Integrated Methods for Cultural Heritage Risk Assessment: Google Earth Engine, Spatial Analysis, Machine Learning

Maria Danese(✉), Dario Gioia, and Marilisa Biscione

ISPC-CNR, Tito Scalo, 85100 Potenza, Italy
maria.danese@cnr.it

Abstract. Cultural heritage risk assessment is an important task. Usually cultural heritage inside the consolidated city is more protected than cultural heritage spread over the territory. In this paper a method is proposed that integrates different technologies and platforms: from Google Earth Engine (GEE) and machine learning to desktop GIS and spatial analysis. The aim is to map the different vulnerability and hazard layers that constitute the cultural heritage risk map of the rural area in the Matera Municipality.

Keywords: Natural hazard · Urban sprawl · Google Earth Engine · Machine learning · Matera · Cultural heritage

1 Introduction

The integration of the new Information and Computer Technologies (ICT) is important not only for knowledge and analysis of cultural heritage, but also for the extraction of information on the basis of their protection. This is the case of cultural heritage risk maps, that are useful to assess future natural or human transformations, that can affect cultural heritage [1].

Analysis and mapping of natural hazards or anthropic impact, are crucial to investigate the vulnerability of a landscape and plan possible mitigation or protection actions of cultural heritage. Such analyses can be done in a more effective way thanks to the integration of spatial and remote sensing technologies, by using GIS and the new GEE platform launched in 2001. With these instruments also a protocol for cultural heritage risk assessment could be defined, based on the definition of risk [2], Eq. 1)

$$R = H \times V \times E \tag{1}$$

In this paper, we considered this definition in particular for the cultural heritage spread over the territory, because it is more exposed to natural and anthropic risks, while cultural heritage inside the urban areas is generally more known and protected.

In Eq. 1, risk R is expressed as the product of the following components:

© Springer Nature Switzerland AG 2021
O. Gervasi et al. (Eds.): ICCSA 2021, LNCS 12951, pp. 605–619, 2021.
https://doi.org/10.1007/978-3-030-86970-0_42

- H is hazard. For the Matera municipality, different type of hazards were considered: the urban growth as anthropic hazard, because it could interfere in a negative way in the presence of elements of archeological interest; landslides, floods, and erosion as natural hazards, for the impact that they have over the cultural heritage in general, but also, more in particular, over the spread archaeological findings.
- V is vulnerability. There are many ways to assess vulnerability maps for cultural heritage. A basic method is represented by the mapping of cultural heritage in a territory [3]; other methods use predictive models to consider also the archaeological potential of the studied area [4–6].
- E is exposure and indicates the value of the elements at risk. In this work, we consider all the cultural heritage elements as exposed to the same level of risk.

2 Datasets and Methods

2.1 Google Earth Engine©

GEE is a free, powerful platform that allows the analysis and monitoring of remote sensing big data at the planetary scale.

The first GEE version went out in 2001 [7], but it was publicly released in 2010. 2001 could be the internal release date of the platform. and today it is still used very much in many application fields [8] in general for the global –scale earth observation data extraction and analysis [9], but more in particular, just to cite some examples, for land use mapping [10] and monitoring and natural risk assessment ([11] and reference therein), for population mapping [12], climate data analysis [13] and large scale management of cultural heritage [14].

The use of GEE offers many advantages. It gives the possibility to work with large dimension data (satellite data but also vector GIS data), without downloading them on pc, even if the download is possible, also because many of these data are freely available. At the same time, the calculation time is very fast, thanks to the cloud elaboration, more than it would be possible on a personal computer. It also gives the possibility to upload the user's data, if needed. Moreover, the GEE platform has an efficient versioning system and offers a flexible access to its code through APIs. Finally, even if the code developed by the user is online, the ownership remains to its developer, even if it could decide to share it.

2.2 Machine Learning: Random Forest Method

Google Earth Engine offers many classification methods inside its platform. Between these, we choose an ensemble method taken from machine learning: the random forest classifier, a method introduced by [15, 16].

Ensemble methods 1) run the same algorithm many times or 2) put different machine learning algorithms, in order to execute a more effective analysis. The random forest classifier belongs to the first category because it uses recursively the decision trees for classification, according to the following simplified steps:

1. A training set is chosen inside the population dataset;
2. The decision tree classification is executed;
3. Then the process is repeated in order to improve the learning process and to return the best classification.

In satellite data analysis random forest recently started to be very used in many applications (see for example [17, 18] and reference therein).

In this paper, the random forest method was used to extract from Landsat data the land uses changes over 30 years and in particular to extract the urban growth [19] of the Matera case study.

2.3 Spatial Analysis

In this paper, map algebra is used as spatial analysis method useful to calculate some of the hazard parameters and then to combine hazard and vulnerability to obtain the risk map.

The classic definition of Map Algebra [20] expresses it as a "high level spatial modeling language, including base elements (operators), complex elements (functions) and formal language (instructions), together with elements needed to program and develop complex models".

Usually map algebra functions are classified in local, focal and zonal [21]. In local functions cell-by-cell operations are done; in focal functions operations are executed inside a defined neighborhood of the input raster; finally in zonal functions, operations are done inside another zonal raster.

3 The Matera Case Study

3.1 The Matera Landscape

The study area includes the administrative territory of Matera, which, from a geological viewpoint, is located between the westernmost sectors of the foredeep of the southern Apennines belt and the carbonate ridges of the Apulia foreland [22, 23]. The landscape is carved in a thick succession of Cretaceous limestones, which exhibit typical features of a karst environment [22, 24, 25]. Starting from the middle-late Pliocene, the subsidence of the western slope of the Apulia foreland promoted the deposition of coarse-grained mixed lithoclastic-bioclastic deposits (i.e. which unconformably overlie the Cretaceous limestone. Calcarenite di Gravina Fm.) passes upward to marine silty clay of the Argille subappennine Fm., Early Pleistocene in age [23]. Starting from the-Middle Pleistocene, this sector is affected by a regional uplift [22, 26], which promoted the progressive emersion of the area and the formation of deep gorges (the so-called Gravine), formed in the Cretaceous bedrock [22].

3.2 The Matera Municipality and its Cultural Heritage

The case study of this research is the territory of Matera, a centuries-old town lo-cated in the Basilicata Region, in Southern Italy. The ancient core of the town, worldwide known as the Sassi and the nearby "Murgia Materana" Park were classified as a UNESCO World Heritage Site since 1993 (http://whc.unesco.org/en/list/670, accessed 12 April 2021). The protected area comprises about 80 small rock-hewn villages that are still almost intact due to the lack of urbanization in the whole area, albeit with significant conservation problems. The park is important for the cultural heritage it mainly preserves architectonic and archaeological heritage in the rupestrian landscape, with many rock-cut settlements (rock churches, farms, dry stone walls, ancient calc-tuff quarry).

We chose the municipality of Matera for three reasons:

1) The Sassi and the Murgia Materana Park represent an extraordinary kind of settlement which have been continuously inhabited by man for two millennia. The cultural heritage is huge and partially not yet discovered and preserved;
2) The UNESCO and 2019 European Capital of Culture acknowledgments on the one hand require an effort for conservation and risk assessment, and on the other, they give a greater tourist, economic and urban impetus to the modern town;
3) The particular history of the territorial planning between the 50s and 70s cer-tainly influenced the growth of the modern city (for example the laws for the re-loca-tion of inhabitants in new residential and rural districts designed by sociologists and urban planners - laws 619/1952, 299/1958, 126/1967, 1043/1971).

4 Vulnerability Analysis

In addition to studies on the loss of cultural heritage [27] and on the appropriate actions to prevent its loss [28], it is important to evaluate previously the effective vulnerability of the cultural heritage at a territorial scale in order to safeguard it and provide precise guidelines for territorial planning and landscape protection [29, 30].

Vulnerability is the aptitude of an artefact to suffer damage, and it is generally associated to the state of conservation. It is usually evaluated with a multidisciplinary study: specific parameters are applied to the artefact characteristics, such as the appearance of the surface, the constructive and structural data, the use and the safety.

In this study, the assessment of vulnerability is on a territorial scale, and it can be classified as a first level vulnerability, based only on the presence of the cultural heritage over the territory. In particular, the consolidated built area was not considered, while all the cultural heritage in the periurban and rural areas were considered to assess vulnerable areas [Fig. 1] in a buffer of 150 m around each element.

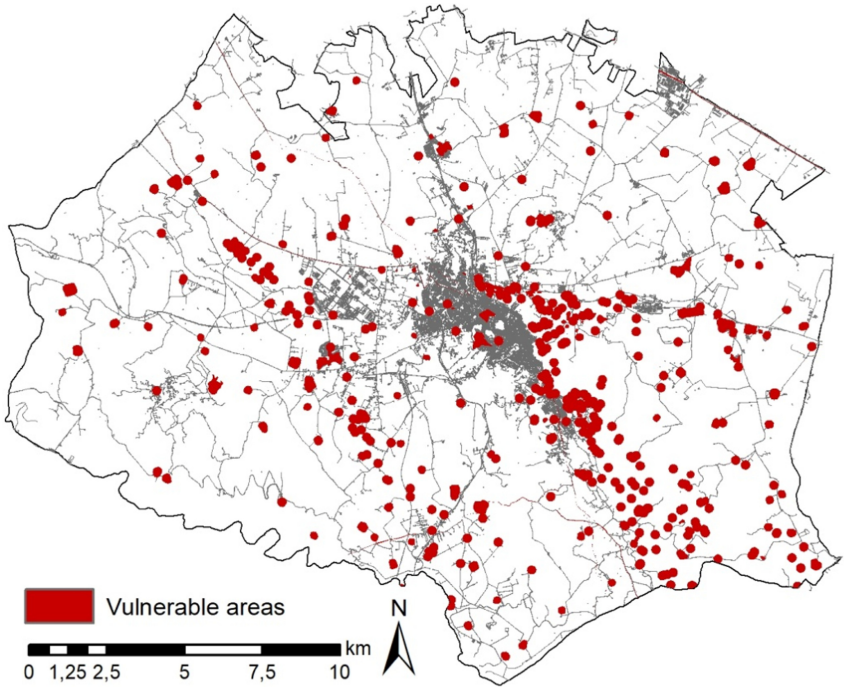

Fig. 1. Vulnerability map based on the presence of the cultural heritage in the extra-urban territory.

5 Hazard Analysis

In this work, we evaluate the following natural and anthropic hazards for the Matera municipality: for what concerns anthropic hazard, the urban growth was considered; whereas natural hazard are represented by soil erosion, landslides, and floods. In the next paragraphs, we explained how each of these hazard layers were evaluated in order to assess a cultural heritage risk map.

5.1 Urban Growth

Land use changes and in particular the urban growth could affect both cultural heritage and archaeological findings distribution over the territory. In the last years, remote sensing and the increasing calculation capability of computers helped very much to monitor this phenomenon, however some limitations remain.

Among them, it is worth to note:

- the monitoring of land use changes is not an easy task, because a robust estimation of the parameter needs landcover maps always updated and with high resolution;
- remote sensing data elaboration, for large dataset, is time consuming and requires large hardware resources.

The GEE platform in this sense is very useful because it allows to use freely high-resolution images and allows to reduce computing time.

In this paper, we analyzed the extension and the variation of the urban area in Matera Municipality by using Landsat in the last three decades (1990–2020).

Landsat is a satellite constellation for the Earth remote observation. There were 8 Landsat launches, from '70s to 2013. It is used for studying many natural and anthropic changes of the Earth surface, but more in particular for land-use changes.

In particular, for the analysis of the urban growth in Matera Municipality Landsat 5 and Landsat 8 (30 m resolution) were: 1) filtered in order to clear pixel affected by the presence of clouds 2) analyzed with the random forest method. Both the algorithm are included in GEE. However, for what concerns the random forest method, it needs a training set, to be performed. Thus, the following four types of areas were chosen and partially vectorized in GEE, in order to be used before as training set and further as test set: 1) urban, 2) water, 3) high density vegetation (indicated as tree) and 4) low-density vegetation, from cultivated areas to uncultivated (indicated as vegetation). In this way, two maps of the land use were obtained. The first is for the period between 1988 and 1991 (Fig. 2a), where we let GEE produce a mediated image in order to obtain the best product in that period; the second map is about the land use in 2020 (Fig. 2b). For each map, an accuracy index that uses the digitized test set and a Kappa statistic was calculated [31]. The Landsat 5 accuracy is equal to 0.92, the Landsat 8 accuracy is equal to 0.89.

As it is possible to observe in Table 1, urban areas grew 24 km^2. Finally, the urban areas were extracted and compared in order to obtain areas where the city is growing (Fig. 3).

5.2 Landslides and Floods

Maps of the hydrogeological (i.e. landslides and floods) hazards were derived through a revision of literature data and available official cartography (see for example [32]). More specifically, the basic information about the spatial distribution of landslide phenomena and flood areas has been extracted from the data stored in the hydrogeological plan of the Basin Authority of Basilicata with local revision based on photo-interpretation. Such data are summarized in Fig. 4.

5.3 Soil Erosion Map

Spatial distribution of erosion and deposition was delineated through the application of the Unit Stream Power Erosion Deposition (USPED) model [33]. The model estimates the average soil loss (A, annual average soil loss [Mg $ha^{-1}yr^{-1}$]) using the structure of the well-known RUSLE empirical equation [34]:

$$A = R \cdot K \cdot LS \cdot C \cdot P \tag{2}$$

where R [MJ mm $h^{-1}ha^{-1}yr^{-1}$] is the rainfall intensity factor, which has been estimated using the following equation:

$$EI_{30} = 0.1087 \cdot (P_{24})^{1.86} \tag{3}$$

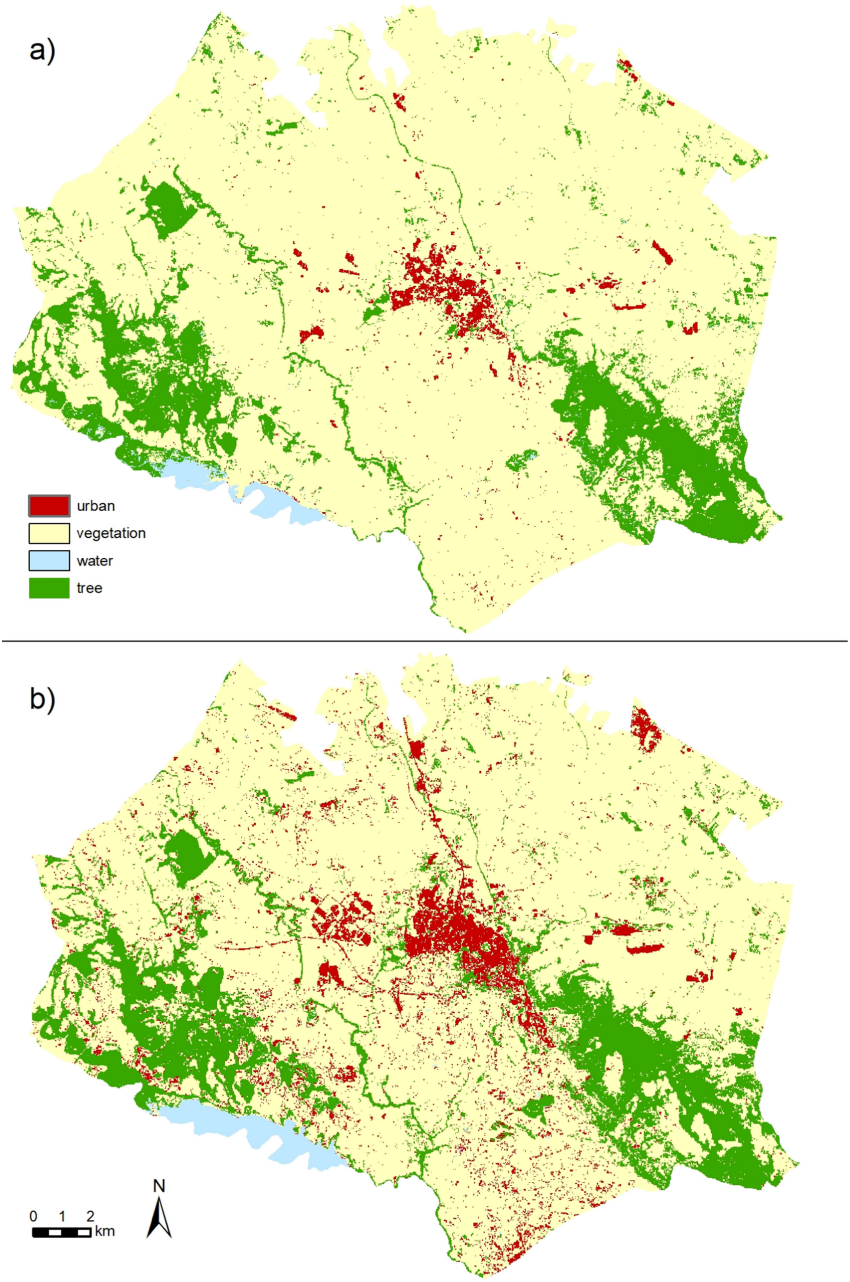

Fig. 2. Random forest classification of the land cover extracted from Landsat in the years a) 1988–1991, b) 2020

Table 1. Land use areas.

Land use	Areas of classes (km^2) extracted from Landsat 5 (1988–1991)	Areas of classes (km^2) extracted from Landsat 8 (2020)
Urban	9.6	34.0
Tree	425.4	385.7
Water	6.0	7.8
Vegetation	71.7	85.3

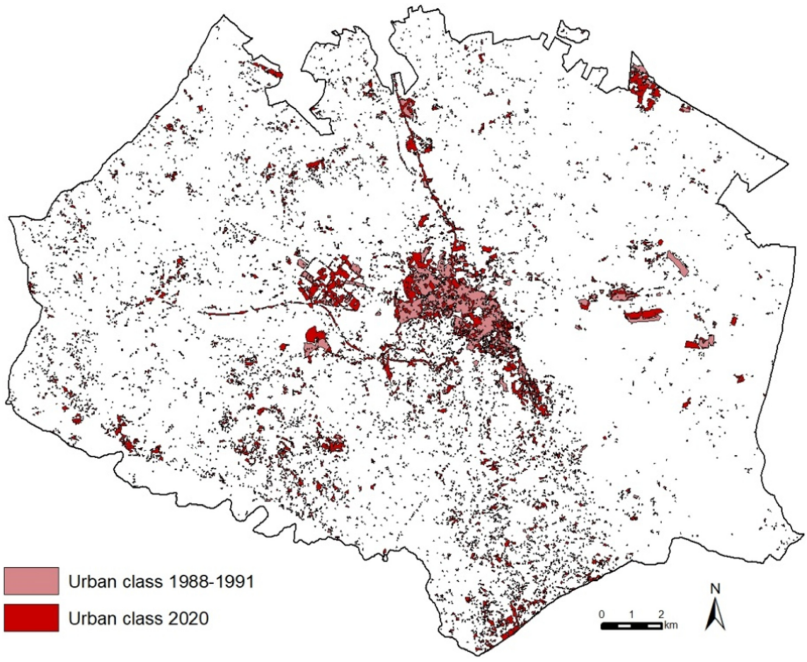

Fig. 3. The map of the urban growth extracted with GEE and random forest classification for the Matera municipality

with, EI30 is the rainfall erosivity in MJ mm h^{-1} ha^{-1} yr^{-1} and P_{24} the daily rainfall amount in mm [35]. Mean annual R-factor value has been estimated for the period from January 2010 to December 2020.

K [Mg h MJ^{-1} mm^{-1}] is the soil erodibility factor, which was estimated using the following equation that includes soil texture and permeability:

$$K = [2.1 \cdot 10^{-4} \cdot (12 - M) \cdot [(Si + fS)(100 - C)] \cdot 1.14 + 3.25(A - 2) + 2.5(P - 3)]/100 \quad (4)$$

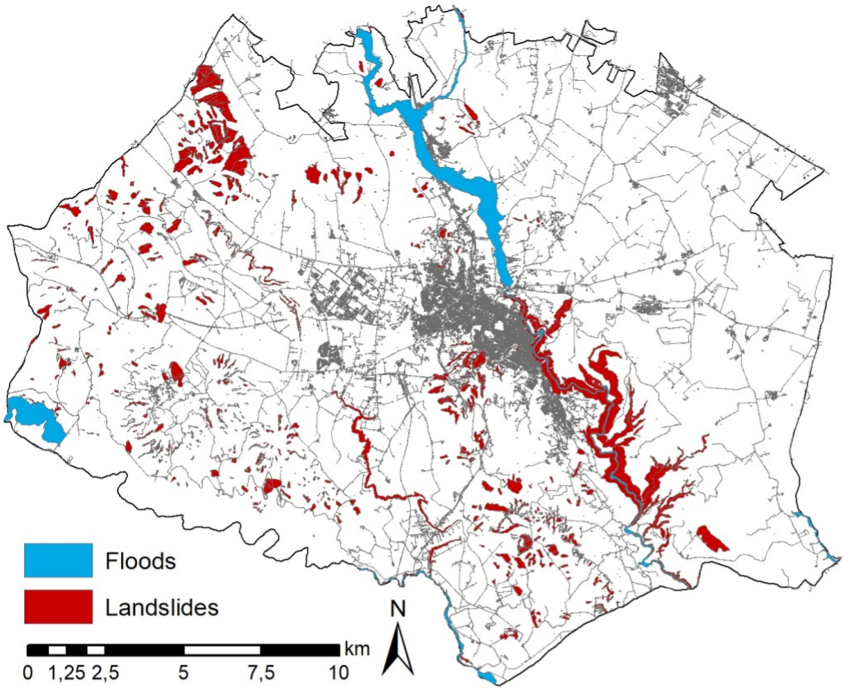

Fig. 4. Natural hazards: the flood and landslide map of the Matera Municipality

where M is the organic matter content (%), Si is the silt content (%), 2 to 50 μm, fS is the fine sand content (%) 50 μ to100 μm, C is the clay content (%) less than 2 μm, S is the sand content (%) 50 μm to 2 mm, A is the structure and P is the permeability class (within the top 0.60 m). K-factor map was drawn by combining lithological and land-use maps with values ranging from 0 to 0.1.

According to the procedures proposed by [33], the topographic factor LS is computed as the divergence of sediment flow (change in sediment transport capacity) using a 5 m DEM.

C [dimensionless] is the land cover factor and was derived starting from a land use map of the study area, provided by the Basilicata Regional Authority (http://rsdi.regione.basilicata.it). Using the basic structure of the Corine Land Cover project [36], the C-factor [dimensionless] has been calculated according to literature data [37].

The results of the application of the USPED model are shown in Fig. 5. Each pixel has been classified using the following three classes of erosion/deposition:

1 - Moderate to extreme erosion (>-5 Mg ha^{-1} yr^{-1}).
2 - Stable ($-5/5$ Mg ha^{-1} yr^{-1}).
3 - Moderate to extreme deposition (>5 Mg ha^{-1} yr^{-1}).

A simple statistical analysis of the frequency distribution of erosion and deposition areas allowed us to evaluate the following data: erosion areas occupy 6.9 km^2, stable areas are the majority (390 km^2) and deposition areas are equal to 7.7 km^2.

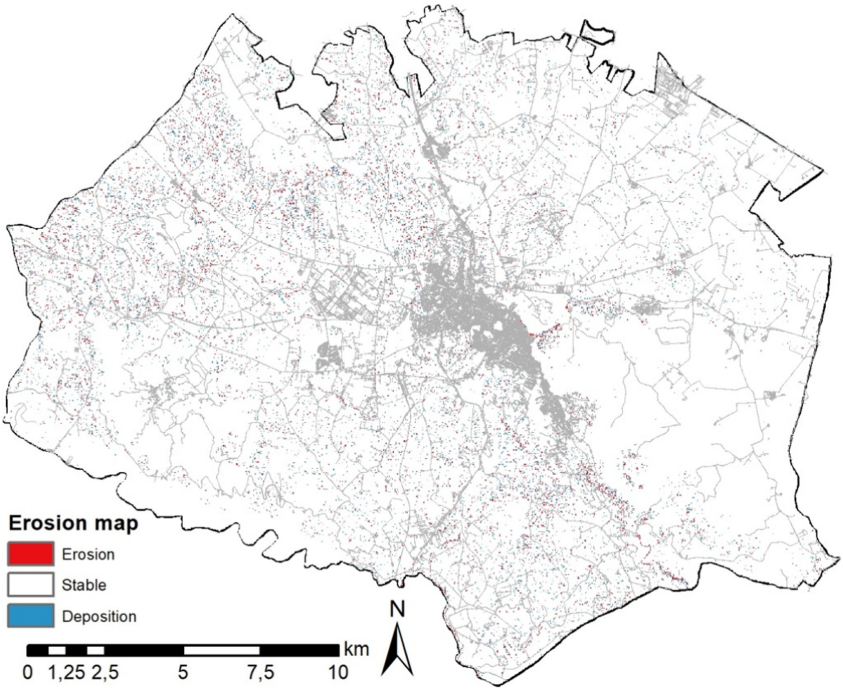

Fig. 5. Natural hazards: the erosion map of the Matera municipality

6 Results and Final Discussion

The different layers obtained in the previous paragraphs were combined with map algebra techniques, in particular with cell-by-cell functions in order to obtain: 1) the hazard map (Fig. 6); 2) the final risk map as a result of the overlay between vulnerability and hazards (Fig. 7).

Quantitative results show that hazard areas cover a total surface of 63.37 km^2 where about 30.7 km^2 are occupied by floods and landslides, 6.9 km^2 by erosion and 25.7 by the urban growth. Moreover, a relevant surface amount is affected by the overlay of two or three hazard factors (Table 2).

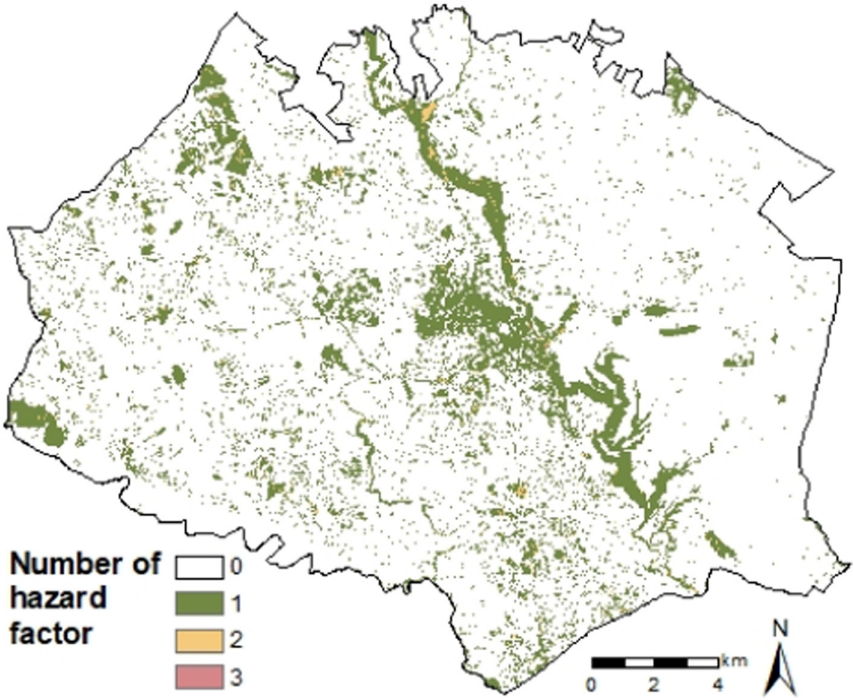

Fig. 6. Hazard factors

Table 2. Areas and percent over the study are occupied by 0, 1, 2 or 3 hazard factors and vulnerable areas

	Area (km^2)	%
Number of hazard factors		
0	330.1	85.3
1	54.6	14.1
2	2.3	0.6
3	0.04	0.01
Vulnerability areas	10.1	2.6

The risk map represents the synoptic view of the overlay of vulnerability and hazard (Main map Fig. 7, Table 3), shows that very few areas are really at risk, even if there are many areas characterized by a complex risk, where different types of hazard coexist (inset maps, Fig. 7). Of course, areas characterized by a more complex and heterogeneous risk should have a priority in the planning strategies of protection. Our results also highlight that natural hazards are mainly located in sectors featured by high-relief, steep slopes,

and along the thalweg of the main gorges, which are not preferentially occupied by archaeological and historical elements.

However, even, fortunately, the cultural heritage of the Matera open territory is quite safe from natural and anthropic risk, the paper showed a methodology that could become an automatic protocol to investigate, extract and assess the risk of any other case study.

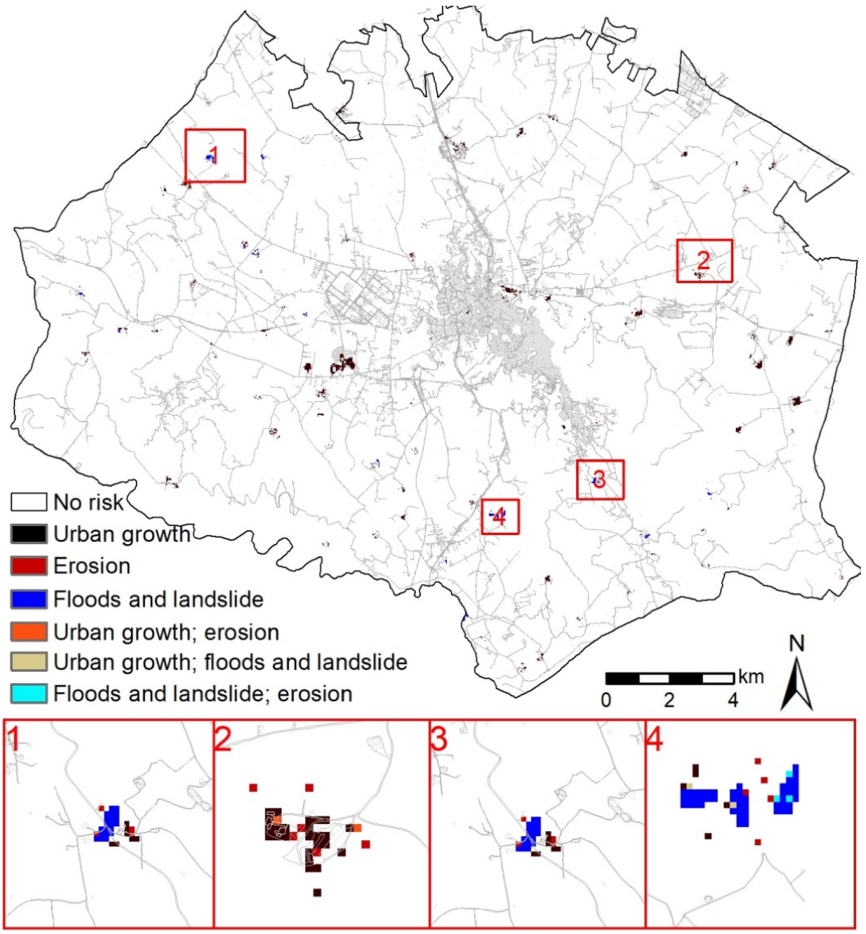

Fig. 7. The final risk map. Inset maps show some areas where different types of risks coexist, in overlay and in proximity.

However, to improve the performance of the model, it would be appropriate to go deepen also in exposure and the state of conservation of cultural heritage, according to the different types of heritage and its intrinsic features.

Table 3. Risk areas and percentage found

	Area (m^2)	% of the vulnerability areas
Type of risk		
Urban growth	1000503	10
Erosion	124356.3	1
Floods and landslides	185827.9	2
Urban growth & erosion	8478.84	0.1
Urban growth & floods and landslides	4945.99	0.1
Floods and landslide & erosion	10598.55	0.1

Acknowledgement. Authors want to thank Guido Ceccarini of the European Commission Joint Research Centre, who freely shared the script which GEE elaboration for urban growth analysis are inspired.

References

1. Serra, M., D'Agostino, S.: Archeologia Preventiva. Agenzia Magna Grecia, Albanella (SA) (2010)
2. Cruden, D.M., Varnes, D.J.: Landslide types and processes. In: Turner, A.K., Shuster, R.L. (eds.) Landslides: Investigation and Mitigation, pp. 36–75. National Academies Press, Washington DC (1996)
3. Biscione, M., Danese, M., Masini, N.: A framework for cultural heritage management and research: the cancellara case study. J. Maps **14**, 576–582 (2018)
4. Danese, M., Gioia, D., Biscione, M., Masini, N.: Spatial methods for archaeological flood risk: the case study of the neolithic sites in the Apulia Region (Southern Italy). In: Murgante, B., et al. (eds.) ICCSA 2014. LNCS, vol. 8579, pp. 423–439. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-09144-0_29
5. Danese, M., Masini, N., Biscione, M., Lasaponara, R.: Predictive modeling for preventive archaeology: overview and case study. Cent. Eur. J. Geosci. **6**(1), 42–55 (2014). https://doi.org/10.2478/s13533-012-0160-5
6. Yan, L., et al.: Towards an operative predictive model for the songshan area during the yangshao period. ISPRS Int. J. Geo-Inf. **10**, 217 (2021)
7. https://blog.google/outreach-initiatives/sustainability/introducing-google-earth-engine/
8. Tamiminia, H., Salehi, B., Mahdianpari, M., Quackenbush, L., Adeli, S., Brisco, B.: Google Earth Engine for geo-big data applications: a meta-analysis and systematic review. ISPRS J. Photogramm. Remote. Sens. **164**, 152–170 (2020)
9. Moore, R.T., Hansen, M.C.: Google Earth Engine: a new cloud-computing platform for global-scale earth observation data and analysis. In: AGU Fall Meeting Abstracts (2011)
10. Ravanelli, R., et al.: Monitoring the impact of land cover change on surface urban heat island through google earth engine: proposal of a global methodology, first applications and problems. Remote Sens. **10**, 1488 (2018)
11. Mutanga, O., Kumar, L.: Google Earth Engine applications. Remote Sens. **11** (2019)
12. Patel, N.N., et al.: Multitemporal settlement and population mapping from Landsat using Google Earth Engine. Int. J. Appl. Earth Obs. Geoinf. **35**, 199–208 (2015)

13. Huntington, J.L., et al.: Climate engine: cloud computing and visualization of climate and remote sensing data for advanced natural resource monitoring and process understanding. Bull. Am. Meteor. Soc. **98**, 2397–2410 (2017)
14. Agapiou, A.: Remote sensing heritage in a petabyte-scale: satellite data and heritage Earth Engine© applications. Int. J. Digit. Earth **10**, 85–102 (2017)
15. Kleinberg, E.M.: On the algorithmic implementation of stochastic discrimination. IEEE Trans. Pattern Anal. Mach. Intell. **22**, 473–490 (2000)
16. Kleinberg, E.M.: An overtraining-resistant stochastic modeling method for pattern recognition. Ann. Stat. **24**, 2319–2349 (1996)
17. Belgiu, M., Drăguţ, L.: Random forest in remote sensing: a review of applications and future directions. ISPRS J. Photogramm. Remote. Sens. **114**, 24–31 (2016)
18. Sheykhmousa, M., Mahdianpari, M., Ghanbari, H., Mohammadimanesh, F., Ghamisi, P., Homayouni, S.: Support vector machine versus random forest for remote sensing image classification: a meta-analysis and systematic review. IEEE J. Selected Topics Appl. Earth Obs. Remote Sens. **13**, 6308–6325 (2020)
19. Cao, G.Y., Chen, G., Pang, L.H., Zheng, X.Y., Nilsson, S.: Urban growth in China: past, prospect, and its impacts. Popul. Environ. **33**, 137–160 (2012)
20. Demers, M.: GIS modeling in raster (2001)
21. Tomlin, C.D.: GIS and Cartographic Modeling. Esri Press, Redlands, California (2013)
22. Beneduce, P., Festa, V., Francioso, R., Schiattarella, M., Tropeano, M.: Conflicting drainage patterns in the Matera Horst Area, Southern Italy. Phys. Chem. Earth **29**, 717–724 (2004)
23. Tropeano, M., Sabato, L., Pieri, P.: Filling and cannibalization of a foredeep: the Bradanic Trough Southern Italy. Geol. Soc. Spec. Publ. **191**, 55–79 (2002)
24. Gioia, D., Sabato, L., Spalluto, L., Tropeano, M.: Fluvial landforms in relation to the geological setting in the "Murge Basse" karst of Apulia (Bari Metropolitan Area, Southern Italy). J. Maps **7**, 148–155 (2011)
25. Teofilo, G., Gioia, D., Spalluto, L.: Integrated geomorphological and geospatial analysis for mapping fluvial landforms in Murge basse karst of Apulia (Southern Italy). Geosciences (Switzerland) **9**, 418 (2019)
26. Gioia, D., Schiattarella, M., Giano, S.: Right-angle pattern of minor fluvial networks from the ionian terraced belt, Southern Italy: passive structural control or foreland bending? Geosciences **8**, 331 (2018)
27. Pérez-Hernández, E., Peña-Alonso, C., Hernández-Calvento, L.: Assessing lost cultural heritage. a case study of the eastern coast of Las Palmas de Gran Canaria city (Spain). L. Use Policy **96**, 104697 (2020)
28. de Noronha Vaz, E., Cabral, P., Caetano, M., Nijkamp, P.: Urban heritage endangerment at the interface of future cities and past heritage: a spatial vulnerability assessment. Serie Research Memoranda 0036, VU University Amsterdam, Faculty of Economics, Business Administration and Econometrics (2011)
29. Fry, G.L.A., Skar, B., Jerpåsen, G., Bakkestuen, V., Erikstad, L.: Locating archaeological sites in the landscape: a hierarchical approach based on landscape indicators. Landsc. Urban Plan. **67**, 97–107 (2004)
30. Agapiou, A., et al.: Impact of Urban sprawl to cultural heritage monuments: the case study of paphos area in Cyprus. J. Cult. Herit. **16**, 671–680 (2015)
31. Congalton, R.: Putting the Map Back in Map Accuracy Assessment, pp. 1–11 (2004)
32. Lazzari, M., Gioia, D., Anzidei, B.: Landslide inventory of the Basilicata region (Southern Italy). J. Maps **14**, 348–356 (2018)
33. Mitasova, H., Hofierka, J., Zlocha, M., Iverson, L.R.: Modelling topographic potential for erosion and deposition using GIS. Int. J. Geogr. Inf. Syst. **10**, 629–641 (1996)
34. Renard, K.G.: Predicting soil erosion by water: a guide to conservation planning with the Revised Universal Soil Loss Equation (RUSLE). United States Government Printing (1997)

35. Capolongo, D., Diodato, N., Mannaerts, C.M., Piccarreta, M., Strobl, R.O.: Analyzing temporal changes in climate erosivity using a simplified rainfall erosivity model in Basilicata (Southern Italy). J. Hydrol. **356**, 119–130 (2008)
36. Büttner, G.: CORINE land cover and land cover change products. In: Manakos, I., Braun, M. (eds.) Land Use and Land Cover Mapping in Europe. RSDIP, vol. 18, pp. 55–74. Springer, Dordrecht (2014). https://doi.org/10.1007/978-94-007-7969-3_5
37. Renard, K.G., Foster, G.R., Weesies, G.A., Porter, J.P.: RUSLE: revised universal soil loss equation. J. Soil Water Conserv. **46**, 30–33 (1991)