area of Rome. The available landslide inventories in the Rome municipality area include the following:

1. The Official inventories covering the national territory, among which the Aree Vulnerate Italiane (AVI 1996) project (1996), the Italian Inventory of landslide phenomena (IFFI 2007) now distributed as open access in the IDROGEO Platform (Iadanza et al. 2021), the Tiber River basin hydrogeological risk mitigation plan (PAI 2012) and Rome's regulatory plan for the urban area (PRG Roma 2007). The IFFI project inventory is the national DB of landslides that collects information from different sources. The DB architecture is designed to provide a wide range of information for each landslide, but often lacks

or has incomplete data. Mapping criteria and landslide geometries are inconsistent, with some landslides identified as points and others as polygons. Additional processing is required to extract the detachment areas, which are necessary for susceptibility assessments. In addition, key information such as the date of (re)activation, volume assessment, geotechnical parameters, landslide type and state of activity are not always available in the area. The ISPRA's "Frane Roma" Project is a database of landslides in Rome, providing location, source of information and description of the event based on Amanti et al. (1995, 2008c, 2013); ISPRA (2014); Ventriglia (1971, 2002). The CERI provides a database of landslides triggered by rainfall in 2014, surveyed on the field (Alessi et al. 2014). The geomorphological map of Rome by Del Monte et al. (2016) covers the downtown area and integrates previous studies with remote analysis and field surveys.

During the collection and review of the available sources, database completeness and homogeneity analyses were approached, analysing the documental sources and comparing the different elements of the ancillary catalogues to get a unique, coherent and corrected data source for subsequent analyses.

**Table 1** Summary of the main geological hazard involving the municipality of Rome

<i>Involved areas</i>	<i>Latest significant event</i>	<i>Event references</i>	<i>Hazard/susceptibility maps</i>	<i>Inventories and maps</i>	<i>Intensity</i>	<i>Return period (yrs)</i>
<i>Subsidence</i> Alluvial plain of Tiber River and its main tributaries		Campolunghi et al. 2008; Bozzano et al. 2015; 2018; Delgado Blasco 2019;	n.a	-	Low	Continuous
<i>Sinkhole</i> Widespread over the urban area, mainly concentrated in areas with relevant thicknesses of volcanic deposits (left bank of Tiber River);	Via Zenodossio in 2021	Esposito et al. 2021	Ciotoli et al. 2015, 2016; Esposito et al. 2021;	Ciotoli et al. 2013	High	2–5
<i>Flood</i> Alluvial plain of Tiber River and its main tributaries	Cremera Creek Flood in 2014	Remedia et al. 1998;	PAI Flood Hazard Map; Salvo et al. 2017	Di Bersani and Bencivenga 2001;	High	20–50
<i>Landslides</i> Spread over the urban area, mainly concentrated in areas featured by sedimentary deposits (right bank of Tiber River)	Diffuse landslides in 2014 following intense rainfall	<i>This study</i> Alessi et al. 2014; Amanti et al. 2012; 2013; Bozzano et al. 2006	n.a	IFFI catalogue; Del Monte et al. 2016	Medium	5–10



Such an integration has been achieved by a sequence of actions:

1. Merging of available data to i) check and erase duplicates and, ii) if needed, correct the location and/or attribute a type of movement, by means of DTM- and orthophoto-based interpretations. Specifically, to check for duplicates between the various DBs, buffers of 20 m were created around each landslide: their intersection, if present, could mean a duplicate of the same process, and thus, all the redundant records by a supervised procedure were deleted. The selection was made starting from those most reliable in terms of precision and checking the quality in the geolocation of the surveyed landslide phenomena.
2. Reviewing and standardising the classification in terms of type of movement, since different descriptions for the same typology have been adopted in the native DBs.
3. Ensuring a geometrical homogeneity by associating at each landslide the minimal informative geometry, i.e., a point (landslide initiation point (LIP)), which is located in the topmost part of the corresponding line (main scarp edge) or polygon (whole instability or accumulation) lends itself better to susceptibility analyses for shallow landslides.

The so-defined inventory includes (where available) both landslide polygon and LIPs, despite only points having been adopted to train the susceptibility model. Given the landslide population and the different landslide types available in the catalogues, shallow landslides and earth slide mechanisms have been considered.

### Spatial component of the landslide hazard

From the information contained in the landslide catalogues and their related attributes, the definition of rainfall-induced landslide hazard was attempted, starting from the definition of the spatial attribute of landslide hazard (i.e., the landslide susceptibility).

To analyse the relationship between causative factors and landslide presence/absence, we relied on a data-driven approach featured by machine learning (ML) techniques aimed at inferring the multivariate combination of causative and predisposing factors over stable and unstable sectors of the investigated area. As machine learning algorithms strongly rely on quality and amount of input data, we decided to boost (i.e., modify and make it reliable for a proper hazard assessment) the landslide database. Specifically, we developed a methodology to assess the extent of the detachment areas and to maximise the amount of information for each landslide generating additional “synthetic” LIPs to better catch the variability of predisposing factors within a given detachment area. This solution adopts a bounding box enveloping landslide crowns or create half-circle buffers around point elements indicating non-mappable landslides. This workflow allows the reconstruction of detachment areas that were used to sample the synthetic LIPs randomly over areas located at elevation lower than the crown and thus within the potential landslide source area. Once obtained, the synthetic LIPs were validated by testing the similarity between the distribution of their features with that of the original LIPs. (Chi-squared (CS test) and inverted Kolmogorov–Smirnov  $D$  statistic (KS test). As regards the factors predisposing landslide initiation, the most important and adopted variables were considered based

on the available products. All the DEMs available in the municipality of Rome were evaluated to extract morphometric and hydrological derived variables. A new and more reliable DTM was defined starting from the extraction of the numeric 1:5000 topographic map of contour lines and quoted points referred to the terrain. The DTM was generated by means of the ANUDEM algorithm implemented in the TopoToRaster tool by ArcGIS. Lithological and land use maps available in open access by the Lazio Region (<https://geoportale.regione.lazio.it/>) were collected to derive litotechnical and land cover classes at a scale of 1:25,000. Given the scarcity and inhomogeneity of the available geotechnical data, additional informative variables about the mechanical parameters of soil cover and bedrock were not considered.

Then, some of the most important and used predisposing variables (Reichenbach 2018) were adopted. These variables and parameters are found in Table 2.

To include preparatory factors (sensu Julian and Anthony 1996) in the analysis of landslide susceptibility thematic variable as the mean annual rainfall (*ord\_rain*) was defined as a proxies of middle-term “hydraulic stress” and weathering efficiency which, in turn, can be related to the material strength decrease. The mean annual rainfall was defined for each rain gauge and then attributed to the related Thiessen polygon.

In pre-processing stages, correlations between features, or multicollinearities, were calculated and variables were systematically removed if a pairwise correlation exceed 75% (Mastrantoni et al. submitted).

Explanatory variables were sampled in landslide initiation points (LIPs) and random stable points after pre-processing. Stable areas were defined as non-landslide areas, which were used to train the statistical model. The LIPs and stable point datasets were balanced with a 35–65% ratio of unstable and stable points, respectively, based on the need of maximising predictive ability on positives but at the same time avoiding an excessive distance between the actual statistical population (i.e., actual proportion of unstable and stable areas in the whole territory) and the sampled training datasets, thus limiting an excessive false positive rate. The datasets were then split into 80% for training and 20% for validation. The susceptibility function was trained and tested on 10 K-fold random partitions of the dataset to estimate the accuracy of the predictive model and avoid selection bias.

Given the spatial concentration of landslide effects in the Monte Mario area caused by many landslides occurred during the 31 January–2 February 2014 meteorological event, the landslide LIPs training dataset was split into two different portions, where the first represents the generic catalogue of instabilities that happened before and after this intense event and the second “nested” subset composed by the 2014 event inventory. Since previous studies remark the exceptionality of this rainfall event in the NW sector of the city (Alessi et al. 2014), the above-mentioned subsets represent the result of ordinary (on an average basis) and rare rainfall events, respectively. On the generic landslide database, a single predictive susceptibility function was trained. To derive the most appropriate predictive function, we tested several ML models, among which the extra trees classifier (Guerts et al. 2006) implemented in the Scikit-learn package (Pedegrosa et al. 2011) outperformed the other ones in the study area (Mastrantoni et al. submitted). The extra trees (abbreviation for extremely randomised trees) is an ensemble

**Table 2** Summary of the explanatory variables adopted in the analysis

Category	Variable	Acronym	Scale/resolution	Source/reference
Lithological and hydrogeological factors	Lithotechnical units	<i>litotec</i>	1:25,000	Lazio Region
	Distance to boundary of permeability contrasts	<i>dist_lim_k</i>	10×10 m (Map 1:100,000)	Hydrogeological map of Lazio Region (Capelli et al. 2012)
	Distance to main river	<i>dist_idro</i>	10×10 m (Map 1:100,000)	ISPRA
Primary and derived morphometric variables	Elevation	<i>dtm</i>	5×5 m (Map 1:5000)	CTR Lazio Region
	Slope	<i>slope</i>	5×5 m	-
	Aspect	<i>aspect_8</i>	5×5 m	-
	Relative relief index	<i>rrelief100</i>	5×5 m	Hesse (2010)
	Total, planar and profile curvature	<i>cur_tot;</i> <i>cur_prof;</i> <i>cur_plan</i>	5×5 m	-
	Flow direction	<i>flow_dir</i>	5×5 m	-
	Flow accumulation	<i>flow_acc</i>	5×5 m	-
	Topographic position index	<i>tpi</i>	5×5 m (neighbourhood 100 m)	Jenness (2006)
	Topographic wetness index	<i>twi</i>	5×5 m	Beven and Kirkby (1979)
Anthropogenic factors	Land use	<i>cus</i>	10×10 m	Lazio Region land use map DGR n.953/2000
	Distance to road	<i>dist_road</i>	10×10 m	Open Street Map (OSM) consortium
	Soil consumption	<i>soil_cons</i>	10×10 m	Municipality of Rome

supervised machine learning method that fits many randomised decision trees on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting. In extra trees, randomness goes one step further in the way splits are computed. It does not come from bootstrapping of data but rather comes from the random splits of all observations.

The extra trees algorithm creates many decision trees, but the sampling for each tree is random, without replacement. This creates a dataset for each tree with unique samples. A random subset of candidate features is used for each tree. The most important and unique characteristic of the extra trees is the random selection of a splitting value for a feature. Instead of calculating a locally optimal value using Gini or entropy to split the data, thresholds are drawn randomly for each candidate feature, and the best of these randomly generated thresholds is picked as the splitting rule. This makes the trees diversified and uncorrelated. With this approach, both bias and variance are handled. The former is reduced by using the whole original sample instead of a bootstrap replica. The latter is dwindled by randomly choosing the split point of each node. Since splits are selected at random for each feature in the extra trees classifier, it is less computationally expensive than a Random Forest (Breiman 2001).

Validation of susceptibility product was also performed on training and test datasets expressing the predictive capability of

the model by confusion matrixes and resulting receiver operating characteristic (ROC) curves as well as other metric performance indicators.

Furthermore, considering the value that the susceptibility prediction assumes in correspondence with the LIPs, detection rate curves (DRCs) were constructed, which represent the cumulative percentage of LIPs correctly predicted for increasing susceptibility values, regarded as the separation threshold between the true and false positive condition of each LIP. This method allows setting susceptibility thresholds corresponding to predetermined detection rates, driving the discrete classification of the susceptibility values. In this paper, to avoid coarse “expert” choice on landslide susceptibility value, fixed thresholds on landslide detection rate were chosen, adopting percentile values of landslide detection equal to 50, 75, 95 and 97.5% to divide very low, low, moderate, high and very high susceptibility classes. Specifically, two DRCs were built: one using only the LIPs of the 2014 event (recognised as extraordinary) and the other with the remaining dataset (which mainly includes events triggered by ordinary rainfall). To attribute temporal recurrence to such scenarios and transpose them into hazard, approaches of data integration have been adopted looking for the best definition of the landslide hazard with the available data, shifting from its theoretical definition to operative practices of use for stakeholders.

## Temporal component of the landslide hazard

Once the landslide susceptibility from the generic catalogue of landslide effects was defined, further steps in the definition of the landslides hazard have been addressed, facing it through a multi-stage and multi-level approach, consistent with the quality and quantity of available data.

The analysis of landslide frequency has been addressed by means of i) a preliminary evaluation of landslide event recurrence on the few dated landslides, ii) quantification of large-scale hazard descriptors (Corominas et al. 2014) and iii) detailed quantitative attribution of return periods (RPs) of the rainfall events that caused landslide triggering basing on a hydrological analysis, hereafter discussed.

Given the composition of the above-mentioned landslide database with both generic and event-based inventories, the here-defined resulting landslide susceptibility product can be referred to as much as common “ordinary” scenarios as to the severe and non-ordinary conditions experienced in 2014.

### Preliminary estimation of landslide recurrence

Temporal hazard was preliminary approached by exploiting the available dating's information retrieved in the catalogues. Selecting landslides of same type featured by multiple reactivations, a gross estimation of the RP was attempted, defining the temporal range between the first and last (re)activation. Because of the reduced number of dated landslides events and the presence of replicas, the number of effects that experienced multiple reactivations was defined in 17 out of the 471 total landslides.

### Large-scale quantitative hazard description

Since we face off with large-scale analysis for advisory purposes, low-resolution hazard descriptors at a regional scale can be defined according to Corominas et al. (2014) based on indices defined as landslide density and frequency or landslides/year/km<sup>2</sup>. For this purpose, Thiessen polygons were delineated for rain gauges featured by a decades-wide time series, to spatialise the area of interest of each weather station (Fig. 1). Hence, dated landslides of the same mechanism were considered to evaluate temporal recurrence within Thiessen polygon and to relate it to the spatial landslide density.

Landslide effects with multiple reactivations, i.e., effects were more than one date of rainfall associated, give us the opportunity to calculate the period  $\Delta T$  between the first and the last rainfall-induced landslide and the number of reactivations of landslides ( $n$ ) within each Thiessen polygon, so the ratio  $n/\Delta T$  provides the minimum landslide temporal frequency. From these results, a synthetic hazard descriptor ( $H_d$ ) for each Thiessen polygons was calculated as the ratio of number of landslide activations over ten years.

### Evaluation of RPs of landslide trigger rainfall events

The most advanced insights on landslide hazard were retrieved from the analysis of the recurrence of the landslide triggering factors by analysing the intensity and RPs of the main rainfall events that caused ground instabilities by proper hydrological analysis.

To this aim, a hydrological analysis on rainfall regime was performed by choosing reference pluviometric stations considering the time span of the historical record and the time sampling of the rainfall series (daily or hourly data). An overall time period of 70 years was considered, despite the analysis of rain gauge stations data revealing a reduced continuity in data logging and a time coverage often reduced to several years, thus not sufficient to perform rigorous hydrological analysis. For this reason, only weather stations with a dataset equal to or broader than ten years were adopted, despite at least 20 years should be considered in the analysis for reliable results (Serrano 2010). The available pluviometric stations are reported in Supplementary Materials.

A statistical analysis of maximum rainfall intensity data was performed on daily and hourly data, to evaluate the return period of heavy rainfalls for different durations. The hydrological-statistical analysis of the maximum values requires the cumulative rainfall at different time intervals to calculate the rainfall probability curves. Daily and hourly rainfall data were used to calculate cumulative rainfalls over the territory of the municipality of Rome for time stages of 1, 2, 5, 10, 20, 30, 60, 90, 120 and 180 days and 1, 3, 6, 12 and 24-h, respectively. The generalised extreme value (GEV) distribution (Jenkinson 1955), widely used in extreme event frequency analysis (Fowler et al. 2003), was adopted, which follows the following function:

$$F(x) = \exp \left\{ - \left( 1 + \xi \frac{x - \mu}{\sigma} \right)^{\frac{1}{\xi}} \right\} \quad (1)$$

where  $\mu$ ,  $\sigma$  and  $\xi$  are referred to as the location, scale and shape parameters, respectively. These parameters have been defined by applying the probability-weighted moments (PWM) method (Hosking et al. 1985) based on the maximum values of the above-mentioned rainfall time periods available from the dataset. First, the RPs of each considered variable were obtained by inverting the probability function. Then, the obtained cumulative rainfall value was fitted by a power law distribution to build the rainfall probability curves.

For every certified landslide (i.e., with a defined date), the daily rainfall value recorded at the closest rain gauge was attributed to infer the range of potential RPs of the intense rainfall causing instabilities and, indirectly, the RPs of the landslide event.

We used the cumulative rainfall values recorded at the relevant rain gauge in correspondence of the date of landslide occurrence, and then, we compared these values with both hourly and daily probability curves. Where the hourly rainfall data or timing of landslide occurrence lacks, we referred to the total cumulative daily rainfall, inferring a posteriori the possible range of admissible RPs if concentrated rainfalls (within 3 h) or distributed precipitations (over 24 h) are considered.

Such an inference is needed due to the absence of the time of occurrence of the landslide events, which makes the precise attribution of RP impractical. Based on the rainfall linked to the dated landslides with respect to the rainfall probability curves, it was possible to establish the order of magnitude of the RPs of each landslide available in the record.

Despite the effort to gain quantitative information about the temporal recurrence of landslides and face off with the definition of landslide hazard, this method is threatened by the scarcity of dated landslides and the intermittent recordings of several rain gauges.

## Persistent scatterer interferometry

The landslide hazard quantification in Rome was completed by evaluating the areas where probability can be considered unitary, since they are experiencing slope movements to date and should be prioritised in the risk analyses. Such areas of interest have been identified using satellite remote sensing technique, and the InSAR (interferometric synthetic aperture radar) technique (Massonnet and Feigl 1998; Hanssen 2005). Such analyses provided information on the distribution of ground movement, giving valuable insights into assessing the state of activity of the detected land movements and additional information on possible landslide reactivations of pre-existing landslides. Furthermore, the contribution of such relevant data can integrate the static landslide susceptibility assessment, providing dynamic information on landslide activity. For these purposes, the advanced differential SAR interferometry (A-DInSAR) approach (Ferretti et al. 2001; Kampes 2006) was applied to the study area, by the processing of C-band Sentinel-1 (European Space Agency, ESA) SAR images, in ascending and descending orbital geometries, covering the Rome municipality area. The acquired images cover a time span of about five years (from October 2014 to April 2019) and have been processed through the Persistent Scatterers Interferometry analysis (Ferretti et al. 2000; 2001; Kampes 2006; Crosetto et al. 2016), that is optimal for detecting deformations in an urban environment, where the density of permanent scatterers (PS) is generally high.

The resulting PSs have been filtered after by applying a high temporal coherence threshold ( $> 0.6$ ). The results are represented by the velocity of displacement maps (mm/yr) for both ascending and descending datasets, where the movement velocity is measured along the satellite line of sight (LOS). These results were post-processed by setting specific and reliable thresholds to identify areas experiencing slope movement and select the “landslide candidates” in the PSs velocity maps. Such thresholds were fixed on velocity values of PSs (velocity  $< -2.5$  or  $> 2.5$  mm/yr) and PS location, considering only PSs in areas with slope  $> 5^\circ$ .

## Data integration

In order to perform an accurate and reliable landslide hazard analysis for the city of Rome and shed lights on the most critical sectors in terms of first-generation landslides and state of activity of existing ones, approaches of data integration can be used, combining the results of spatial and temporal hazard, adding on its evidence of slope movements derived by PSI analysis.

The raw attribution of landslide recurrence in each Thiessen polygon has been combined with its landslide susceptibility, thus attributing it to the highest susceptible areas, zoning the territory by landslide hazard indicators. Specifically, from the above-described susceptibility analysis, a landslide hazard index ( $H_i$ ) was defined combining, cell-by-cell, the susceptibility level with the classes values of the large-scale hazard descriptors ( $H_d$ ) expressing the mean landslide temporal recurrence in every Thiessen area.  $H_i$  has been thus defined as follows:

$$H_i = \text{Susceptibility} / (1 - H_d)$$

The results of  $H_i$  has been after reclassified to rank the hazard in view of final user requirements. In this paper, a 5-class reclassification based on natural breaks was adopted. Then, potential landslide source areas have been extracted within the high and very high susceptibility classes defined considering the intensity-based DRC classification.

Hence, if two intensity scenarios are considered and two DRC-based classification adopted, it is possible to obtain hazard maps concerning different intensity levels, delineating potential landslide source areas reactivable under different ranges of RPs, achieving a first robust landslide hazard assessment in Rome.

## Results

### Inventory review and dataset preparation

From the collection and comparison of the available databases, we noticed a certain degree of fragmentation of information, such as lack of heterogeneity and consistency, record repetition, and different attributes related to landslide effects. The differences are evident also about number of elements, publication date (from 1995 to 2018), description of the type of movement (if present), updating, purposes and scale of observation (from national to local). Moreover, the data associated with the landslides are not homogeneous for what concerns temporal information. Certified (i.e., dated) landslides are available only in the “Frane Roma” ISPRa Project. After the data collection, the geometry of the databases also appears heterogeneous: within the same database, shapefiles can vary from points or polygons to points–lines–polygonal vectors. However, many inventories use a single geometry as a comprehensive outline for source transit and deposit areas.

According to all the available sources, the uniformed DB is composed by as many points as the filtered landslides (566), including earth slope instabilities (earth slides and shallow, i.e., soil slip) movements, rockfalls and remaining n.d. effects. Most of the landslides in the city of Rome occurred in the last decades (Fig. 2), with slope failures that have almost always been associated with heavy rainfall between 2008 and 2014 (67 landslides out of the total occurred during the storm of 31 January 2014; Alessi et al. 2014).

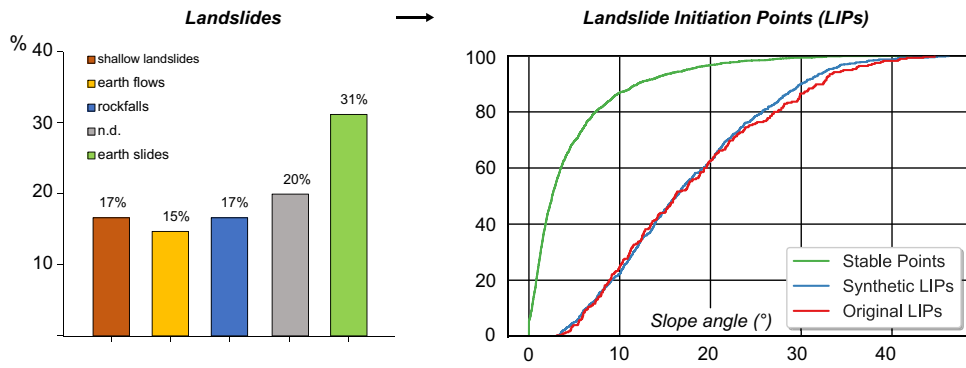
Based on the reviewed catalogue, the dataset of LIPs to be considered for the susceptibility analysis was defined, accounting for shallow earth failure mechanisms only, hereafter considered as a unique ensemble. Test of validation of the synthetic vs original LIPs was performed through the evaluation of similarity of the slope angle distribution between synthetic and original LIPs (Fig. 2). CS and KS tests resulted in scores of 0.666 and 0.923, respectively.

Hence, the final point-based landslide database is represented by 1099 LIPs (289 original and 810 synthetic), excluding the 67 related to the January 2014 extreme rainfalls.

### Spatial component of the landslide hazard

The here conducted analysis allowed us to point out the landslide susceptibility to earth slide and shallow failures in the municipal area of Rome. The continuous landslide susceptibility map shown in Fig. 3 reveals local maxima in the right embankment of the Tiber River, in the Monte Mario and Monte Ciocci ridges, where the M. Vaticano and M. Mario sedimentary units crop out. The portion of Rome west of the Tiber River shows higher susceptibility with

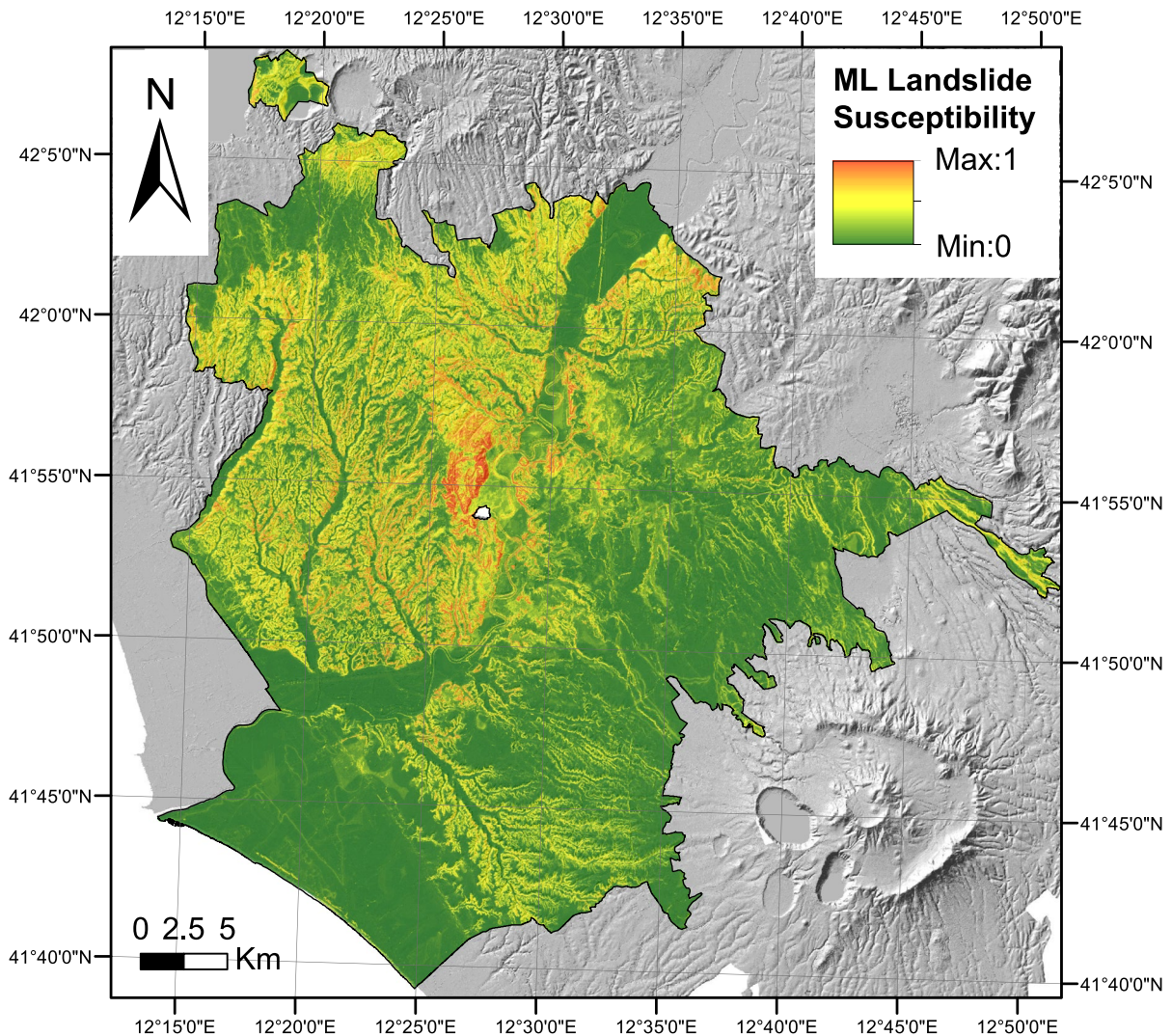




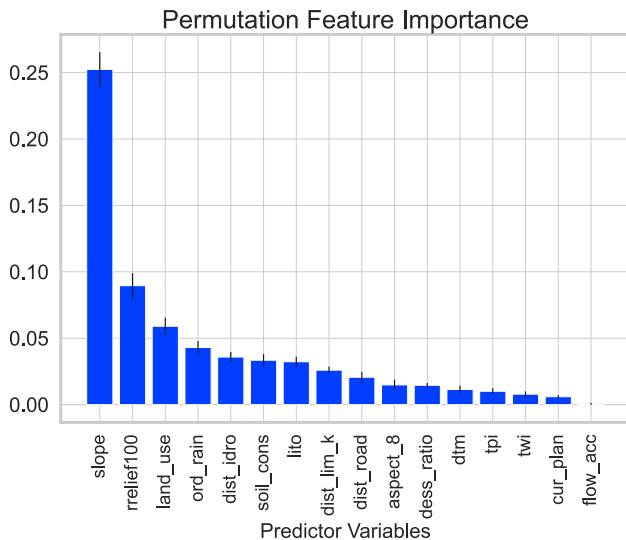
**Fig. 2** a) Distribution of landslide effects available in the catalogues over the territory of Rome. b) Distribution of LIPs and stable points vs. slope angle adopted in the susceptibility model

respect to the eastern one because of the presence of tuffs and pyroclastic deposits of Sacrofano and La Storta unit from the Sabatini Volcanic District (Sottili et al. 2004).

The most important predisposing and preparatory variables can be identified in the feature importance permutation graph in Fig. 4, where slope angle and relative relief (*Rrelief100*) are the most



**Fig. 3** Continuous map of landslide initiation susceptibility of Rome area according to the adopted ML model



**Fig. 4** Feature importance resulting from the extra tree classifiers model

conditioning features, which showed a Pearson correlation coefficient equal to 0.7, thus acceptable since lower than the threshold of 0.75 recommended by Kuhn and Johnson (2013). Relevant importance can be attributed to lithology and land use types as well as to the distances to rivers and permeability limits, which can be considered a proxy of hydraulic circulation and the presence of temporary water tables. A marked preparatory role was played by the predictive variable of the annual rainfall, whose distribution can be considered a factor capable of controlling soil moisture, a stressor for slope stability.

The quantification of susceptibility performance revealed on the test dataset by the confusion matrix and resulting ROC curves highlighted the very high quality of the function in predicting both stability and instability (Fig. 5), especially after the definition of the hyperparameters. An area under curve (AUC) value of 0.96 was found, with excellent specificity and sensitivity.

Afterwards, the susceptibility map was reclassified according to the DRCs extracted with respect to subsets of landslides available in the catalogue and referable to the 2014 event and

the rest of the catalogue. From this point, two different scenarios were considered adopting the two reclassification criteria that account for the different susceptibility conditions referable to events of different intensity.

Two different detection rate curves were thus extracted from the dataset, including LIPs of the 2014 event and the other antecedent 2014 (pre-2014 in Fig. 6), to fix susceptibility thresholds under fixed detection rates. Such thresholds drove the discrete classification of the susceptibility in five classes. For example, moderate to very high susceptibility classes moved from susceptibility equal to 0.27 to values greater than 0.65 if the rare scenarios are considered (Fig. 6).

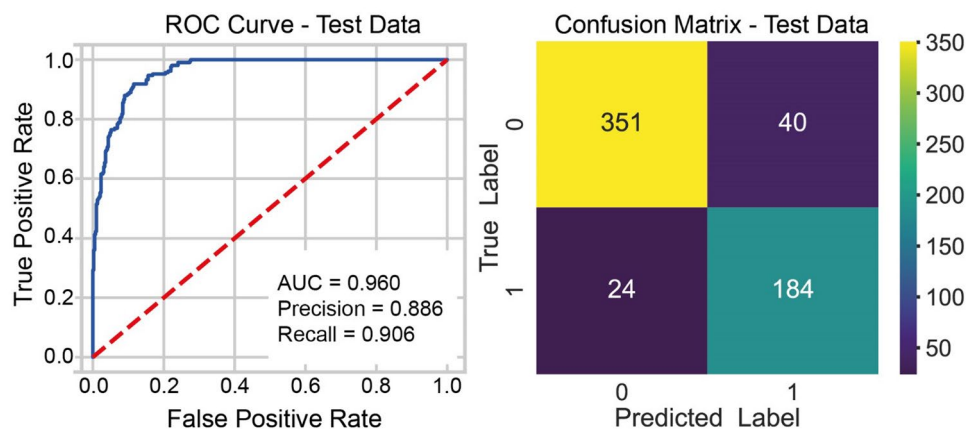
The spatialisation of the so-defined susceptibility classes points out how Monte Mario and Monte Ciocchi hills face a limited increase in landslide susceptibility when moving from the ordinary to the rare scenario that reflected in limited variations of the areal extension of the high and very high susceptibility classes (Fig. 7). Larger differences can be found in the right embankment of the Tiber River, where wide areas fall into the moderate and high classes given the reduction of the class breaks from 0.65 to 0.27 (Fig. 7).

The moderate, high and very high classes in the ordinary cover a percentage of the area of 3.73%, 0.32% and 0.10%, respectively, which increase up to 10.50%, 1.64% and 0.37% in the rare scenarios. Adopting the two different classifications, it is possible to consider two different landslide intensities (intrinsic in the event-based inventory) and refer the susceptibility analysis to pluviometric events with increasing RPs.

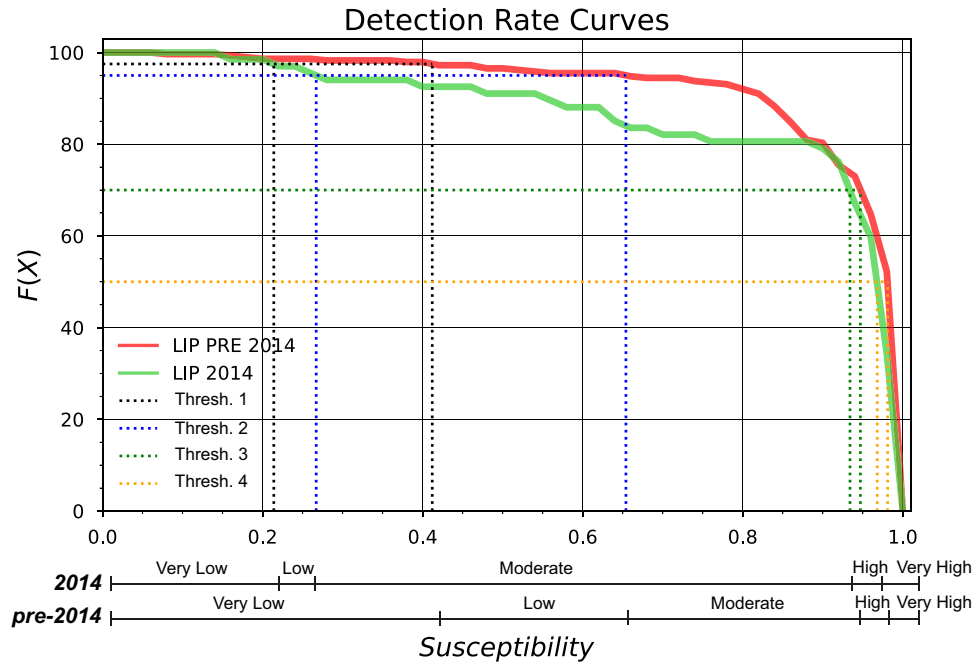
From the obtained results it is possible to state that limited differences among the two intensity scenarios can be found. This similarity relies on the common average size of the landslides inventoried in the generic and 2014 event catalogues, as well as on the similarity of landslide triggering pluviometric inputs. For these reasons, a univocal landslide scenario and a resulting map can be considered valid for depicting landslide hazard in Rome.

### Temporal component of the landslide hazard

With the aim of associating RPs to the different zones of the territory, a preliminary definition of landslide recurrency was



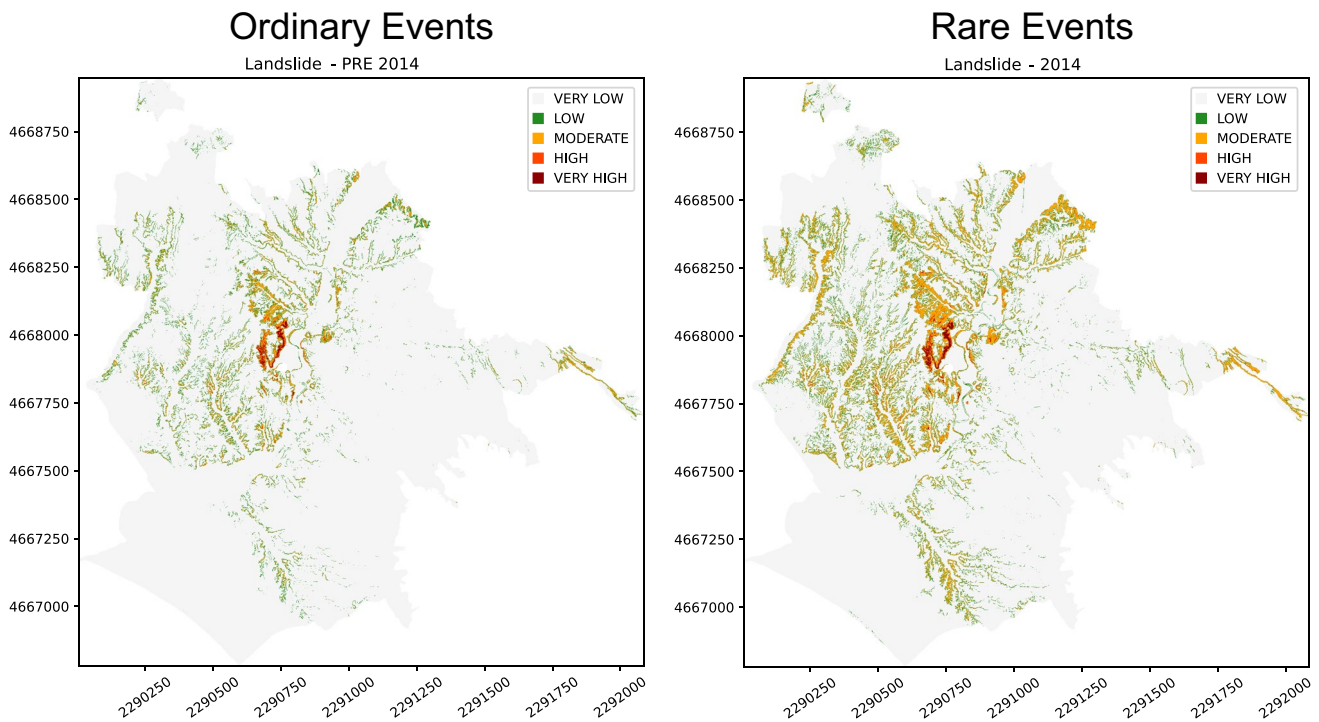
**Fig. 5** ROC curve defined by the confusion matrix on the testing datasets. AUC, Precision and Recall are also reported



**Fig. 6** Detection rate curves obtained on the susceptibility results considering LIPs antecedent the 2014 meteorological event, and dated at 31/01/2014, drive the discrete classification of the susceptibility

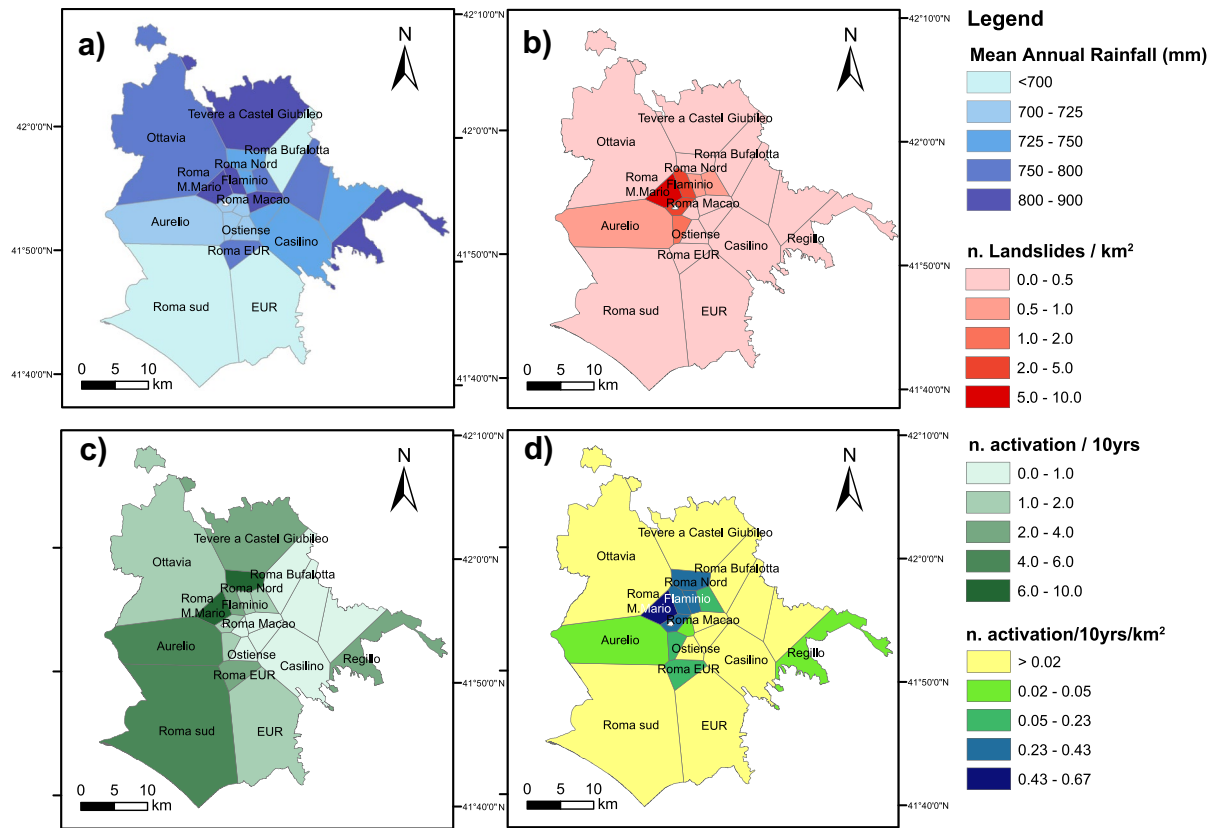
approached by the analysis of temporal range among subsequent occurrences of landslide-triggering meteorological events; first evidence from the few reactivated landslides came out by the preliminary attempts: on the 17 reactivated effect, the return period ranges between 1 and 25 years with a mean value of 5 years.

The large-scale analysis on dated landslides, which uses hazard descriptors, pointed out the higher landslide density in the W and NW sector of Rome for shallow and earth slide failure mechanisms. The landslide density is in a strict relationship with the mean annual and maximum daily rainfall (Fig. 8).



**Fig. 7** Classification of landslide susceptibility for earth failure mechanisms in the municipal area of Rome according to detection rate thresholds defined for ordinary scenarios (left) and rarer events (right) like the one that occurred in 2014





**Fig. 8** **a** Mean annual rainfall in Rome according to rain gauges with a minimum time coverage of 20 yrs. **b** Spatial density (landslide/km<sup>2</sup>) of shallow landslides and earth slides over the total area of Rome. **c** Landslide temporal return period (no. of activations/10 years) of shallow landslides and earth slides over the period from first to last dated landslides. **d** Synthetic descriptors of hazard ( $H_d$  no. activations/10 years/km<sup>2</sup>)

The resulting landslide frequency fairly matches the one obtained by the preliminary analysis on return periods of reactivated (and multi-dated) landslide, whose values range from 2 to 10 years.

In-depth information on temporal hazard has been also derived from the hydrological analysis of daily and hourly rainfall data, resulting in a series of rainfall probability curves enclosed in the Supplementary material. Some of these curves are also reported in Fig. 9. The here-conducted analysis allows to evaluate the cumulative rainfall for every pluviometric station assuming different RPs, updating the analysis carried out by Alessi et al. (2014) to the end of March 2021 and confirming that the January 2014 storm is characterised by RPs up to 100 years.

Based on the few landslides featured by a certain date and assuming as critical rainfall the cumulative, daily rainfall fallen until the landslide date, it is possible to qualitatively evaluate the range of RPs of the triggering rainfall by comparing the daily and hourly cumulative rainfall obtained from the GEV analysis (Fig. 10).

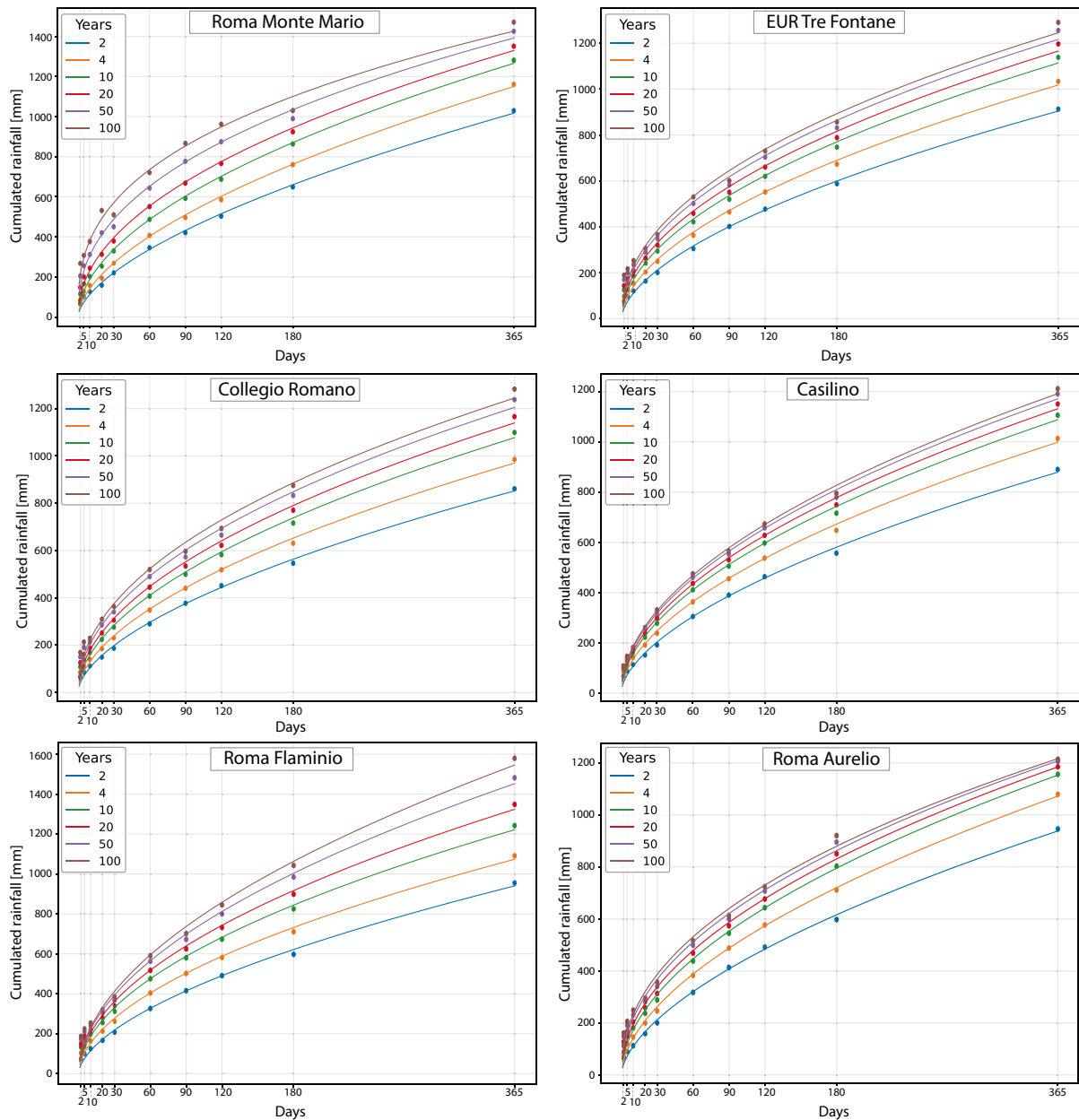
It is possible to note that in the municipality of Rome, the pluviometric events responsible for landslide triggering are featured by RPs below 2 years for most of the rainfall-induced effects. On the contrary, RPs for January 2014 and December 2008 rainfall events that, resulted in RPs always above 10 years.

In detail, the graphs in Fig. 11 show the results of the rainfall data processing plotted on the rainfall probability curves. It is evident that  $TR < 2$  values prevail for rainfall-induced effects that occurred before January 2014, as well as the large amount of “No Data” testifies the lack of rainfall record (Fig. 11b). For ordinary rainfalls, instead, the comparison with rainfall probability curves obtained by GEV highlighted that more than 60% of dated landslides are associated with RPs among 2 and 10 years (Fig. 11b). The landslide that occurred on 31/01/2014 significantly show the exceptionality of the event, with more than 25% of the landslides with RP values between 10 and 50 years and complementary ones ranging between 50 and 100 years (Fig. 11c).

This confirms the exceptionality of the 2014 rainfall event, as stated by Alessi et al. (2014), that, however, reflects in average sized landslides. Furthermore, they retrieved its uneven areal distribution and the peak of rainfall in the NW sector of Rome, which is the most susceptible and where the higher landslide density and frequency exist (Figs. 1, 3 and 8).

Given the results of preliminary evaluation, hazard descriptors at large scale and detailed hydrological analysis, site-specific landslide scenarios can be defined, attributing an average landslide intensity and associating temporal attributes to the static susceptibility.





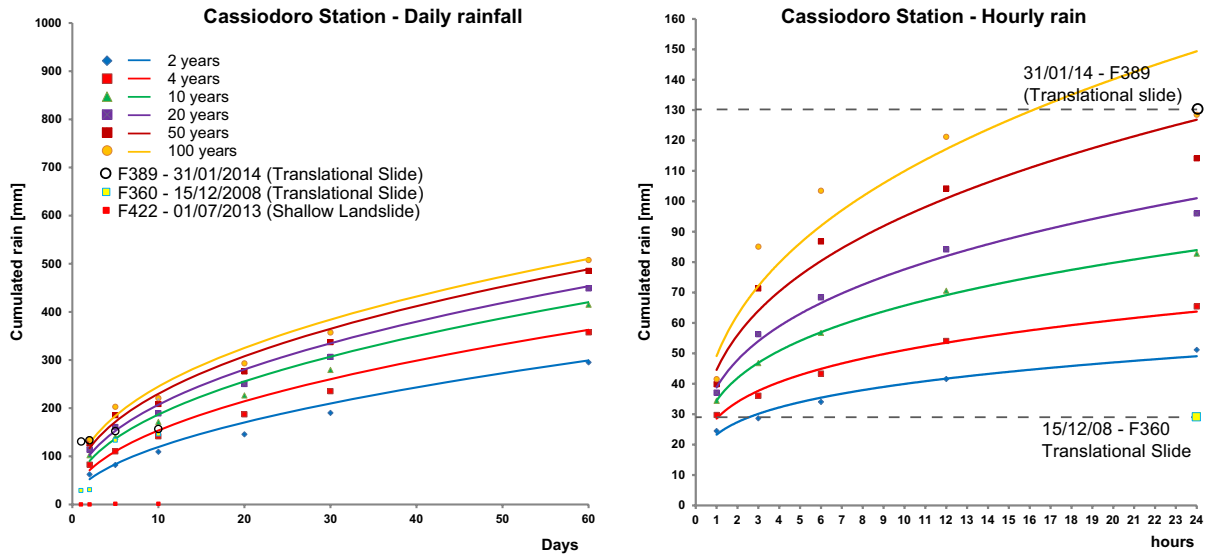
**Fig. 9** Exemplary rainfall probability curves obtained in the municipality of Rome by the GEV analysis on daily rainfalls

### Persistent scatterer interferometry

The complementary interferometric analysis carried out and the resulting A-DInSAR velocity maps referred to the period October 2014–April 2019 (Fig. 12) represents the first steps in the dynamic update of the static landslide susceptibility map.

The detected clusters of PSs affected by displacement allowed us to identify a total of 19 moving areas featured by slope angle  $> 5^\circ$  (only 3 clusters are in areas with slope  $< 10^\circ$ ). It is worth noting that the most significant part of the urban slopes, where the highest slope angles and susceptibility values were found, such as those of Monte Mario, are generally covered by vegetation that limits the possibility of retrieving PSs. Further analysis on these clusters have to be conducted.

A-DInSAR analysis identifies the areas where a slope movement is taking place during the analysed time span (landslide candidates) and reveals the active deformation phenomena. The comparison of the results of the PSI analysis with the landslide susceptibility map highlights the convergence of the landslide candidates and the area more susceptible to landslides, as well as potential source areas in the very high susceptibility classes (Fig. 13). The PSI results were also compared with the lithology outcropping in the study area (scale of 1:5000). Different clusters of PSs in motion are found in correspondence with tuffaceous and clayey lithologies or generally in high susceptibility areas (Table 3).



**Fig. 10** Extraction of range of RPs from the comparison between daily rainfall and results of the GEV analysis derived from daily and hourly rainfall records

**Data integration**

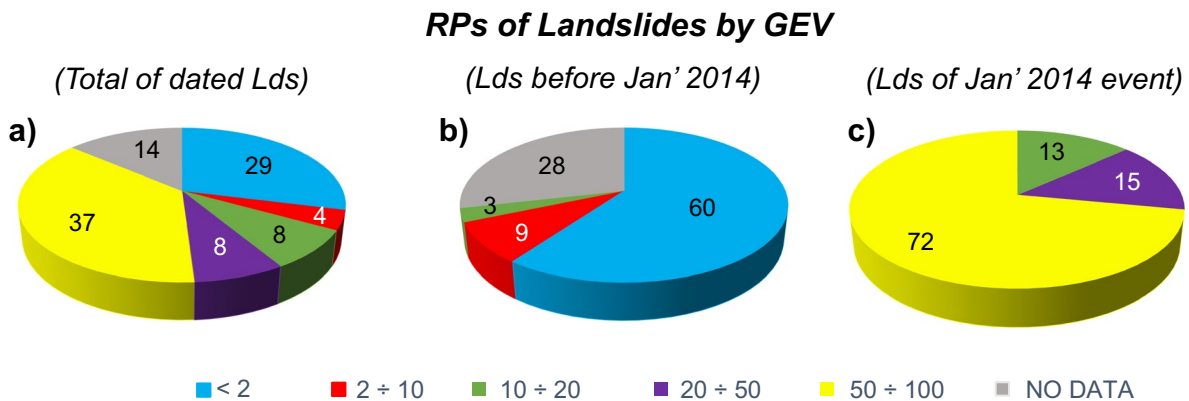
Thanks to the integration of spatial susceptibility with temporal attributes resulting from the definition of hazard descriptors or hydrological analysis, a first landslide hazard quantification results from the reclassification and hierarchization of susceptibility, according to their intensity scenario and mean recurrency of the triggering input.

Landslide susceptibility results were combined and integrated with temporal attributes derived from hazard analysis on the subset of dated landslides. Hazard index ( $H_i$ ) map pointed out the maximum hazard in the ridges of Mt. Mario and Mt. Ciocci, where the maximum value of hazard was derived, strengthening the results of spatial hazard analysis. Similar impact can be assessed in the susceptible slopes falling in the Roma Flaminio, Acqua Acetosa and Roma Nord areas, where relatively higher  $H_d$  were found. The hazard index classification obtained and compared with the reclassified susceptibility map was reported in Fig. 13.

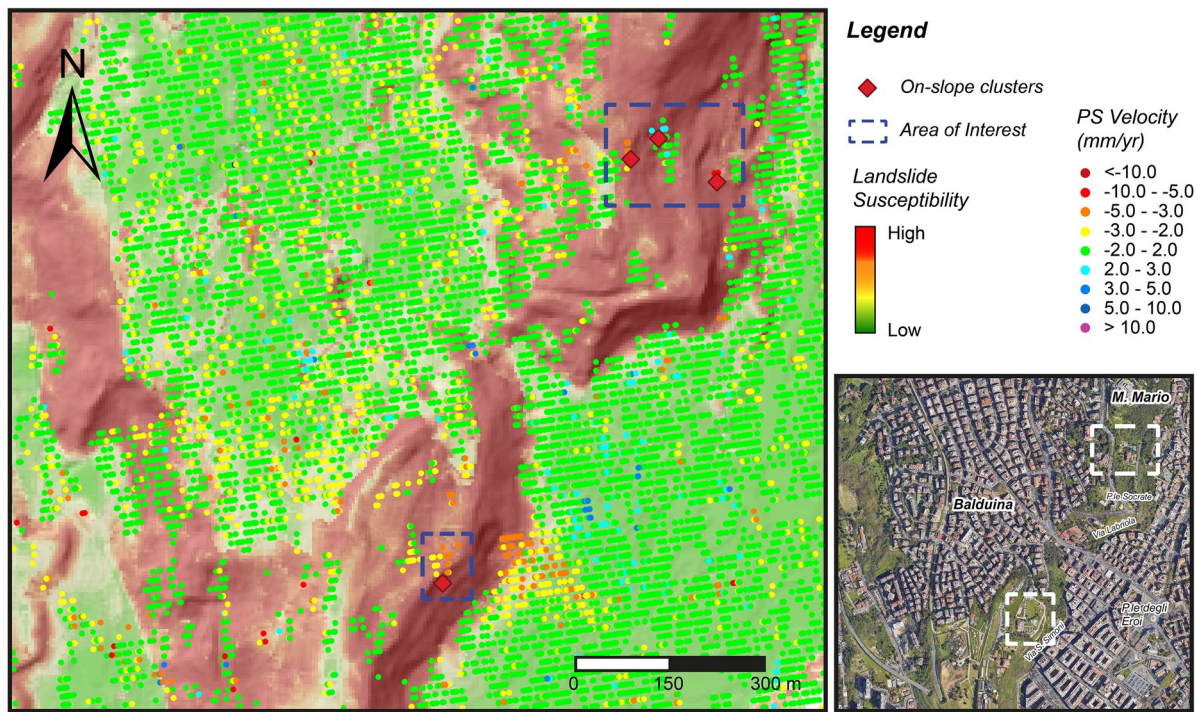
**Discussions**

Multivariate statistical analysis, including machine learning models, can identify landslide patterns by analysing various input data, such as an ancillary landslide database. However, data quality is often inadequate due to incomplete, heterogeneous and erroneous data, leading to biased susceptibility maps (Steger et al. 2017). Spatial heterogeneity in DBs has proved to cause bias in landslide susceptibility maps (Loche et al. 2022). Open-source landslide inventories may have erroneous geometries and positional errors, resulting in sparse and unreliable data (Steger et al. 2016). To improve reliability, low-accuracy datasets must be integrated adequately in terms of quality and quantity (Mastrantoni et al. 2022; Titti et al. 2021).

The study involved a GIS- and ML-based combined approach to collect, check, cross-validate and integrate open-source landslide inventories into a single database. However, the number of LIPs was too low to train reliable ML models, so we developed a methodology to derive synthetic LIPs. This helped to improve the overall



**Fig. 11** Pie chart showing the percentage of landslides with different RPs considering the entire database (a), before the 2014 event (b) and during January 2014 (c) meteorological event

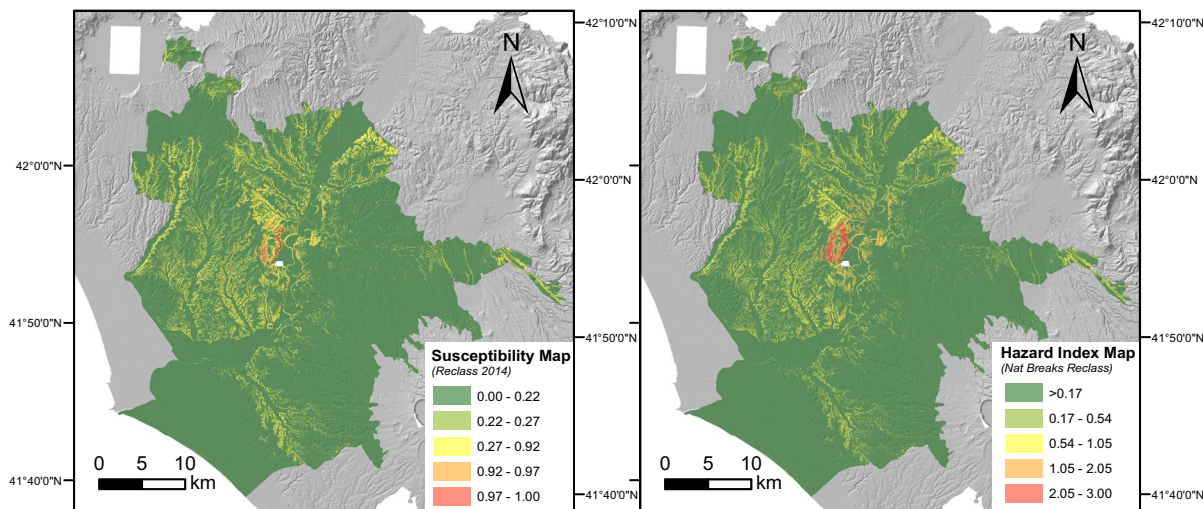


**Fig. 12** PSs map and moving on slopes clusters extracted from PSI analyses on Sentinel-1 images from October 2014 to April 2019. Ascending orbital geometry is taken as an example

consistency of the original database and allowed for more accurate ML model training.

Spatial hazard was resolved by means of ML approaches, providing the first advanced landslide susceptibility zoning useful at a statutory level. From the continuous landslide susceptibility map, a discrete reclassification was performed by means of DRCs to encompass the different intensity and account for different intensity scenarios. Results pointed out how similar is the areal extension of the two classifications within high and very high

susceptibility classes. These outcomes rely on the most common type of landslides in Rome, which are mainly shallow (soil slips and translational slides) and comparable in volumes (from some cubic metres to several tens of cubic metres). On these assumptions, we can consider the obtained hazard maps as landslide scenarios of a specific intensity and a certain pluviometric event (i.e., a certain number of landslides simultaneously triggered during a rainfall event) rather than a uniform hazard analysis. Landslide scenarios associated to ordinary and rarer pluviometric input were evaluated,



**Fig. 13** Susceptibility (left) and hazard index map (right) resulting from the hazard analysis in the municipality of Rome



**Table 3** List of the areas of Interest resulting by combination of susceptibility analysis and persistent scatterer interferometry

Area of interest	PS (mm/year)	Susceptibility class
Via Simone Simoni	3–5	5
Via Fedro	3–5	5
Viale Platone	5–10	5
Passeggiata del Gianicolo	3–5	5
Via Misurina	5–10	4
Via F. Scaduto–Villa Veschi	5–10	5
Via Trionfale–Via Igea	3–5	2

assuming the temporal information either from preliminary evaluation on landslide reactivation recurrency or from large scale hazard analysis or detailed hydrological analysis. These data help us to integrate the spatial hazard, by attributing to the different areas of the municipality of Rome average RPs and hazard descriptors. Given the low number of dated landslides, more detailed analysis on landslide triggering conditions was not approached.

In this study, we constrained order of magnitudes for ordinary landslide frequency, as the estimations are based only on the RPs of the recurrent landslides in which multiple dates are available. To overcome such limitations and provide RP and critical rainfall intensity ranges for the few dated landslides, the antecedent rainfall

registered before the event was compared with the standard rainfall probability curves obtained for every pluviometric station. Despite the scarcity of data, landslide frequency in Thiessen polygons and comparison between dated landslides and rainfall probability curves provided similar results.

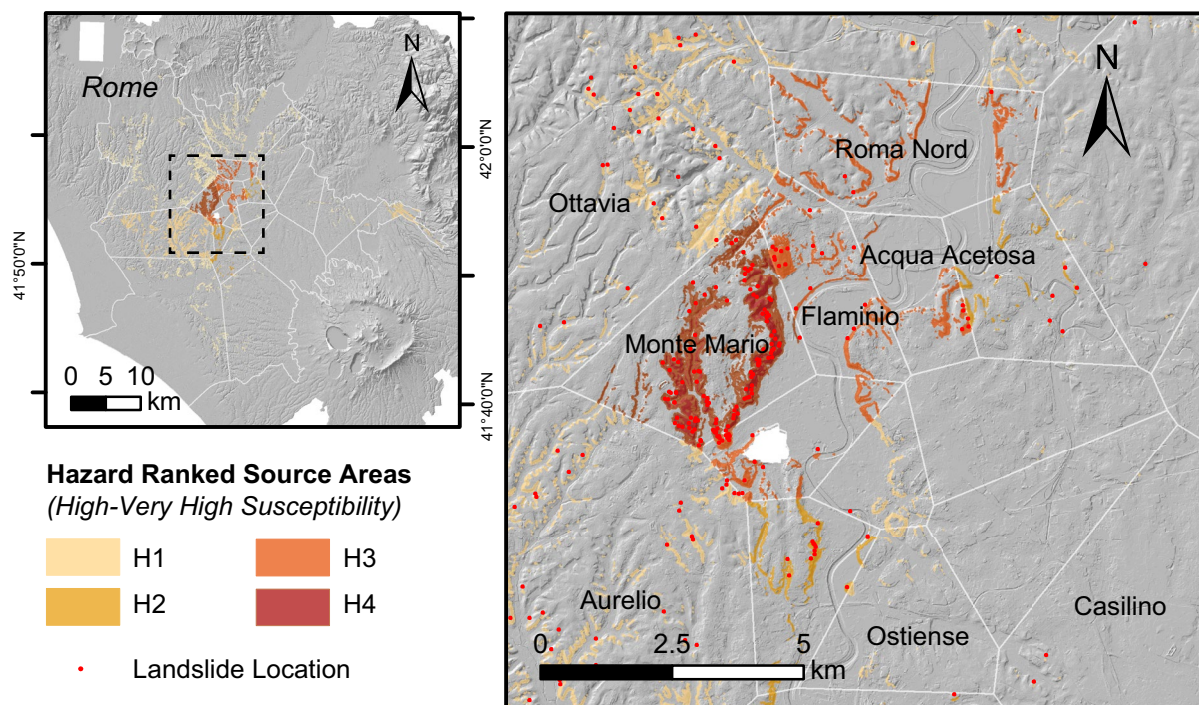
Despite the effort to gain quantitative information about the temporal recurrence of landslides and face off with the definition of landslide hazard, this method is intrinsic limited by the scarcity of dated landslides and the intermittent recordings by some of the rain gauges. The lack of detailed information from DBs mainly affects the temporal hazard and minorly the spatial hazard.

Afterwards, potential landslide source areas were extracted from high susceptibility pixels, associating temporal attributes and ranking the hazard zones (Fig. 14).

For landslide risk management purposes and prevention of landslide risk in a highly urbanised area like Rome (Italy), the definition of spatial and temporal hazards and their variations over time can be crucial; however, detailed multi-annual landslide inventories are a necessary and essential fact (Zieher et al. 2016; Fang et al. 2022).

The here-released product can be hence integrated either in case of availability of updated landslide inventories (e.g., insertion of new events and/or refinement of thematic maps), or additional and/or higher-resolution ancillary data become available. This updating process can be followed by adopting defined protocols, aimed at testing the validity of susceptibility models or training new ones. In this sense, cloud-based solutions, tools or routines can be of help to both geoscientists and administrators (e.g., Titti et al. 2022).

By the proposed approach, we provide a way to distinguish different landslide hazard scenarios in Rome based on the fragmented

**Fig. 14** Ranking of hazard in the potential source areas extracted by high and very high susceptibility classes. Locations of landslides sites are also reported



and incomplete data. The reclassification of susceptibility by DRCs also pointed out as landslides in Rome are featured by common intensity (i.e., basically given by landslide size). Furthermore, the spatial susceptibility was joined with the outcomes of the hydrological analysis, which allow linking such scenarios to RPs of 5–10 years for an ordinary rainfall and of 20–50 years for rarer events (like the 2014 event).

In the hazard map so defined, it is possible to locate on map and extract potential source areas for first-time landslides according to their spatial probability, that increase in intensity (i.e., area) the lower is their associated hazard. On these source areas, an additional classification can be performed to differentiate with hazard attributes the proneness to landslides of specific areas of interest (Fig. 14).

The extraction of potential source areas can be done by an arbitrary choice or supervised procedures from high to very high susceptibility classes, or, precautionarily, considering a high rate of landslide detection (e.g., assuming as potential source areas also lower susceptibility values). Alternatively, potential source areas for shallow or translational sliding can be identified over portions of the slope not yet failed and extracted by fixing a defined percentile in the distribution of susceptibility values within class 5 (maximum susceptibility). For a complete overview of the hazard condition in the urban area of Rome, first time occurrence hazard map must be overlaid with location of existing landslides, where the hazard in unitary and landslide reactivations must be expected (Fig. 14). In these hazardous areas, the static landslide susceptibility was linked to the PSI analysis: such an integration allowed to locate and update, in a dynamic way, zones of ground deformations potentially related to landslides. This analysis consists in a second-order check on the current state of activity that allowed to quantify the rate of activity within potential landslide source areas. Thus, the multitemporal analysis allowed to convert the static product into dynamic thematic layers, able to steer the risk managers and infrastructure stakeholders towards the most hazardous sectors in the urban area and size the intensity of the gravitational processes (Esposito et al. 2021). According to the location, distribution and displacement characteristics of the observed PSs, some interesting evidence of slope movements were collected, at least where the PSs coverage allowed it (not in vegetated slopes). Data deriving from the European Ground Motion Service (EGMS) platform by Copernicus will also allow a dynamic update on an annual time scale of the state of activity to rank the most critical situations (i.e., slopes with actual ongoing deformations) among the landslide-prone areas, as suggested for Sinkholes in Esposito et al. 2021. Furthermore, advances in PSI analysis can availing of COSMO Sky-Med (ASI) images that ensure higher resolution and spatial coverage to catch in detail landslide process. For these purposes, approaches that integrate different constellations can be adopted by data fusion algorithms or by adopting advanced photo-monitoring approaches (Caporossi et al. 2018).

## Conclusions

This paper presents the results of a study aimed at defining the landslide hazard in the municipal territory of Rome, with the dual objective of: i) providing a first “integrated” product useful for the prediction—at least in terms of spatial hazard—of first-generation landslides, through the systematisation of all the knowledge available for the area and ii) identifying methods which, starting from

an available database that is not optimal in terms of quality and quantity, would allow to obtain reliable results for implementation on an operational level, e.g., land management by policy making institutions.

As for the first objective, this study made it possible to highlight and quantify the potentially critical conditions for landslides affecting the Roman area. This geohazard factor, although considered in the official planning documents, has always been strongly underestimated with respect to the actual risk conditions: in this sense, only 27 already existing landslides are officially “certified” as situations of potential risk.

At the same time, the work carried out allowed us to develop a methodological approach capable of exploiting the available data. The review of the DB and inventory boosting made it possible to optimise the input information, while the application of ML techniques and the comparison between the results allowed us to select high-performance susceptibility functions. The integration of historical–statistical analyses on the few dated landslides and on the recurrence of triggering factors has also made it possible to define the temporal hazard, at least in terms of average frequency of occurrence. The result of the above-mentioned workflow is an integrated map which identifies areas of concern, classified based on the combination of spatial hazard with high reliability and expressed in quantitative terms, and an expected temporal frequency with lower reliability and measured in terms of “qualitative–quantitative” estimates. The here-defined components of the landslide hazard fulfil the requirements of advisory purposes, thus providing non-binding strategic advice to the management of the territory to address further (and focused) detailed studies in areas where the River Basin Authority could place impeding or prescriptive constraints, based on the severity of the instability. As a matter of fact, the results of this study have already been presented to and shared with local decision-makers and the River Basin Authority of the Central Apennines, starting dedicated interlocutions aimed at its operative transposition. Furthermore, in the frame of an ongoing research contract with the municipal Civil Protection this research represents the basis for the identification of sub-areas of high susceptibility where to evaluate rainfall thresholds for landslide initiation useful for Civil Protection duties (Segoni et al. 2018). The reported products point out the state of activity and the framework of landslide proneness for shallow and earth slide failure mechanisms, representing the operative products supporting decision-makers in the management of criticalities in the territory for an aware setup of monitoring solutions and/or prioritization of investments for prevention and mitigation.

## Acknowledgements

The authors wish to acknowledge Dr Luigi Lombardo and one anonymous reviewer for their effort in revising the manuscript and for valid and expert comments on the topic, which significantly improved the manuscript.

## Author contributions

C.E. conceptualised and coordinated the research activities. A.P., P.C., G.M. and B.A. reviewed the landslide databases. G.M. and G.M.M. prepared the training and test datasets and performed the susceptibility analysis. G.M., L.S., P.C. and B.A. performed the

hydrological analysis. G.M.M., C.E. and P.C. took care of temporal hazard descriptors and data integration. B.A. and F.B. analysed and processed the interferometric data by PSI. G.M.M. prepared the figures and wrote the manuscript. C.E. and F.B. coordinated specific research activities. F.B. provided research grants as principal investigator of the acknowledged fundings. All authors reviewed the final version of the manuscript.

### Funding

Open access funding provided by Università degli Studi di Roma La Sapienza within the CRUI-CARE Agreement. This research has been partly funded by the Italian Ministry of University and Research in the frame of the PRIN project “URGENT-Urban Geology and Geohazards: Engineering geology for safer, resilient and smart cities”.

### Data availability

The data that support the findings of this study are available from the corresponding author upon request.

### Declarations

**Competing interests** The authors declare no competing interests.

**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

Alessi D, Bozzano F, Di Lisa A, Esposito C, Fantini A, Loffredo A, Varone C (2014) Geological risks in large cities: the landslides triggered in the city of Rome (Italy) by the rainfall of 31 January–2 February 2014. *Ital J Eng Geol Environ* 1:15–34

Amanti et al (2008c) In Memorie Descrittive della Carta Geologica d’Italia: La geologia di Roma. Dal centro storico alla periferia 80

Amanti M, Cesi C, Vitale V (2008a) Le frane nel territorio di Roma in La Geologia di Roma – Dal centro storico alla periferia. *Descr. Carta Geol. d’Italia, Vol, Mem, p 80*

Amanti M, Cesi C, Vitale V (2008b) Landslides distribution in the Roma municipality. *Carta Geol. d’Italia, Vol, Descr, p 80*

Amanti M, Chiessi V, Guarino PM (2012) The 13 November 2007 rockfall at Viale Tiziano in Rome (Italy). *Nat Hazard* 12(5):1621–1632

Amanti M, Gisotti G, Pecci M (1995) I dissesti a Roma in La Geologia di Roma – Il centro storico. *Descr. Carta Geol. d’Italia, Vol, Mem, p 50*

Amanti M, Troccoli A, Vitale V (2013) Pericolosità geomorfologica nel territorio di Roma Capitale Analisi critica di due casi di studio: la Valle dell’Inferno e la Valle dell’Almone. *Mem Descr Carta Geol d’Italia Vol. 93 AVI (1996) - Aree vulnerate italiane. A cura di Coordinamento della Protezione Civile al Gruppo Nazionale per la Difesa dalle Catastrofi Idrogeologiche (GNDCI) del Consiglio Nazionale delle Ricerche (CNR) <http://avi.gndci.cnr.it>*

Bersani P, Bencivenga M (2001) Le piene del Tevere a Roma: dal 5. secolo a. C. all’anno 2000

Beven KJ, Kirkby MJ (1979) A physically based, variable contributing area model of basin hydrology/Un modèle à base physique de zone d’appel variable de l’hydrologie du bassin versant. *Hydrol Sci J* 24(1):43–69

Bozzano F, Andreucci A, Gaeta M, Salucci R (2000) A geological model of the buried Tiber River valley beneath the historical centre of Rome. *Bull Eng Geol Env* 59(1):1–21

Bozzano F, Esposito C, Franchi S, Mazzanti P, Perissin D, Rocca A, Romano E (2015) Understanding the subsidence process of a quaternary plain by combining geological and hydrogeological modelling with satellite InSAR data: the Acque Albule Plain case study. *Remote Sens Environ* 168:219–238

Bozzano F, Esposito C, Mazzanti P, Patti M, Scancella S (2018) Imaging multi-age construction settlement behaviour by advanced SAR interferometry. *Remote Sensing* 10(7):1137

Bozzano F, Martino S, Priori M (2006) Natural and man-induced stress evolution of slopes: the Monte Mario hill in Rome. *Environ Geol* 50(4):505–524

Breiman L (2001) Random Forests *Machine Learning* 45(1):5–32

Campolunghi MP, Capelli G, Funicello R, Lanzini M, Mazza R, Casacchia R (2008) Un caso esemplare: la stabilità degli edifici nell’area intorno a Viale Giustiniano Imperatore (Roma, IX Municipio). In: *Memorie descrittive della Carta Geologica D’Italia. Volume L. La geologia di Roma dal centro alla periferia. Istituto Poligrafico dello Stato*

Caporossi P, Mazzanti P, Bozzano F (2018) Digital image correlation (DIC) analysis of the 3 December 2013 Montescaglioso landslide (Basilicata, southern Italy): results from a multi-dataset investigation. *ISPRS Int J Geo Inf* 7(9):372

Capelli G, Mastroiello L, Mazza R, Petitta M, Baldoni T, Banzato F et al (2012) Carta idrogeologica del Territorio della Regione Lazio, scala 1:100.000 (4 fogli)

Cascini L (2008) Applicability of landslide susceptibility and hazard zoning at different scales. *Eng Geol* 102(3–4):164–177

Ciotoli G, Corazza A, Finoia MG, Nisio S, Serafini R, Succiarelli C (2013) Sinkholes antropogenici nel territorio di Roma Capitale. I Sinkholes: metodologie di indagine, ricerca storica, sistemi di monitoraggio e tecniche di intervento. Centri abitati e processi di instabilità naturale: valutazione, controllo e mitigazione. *Mem Descr Carta Geol D’it* 93:143–181

Ciotoli G, Di Loreto E, Finoia MG, Liperi L, Meloni F, Nisio S, Sericola A (2016) Sinkhole susceptibility, Lazio region, central Italy. *J Maps* 12(2):287–294

Ciotoli G, Nisio S, Serafini R (2015) Analisi della suscettibilità ai sinkholes antropogenici nel centro urbano di Roma: analisi previsionale. *Mem Descr Carta Geol D’it* 99:167–188

Corominas J, van Westen C, Frattini P, Cascini L, Malet JP, Fotopoulou S, Smith JT (2014) Recommendations for the quantitative analysis of landslide risk. *Bull Eng Geol Environ* 73(2):209–263

Crosetto M, Monserrat O, Devanthery N, Cuevas-González M, Barra A, Crippa B (2016) Persistent scatterer interferometry using Sentinel-1 data. *Int Arch Photogramm Remote Sens Spat Inf Sci ISPRS Arch* 41:835–839. <https://doi.org/10.5194/isprsarchives-XLI-B7-835>

Delgado Blasco JM, Fomelis M, Stewart C, Hooper A (2019) Measuring urban subsidence in the Rome metropolitan area (Italy) with Sentinel-1 SNAP-StaMPS persistent scatterer interferometry. *Remote Sensing* 11(2):129

Del Monte M, D’Orefice M, Luberti GM, Marini R, Pica A, Vergari F (2016) Geomorphological classification of urban landscapes: the case study of Rome (Italy). *J Maps*. <https://doi.org/10.1080/17445647.2016.1187977>

Di Salvo C, Ciotoli G, Pennica F, Cavinato GP (2017) Pluvial flood hazard in the city of Rome (Italy). *J Maps* 13(2):545–553

Esposito C, Belcecchi N, Bozzano F, Brunetti A, Marmoni GM, Mazzanti P, Spizzirri M (2021) Integration of satellite-based A-DInSAR and geological modeling supporting the prevention from anthropogenic sinkholes: a case study in the urban area of Rome. *Geomat Nat Hazards Risk* 12(1):2835–2864

Esposito C, Marmoni GM, Scarascia Mugnozza G, Argentieri A, Rotella G (2019) A preliminary large-scale assessment of landslide susceptibility in the territory of the Metropolitan City of Rome (Italy) *Proceedings of 2019 IPL Symposium on Landslides, Paris, France 16 – 19 September 2019. ISBN 978-4-9903382-5-1*

- Fang Z, Wang Y, van Westen CJ, Lombardo L (2022) Space-time landslide susceptibility modelling in Taiwan
- Fell R, Corominas J, Bonnard C, Cascini L, Leroi E, Savage WZ (2008) on behalf of the JTC-1 Joint Technical Committee on Landslides and Engineered Slopes (2008) Guidelines for landslide susceptibility, hazard and risk zoning for land use planning. *Eng Geol* 102(3–4):85–98
- Ferretti A, Prati C, Rocca F (2000) Nonlinear subsidence rate estimation using permanent scatterers in differential SAR interferometry. *IEEE Trans Geosci Remote Sens* 38:2202–2212. <https://doi.org/10.1109/36.868878>
- Ferretti A, Prati C, Rocca F (2001) Permanent scatterers in SAR interferometry. *IEEE Trans Geosci Remote Sens* 39:8–20. <https://doi.org/10.1109/36.898661>
- Fowler HJ, Kilsby CG, O’Connell PE (2003) Modeling the impacts of climatic change and variability on the reliability, resilience, and vulnerability of a water resource system. *Water Resour Res* 39(8)
- Funciello R, Giordano G (2008) La nuova carta geologica di Roma: litostratigrafia e organizzazione stratigrafica. La geologia di Roma dal centro storico alla periferia. *Memorie Descrittive della Carta Geologica d’Italia* 80(1):39–85
- Geurts P, Ernst D, Wehenkel L (2006) Extremely randomised trees. *Mach Learn* 63:3–42. <https://doi.org/10.1007/s10994-006-6226-1>
- Glade T (2001) Landslide hazard assessment and historical landslide data—an inseparable couple?. In: *The use of historical data in natural hazard assessments* (pp 153–168). Springer, Dordrecht
- Hanssen RF (2005) Satellite radar interferometry for deformation monitoring: a priori assessment of feasibility and accuracy. *Int J Appl Earth Obs Geoinf* 6:253–260. <https://doi.org/10.1016/j.jag.2004.10.004>
- Hesse R (2010) LiDAR-derived Local Relief Models - a new tool for archaeological prospection. *Archaeol Prospect* 17:67–72
- Hosking JRM, Wallis JR, Wood EF (1985) Estimation of the generalized extreme value distribution by the method of probability weighted moments. *Technometrics* 27:251–261. <https://doi.org/10.1080/00401706.1985.10488049>
- Iadanza C, Trigila A, Starace P, Dragoni A, Biondo T, Roccisano M (2021) IdroGEO: a collaborative web mapping application based on REST API services and open data on landslides and floods in Italy. *ISPRS Int J Geo-Inf* 10:89. <https://doi.org/10.3390/ijgi10020089>
- IFFI (2007) Inventario dei fenomeni franosi in Italia realizzato dall’ISPRA e dalle Regioni e Provincie Autonome. <http://www.isprambiente.gov.it/it/progetti/iffi-inventario-dei-fenomeni-franosi-in-italia>
- ISPRA (2014) Progetto frane Roma. Inventario dei fenomeni franosi nel territorio di Roma Capitale. Italian National Institute for Environmental Protection and Research, (ISPRA). Retrieved September 17, 2015, from: <http://sgi.isprambiente.it/franeroma/default.htm>
- Jenness J (2006) Topographic position index (tpi\_jen. avx) extension for ArcView 3. x, v. 1.3 a. Jenness Enterprises
- Jenkinson AF (1955) The frequency distribution of the annual maximum (or minimum) values of meteorological events. *Q J Royal Meteorol Soc* 81:158–171. <https://doi.org/10.1002/qj.49708134804>
- Julian M, Anthony E (1996) Aspects of landslide activity in the Mercantour Massif and the French Riviera, southeastern France. *Geomorphology* 15(3–4):275–289
- Kampes BM (2006) Radar interferometry: persistent scatterer technique. Vol. 12 ISBN 140204576X.
- Kuhn M, Johnson K (2013) Applied predictive modeling (Vol. 26, p. 13). New York: Springer
- La Vigna F, Mazza R, Amanti M, Di Salvo C, Petitta M, Pizzino L (2015). The synthesis of decades of groundwater knowledge: the new hydrogeological map of Rome. *Acque Sotterranee-Ital J Groundwater* 4(4)
- Loche M, Alvioli M, Marchesini I, Bakka H, Lombardo L (2022) Landslide susceptibility maps of Italy: lesson learnt from dealing with multiple landslide types and the uneven spatial distribution of the national inventory. *Earth Sci Rev* 104125
- Marra F, Rosa C (1995) Stratigrafia e assetto geologico dell’area romana. *Memorie Descrittive della Carta Geologica d’Italia* 50:49–118
- Massonnet D (1998) Feigl, KL Radar interferometry and its application to changes in the earth’s surface. *Rev Geophys* 36:441–500. <https://doi.org/10.1029/97RG03139>
- Mastrantoni G, Caprari P, Esposito C, Marmoni GM, Mazzanti P, Bozzano F (2022) Data requirements and scientific efforts for reliable large-scale assessment of landslide hazard in urban areas (No. EGU22–4669). Copernicus Meetings
- Mastrantoni G, Marmoni GM, Esposito C, Bozzano F, Scarascia Mugnozza G, Mazzanti P (submitted) Reliability assessment of open-source multi-scale landslide susceptibility maps and effects of their fusion. *Georisk PAI* (2012) - Piano di Assetto Idrogeologico a cura dell’Autorità di bacino del Tevere
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Duchesnay E (2011) Scikit-learn: machine learning in Python. *J Mach Learn Res* 12:2825–2830
- PRG Roma (2007) Carta di pericolosità e vulnerabilità geologica del territorio comunale. Piano Regolatore Generale del Comune di Roma a cura di Mordigliani D
- 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
- Remedia G, Alessandrini MG, Mangianti F (1998) Le piene eccezionali del fiume Tevere a Roma Ripetta. Università degli Studi dell’Aquila, Dip. di Ingegneria delle Strutture, delle Acque e del Terreno (DISAT n. 3)
- Schilirò L, Poueme Djueyep G, Esposito C, Scarascia Mugnozza G (2019) The role of initial soil conditions in shallow landslide triggering: insights from physically based approaches. *Geofluids*
- Segoni S, Piciullo L, Gariano SL (2018) A review of the recent literature on rainfall thresholds for landslide occurrence. *Landslides* 15(8):1483–1501
- Serrano SE (2010) Hydrology for engineers, geologists, and environmental professionals: an integrated treatment of surface, subsurface, and contaminant hydrology, 2nd edn. Hydroscience Inc., USA, p 590
- Sottili G, Palladino DM, Zanon V (2004) Plinian activity during the early eruptive history of the Sabatini Volcanic District, Central Italy. *J Volcanol Geoth Res* 135(4):361–379
- Steger S, Brenning A, Bell R, Glade T (2016) The propagation of inventory-based positional errors into statistical landslide susceptibility models. *Nat Hazard* 16(12):2729–2745
- Steger S, Brenning A, Bell R, Glade T (2017) The influence of systematically incomplete shallow landslide inventories on statistical susceptibility models and suggestions for improvements. *Landslides* 14:1767–1781
- Thiery Y, Terrier M, Colas B, Fressard M, Maquaire O, Grandjean G, Gourdiere S (2020) Improvement of landslide hazard assessments for regulatory zoning in France: state-of-the-art perspectives and considerations. *International Journal of Disaster Risk Reduction* 47:101562
- 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
- Titti G, van Westen C, Borgatti L, Pasuto A, Lombardo L (2021) When enough is really enough? On the minimum number of landslides to build reliable susceptibility models. *Geosciences* 11(11):469
- Ventriglia U (1971) La geologia della città di Roma. *Amm Prov Roma*, p 417
- Ventriglia U (2002) Geologia del territorio del Comune di Roma. *Amm Prov Roma*, p 810
- Zieher T, Perzl F, Rössel M, Rutzinger M, Meißl G, Markart G, Geitner C (2016) A multi-annual landslide inventory for the assessment of shallow landslide susceptibility—two test cases in Vorarlberg, Austria. *Geomorphology* 259:40–54

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10346-023-02095-7>.

**C. Esposito · G. Mastrantoni · G. M. Marmoni (✉) · B. Antonielli · P. Caprari · A. Pica · P. Mazzanti · F. Bozzano**

Earth Sciences Department, Sapienza University of Rome, CERI Research Centre for Geological Risks. P. le Aldo Moro, 00185 Rome, Italy

Email: gianmarco.marmoni@uniroma1.it

**L. Schilirò**

Institute of Environmental Geology and Geoengineering (IGAG), National Research Council (CNR), Rome, Italy

**P. Mazzanti**

NHAZCA S.R.L. start-up, Sapienza University of Rome, 00185 Rome, Italy