sciendo

Thematic paper: Earth Observation for Smart City and Smart Region

GEOSCAPE

Using open data to reveal factors of urban susceptibility to natural hazards and man-made hazards: case of Milan and Sofia

Alberto Vavassori[1] ✉ — Angelly De Jesús Pugliese Viloria[1] — Maria Antonia Brovelli[1]

[1]Dipartimento di Ingegneria Civile e Ambientale, Politecnico di Milano, Italy
✉ *alberto.vavassori@polimi.it*

Abstract

Multi-hazard mapping in urban areas is relevant for preventing and mitigating the impact of nature- and human-induced disasters while being a challenging task as different competencies have to be put together. Artificial intelligence models are being increasingly exploited for single-hazard susceptibility mapping, from which multi-hazard maps are ultimately derived. Despite the remarkable performance of these models, their application requires the identification of a list of conditioning factors as well as the collection of relevant data and historical inventories, which may be non-trivial tasks. The objective of this study is twofold. First, based on a review of recent publications, it identifies conditioning factors to be used as an input to machine and deep learning techniques for single-hazard susceptibility mapping. Second, it investigates open datasets describing those factors for two European cities, namely Milan (Italy) and Sofia (Bulgaria) by exploiting local authorities' databases. Identification of the conditioning factors was carried out through the review of recent publications aiming at hazard mapping with artificial intelligence models. Two indicators were conceived to define the relevance of each factor. A first research result consists of a relevance-sorted list of conditioning factors per hazard as well as a set of open and free access data describing several factors for Milan and Sofia. Based on data availability, a feasibility analysis was carried out to investigate the possibility to model hazard susceptibility for the two case studies as well as for the limit case of a city with no local data available. Results show major differences between Milan and Sofia while pointing out Copernicus services' datasets as a valuable resource for susceptibility mapping in case of limited local data availability. Achieved outcomes have to be intended as preliminary results, as further details shall be disclosed after the discussion with domain experts.

Keywords

Disasters,
Multi-hazard mapping,
Urban susceptibility,
Open data,
Machine and deep learning

Received:
30 May 2022
Received in revised form:
09 October 2022
Accepted:
15 October 2022

Highlights for public administration, management and planning:

- We identify the factors conditioning urban susceptibility to different nature- and man-induced hazards and a set of open data describing such factors for two case studies (Milan and Sofia). Data shall be used as input to artificial intelligence models for urban susceptibility mapping.
- Obtained results are key to a thorough assessment of urban susceptibility to single and multi-hazard scenarios. Indeed, they can contribute to identifying the actions needed to increase urban preparedness and resilience.
- Research outcomes are meant to provide local stakeholders and decision-makers with valuable tools to improve urban planning and development strategies with the purpose of mitigating hazards' negative effects on existing activities and assets.

Available online at content.sciendo.com GeoScape 16(2) — 2022: 93—107 doi: 10.2478/geosc-2022-0008

The paper was originally presented at the "GIS Ostrava 2022 Earth Observation for Smart City and Smart Region" conference held on-line in March, 2022 (https://gisak.vsb.cz/gisostrava/index.php). Selected presentations from the conference were significantly extended and are now published in this volume as thematic papers exploring various topics related to usage of Earth Observation in smart city and smart region applications.

1 Introduction

Nature- and human-induced hazards are extreme phenomena that may have severe impacts on both the natural and man-made environment. Overpopulation, climate change, and urban development in areas that are susceptible to this kind of hazards may result in disasters affecting the environment and communities (Alexander 1995; Adger 2006; Kelman et al. 2015; Skilodimou & Bathrellos 2021).
The annual reported number of natural disasters by type from 1970 to 2019 shows an increasing trend starting from the 70s, being floods the most frequent hazard (EM-DAT 2020). In 2020, there were 193 major flood events globally, accounting for 60% of the major disasters of that year, affecting a significant amount of people (Academy of Disaster Reduction & Emergency Management 2021). Secondary peril events – i.e., natural disasters that tend to happen fairly frequently and imply low to medium losses (e.g., floods, landslides, and wildfires) – have been increasing during the last five years. Indeed, 2021 was the first year in which two secondary peril events generated losses above USD 10 billion, namely, Uri winter storm and the flooding in Europe (Bevere & Remondi 2022).
The urban resilience concept refers to the capacity that individuals and communities have within a city to survive, adapt, and grow despite the challenges that they may experience (Resilient Cities Network 2022). Urban resilience is a response to three main trends: climate change, urbanization, and globalization. The three of them lead to risks such as the increase of extreme events, e.g., floods and landslides, and environmental challenges, e.g., the increase in deforestation and greenhouse gas emissions. Therefore, proper urban planning is of paramount importance to prevent the negative consequences of nature- and human-induced hazards as well as to mitigate the associated risk.
Most of the published studies focused on the analysis of single hazards (Raška et al. 2020), but urban areas are typically susceptible to numerous hazards that may occur simultaneously or consecutively (Skilodimou et al. 2019), leading to much worse consequences on activities, people, and assets. For this reason, the development of a state-of-the-art method for an effective multi-hazard assessment is crucial. This is especially true for the urban centers, where the amount of exposed and vulnerable elements, such as people, settlements, and infrastructures, is particularly significant. As multi-hazard maps are ultimately derived from a proper combination of single-hazard maps (Skilodimou et al. 2019; Nachappa et al. 2020), a thorough understanding of the factors driving the susceptibility of the single hazards in a certain urban area is key to an exhaustive multi-hazard assessment. To that end, this work investigates through an in-depth literature review the conditioning factors that play a role in the most typical hazards that may threaten the urban environment. The obtained list of conditioning factors is promising to support the production of single and multi-hazard susceptibility maps at the urban level using machine and deep learning techniques.
This work was carried out in the framework of the Harmonia project (HARMONIA 2022, https://harmonia-project.eu/), which aims at providing stakeholders and urban planners with a decision support system to improve urban resilience and climate change mitigation strategies. The four European pilot cities of the project are Milan (Italy), Sofia (Bulgaria), Ixelles (Belgium), and Piraeus (Greece). Specifically, this paper focuses on Milan and Sofia, creating cases of study to assess the conditioning factors that take a part in the hazards that affect each city and provide an associated list of datasets. Furthermore, a list of detailed conditioning factors is provided for every single hazard, namely ground subsidence, landslide, flood, earthquake, heat island, and air pollution.
The remainder of this paper is structured as follows. The methodology followed to list single hazards' conditioning factors and the corresponding results are presented in Section 2. Case studies are discussed in Section 3, where the most frequent hazards and the available datasets for each city are presented. Finally, the final discussion and conclusions are reported in Section 4.

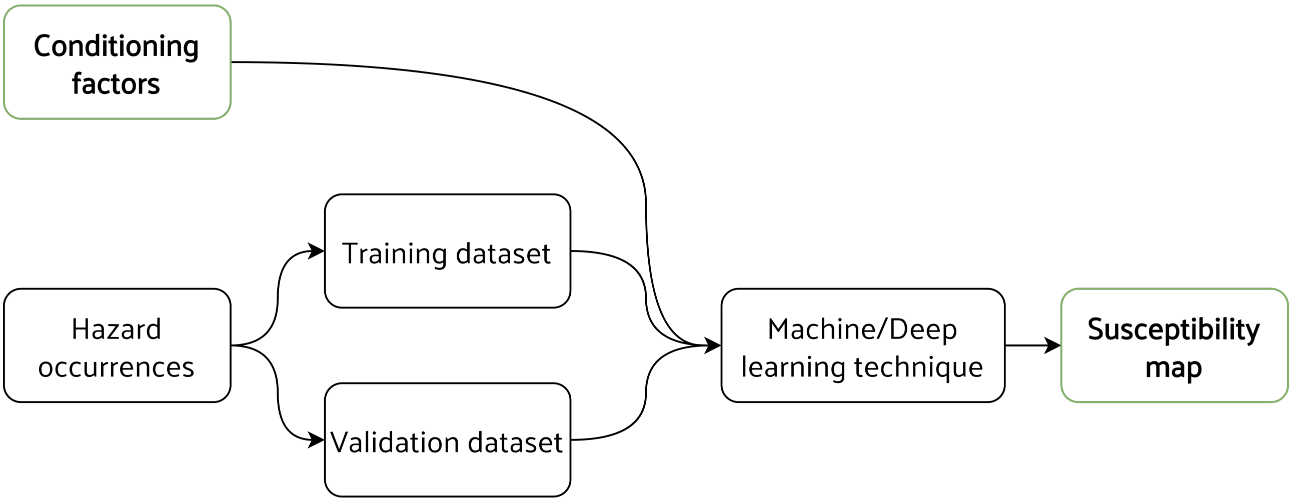

Fig. 1 Procedure adopted in the literature for the production of single-hazard susceptibility maps with machine and deep learning algorithms.

2 Definition of single hazards' conditioning factors

2.1 Methodology

This work focuses on the identification of the factors conditioning the susceptibility of an urban context to different nature- and man-induced hazards. The list of hazards considered in the paper encompasses geological and seismic hazards (ground subsidence, landslides, and earthquakes), hydrological hazards (floods), meteorological and climate hazards (heat islands), and man-made hazards (air pollution). Drought and extreme precipitation were not considered in this work because they cannot be modeled at the city level, despite the increase of drought events in the last 40 years (EM-DAT 2020) and the worldwide annual precipitation anomalies (United States Environmental Protection Agency (EPA) 2021). Nevertheless, extreme precipitation shall be considered as a conditioning factor for other possible hazards, such as floods. For a different reason, fires were not considered in this paper. Although the susceptibility to this hazard may be modeled at a local level, most of the reviewed papers deal with the susceptibility to wildfires. Urban fires are typically induced by different phenomena, such as industrial accidents and explosions. Despite other types of hazards might affect the urban environment (e.g., pandemics and industrial accidents), the investigation was restricted to those stated in the frame of the project Harmonia.

The conditioning factors playing a role in the occurrence of each hazard were identified through a review of recent scientific publications. The literature review was carried out by employing two popular research engines, namely Web of Science (https://www.webofscience.com) and Scopus (https://www.scopus.com/home.uri), according to specific criteria. First, the research was limited to the most recent publications (i.e., from 2018 to 2022) dealing with single and multi-hazard mapping with machine/deep learning models. Separate research was carried out for the different hazards, however, a set of common keywords was defined and used to optimize the review process, namely, "conditioning factors", "machine learning", "deep learning", "susceptibility", "hazard", and "mapping". The research keywords were then manually adjusted and refined for each hazard-specific review based on the research outcome. A more in-depth analysis of the publication's content was performed only for papers providing a relevant list of conditioning factors and exploiting the workflow described in the following.

A straightforward and common procedure to produce susceptibility maps with artificial intelligence algorithms was employed in most of the works found in the literature. The procedure can be described as follows (see Fig. 1). Firstly, a list of conditioning factors for single hazards is defined. Secondly, two types of data are collected, either from national/local inventories or from freely and publicly accessible databases. Specifically, data regarding conditioning factors and past hazard occurrences are retrieved. Past event occurrences are split into a training dataset and a validation dataset in the modeling process. Eventually, machine and deep learning algorithms are applied to produce a single-hazard susceptibility map.

Although in principle conditioning factors should be independent of the method used to compute susceptibility maps, only papers dealing with machine/deep learning techniques were investigated in order to detect the input variables leading to the best performance of artificial intelligence techniques.

The above-described procedure is quite standard (e.g., Choubin et al. 2020; Dang et al. 2020; Ebrahimy et al. 2020; Ahmad et al. 2021; Chen et al. 2022). However, the definition of a list of conditioning factors is not so straightforward as a variety of factors directly or indirectly play a role in the urban susceptibility to the hazard occurrence. Some factors are common to most of the literature works, but some relevant differences can be found in the different publications. For the sake of completeness, all the factors cited in the literature were taken into consideration and a degree of relevance was assigned to each of them.

For each hazard, the conditioning factors were methodically reported in a table containing the type of conditioning factor (e.g. hydrogeological, meteorological, topographical), the corresponding physical variable (e.g. groundwater level, maximum daily temperature, slope angle), its unit of measurement (e.g. meters, Celsius degrees), and the papers where a reference to such a factor was found.

A method for assigning the degree of relevance to each factor was conceived and applied. Specifically, the degree of relevance was assigned based on two indicators (Fig. 2). The first indicator (Indicator 1) corresponds to the number of publications mentioning each factor. A factor cited in a higher number of publications was considered more relevant. Conditioning factors having the same number of citations were then sorted based on a second indicator (Indicator 2), which takes into consideration the number of citations of each publication. Specifically, the second indicator is defined as the ratio between the number of citations of each paper and the number of years passed since it was published. Therefore, factors mentioned by highly-cited publications were labeled as most relevant.

For the sake of clarity, Fig. 2 depicts an example. Conditioning factors are sorted according to the first indicator (the first factor in the table has the highest value of Indicator 1). The second and the third factors are characterized by the same number of publications, thus they are sorted according to the second indicator. In particular, the second conditioning factor is cited in Paper 1, which has the highest value of Indicator 2. For this reason, it is considered more relevant than the third conditioning factor, which is not cited in the same publication. The two indicators were conceived to give an objective though preliminary evaluation of the importance of each factor, however, additional considerations regarding the specific local context and data availability should be made when selecting the factors to be employed for susceptibility mapping. Accordingly, the list of variables may vary depending on the specific case study. This aspect will be further emphasized in the following sections.

2.2 Results

A total number of 49 papers (48 dealing with single-hazards and 1 involving multiple-hazards) were initially retrieved and screened. Only 34 papers actually met all the research requirements (33 single-hazards and 1 multiple-hazards) and were therefore finally subject to in-depth analysis. Table 1 provides a summary of the number of reviewed papers dealing with each hazard.

The sorted lists of conditioning factors for each hazard are presented in the form of tables in the Supplementary Material. Conditioning factors were arranged according to the methodology described in the previous section, however, the actual value of the two indicators per factor and publication are not reported for space constraints. Nevertheless, essential information concerning each factor is provided. Specifically, tables provide a reference to the hazard definition as well as the references to the reviewed papers. To best understand which kind of variable affects the susceptibility to the different hazards, conditioning factors were grouped into categories (e.g., geological, hydrological, meteorological factors), so that similarities and differences among the various hazards could be disclosed.

Table 1 Number of papers initially screened and finally considered for each hazard

Hazard	No. of initial papers	No. of papers considered
Ground subsidence	9	9
Landslides*	7	7
Floods*	6	6
Earthquakes	4	4
Heat island	16	5
Air pollution	8	4

* The paper (Nachappa et al. 2020) was counted for both landslides and floods as it deals with multiple-hazard assessment

GeoScape 16(2) — 2022: 93—107 doi: 10.2478/geosc-2022-0008 Available online at content.sciendo.com

Conditioning factors			Indicator #1: Number of papers	Indicator #2: Number of citations over years since publication			
Type of factor (TF)	Physical Variable (PV)	Unit of measurement (U)		Paper #1	Paper #2	Paper #3	Paper #4
				8	6	5	3
TF #1	PV #1	U #1	4	•	•	•	•
TF #2	PV #2	U #2	3	•	•	•	
TF #3	PV #3	U #3	2		•	•	•
TF #4	PV #4	U #4	1				•
...	...	...	...	...	...	...	...

Fig. 2 Methodology applied to assign a degree of relevance to each conditioning factor.

Ground subsidence, landslides, and floods share a similar list of conditioning factors. Most of these factors are topographical variables that can be directly derived from a Digital Terrain Model (DTM) of the study area (e.g., slope, aspect, elevation). The susceptibility to these hazards is also conditioned by similar hydrological variables (e.g., rainfall) as well as the geological characteristics of the area (e.g. lithology). This type of data is provided by local or regional agencies and its availability is strictly dependent on the specific context.
Despite the similarities, some hazard-specific conditioning factors were found in the literature. In particular, the aquifer unit characteristics (e.g., permeability and sedimentary cover thickness) for ground subsidence, the slope and soil characteristics (e.g., slope length, convergence index, and soil type) for landslides, and the river catchment characteristics (e.g., flow accumulation and sediment transport index) for floods. Factors related to the slope and river catchment properties may be derived by leveraging the study region DTM. Data about the aquifer unit properties are obtained by in-situ surveys and distributed by local authorities.
Some of the above-mentioned factors, namely DTM-related variables (slope, elevation, aspect, and curvature) and geological variables (geology and proximity to faults), affect the susceptibility to earthquakes as well. However, other relevant conditioning factors for earthquakes are closely related to the region seismicity (e.g., magnitude, epicenter, and fault densities). Data about seismic variables are typically provided by national or research agencies, such as the INGV (National Institute of Geophysics and Volcanology) for Italy.
Heat island and air pollution are quite different from the other hazards, as their occurrence is not related to a specific event, such as a landslide or flooding, but to the exceedance of a certain physical variable threshold. Specifically, in the case of heat island, the two target physical variables are air temperature and air relative humidity, whereas the target variable of air pollution is the pollutants' concentration.
Despite these differences with respect to the hydrogeological hazards, a similar approach can be adapted to produce susceptibility maps of air heat islands and air pollution. Tables point out that these two phenomena are conditioned by similar meteorological variables, primarily related to temperature (e.g., maximum temperature), wind (e.g. wind speed), and topographical factors (e.g. elevation). However, some differences may be pointed out. Susceptibility to heat islands is conditioned by the urban morphology in terms of buildings and streets orientation and density (city canyons), and anthropogenic heat fluxes. On the other hand, air pollution is affected by other meteorological factors (e.g., evapotranspiration, soil moisture, dew point). Detailed data about urban morphology can be derived from the Topographic Database (TDB) which is generally provided by the single municipalities.
The analysis brought to light the relevance of land cover as a crucial factor in determining the susceptibility to most of the considered hazards. This type of information is provided by international agencies through dedicated services (such as the Copernicus Land Cover service) as well as local authorities.
As a final note, the reviewed papers dealing with susceptibility to landslides consider different types of movements, including rockfalls/rockslides, debris flows, shallow landslides, and complex movements (Nachappa et al. 2020; Emami et al. 2020; Ahmad et al. 2021), without differentiating variables more relevant to the different types of landslides. Accordingly, the list of physical variables reported in this work may be considered as a comprehensive result,

Available online at content.sciendo.com GeoScape 16(2) — 2022: 93—107 doi: 10.2478/geosc-2022-0008

however, more insightful evaluations should be performed depending on the specific context.

3 Case studies: open data for Milan and Sofia

Once defined the relevant conditioning factors to be leveraged for urban susceptibility mapping, the availability of open data was investigated, as the physical variables that will eventually be used as an input to machine and deep learning algorithms strictly depend on the case study and data characteristics. For this reason, the availability of data for two pilot cities was investigated. Specifically, Milan (Italy) and Sofia (Bulgaria) were considered as case studies in this work. The reason for this choice is twofold. On the one hand, both Milan and Sofia are pilot cities of the project Harmonia. Furthermore, a comparison between cities characterized by a significant difference in terms of data availability can be carried out.

For each case study, potential nature- and man-induced hazards were identified based on existing official documentation. Sources of open data that may be leveraged for the description of conditioning factors for each hazard were then pointed out. To best understand which conditioning factors, and, consequently, which hazards could be described and modeled through each dataset, relationships between data, factors, and hazards were represented through Sankey diagrams. Despite significant differences between the two case studies in terms of data availability, most local open datasets may be retrieved either from the municipality and regional Geoportals or from the local government plans. Useful global coverage datasets provided by the Copernicus Program Services were also investigated, as they may be adopted either as useful complementary information or in case of missing local data.

3.1 Milan case study

Among the hazards considered in this study, nature- and man-induced hazards that may affect the city of Milan and its metropolitan area are essentially ground subsidence, floods, earthquakes, heat island, and air pollution.

Specifically, some parts of the metropolitan city are affected by ground subsidence, primarily due to the significant groundwater withdrawal across the urban area (ISPRA 2020). Some portions of the major rivers' neighboring areas are characterized by a high flood probability, which makes flood a relevant hazard to be taken into consideration. On the contrary, being the metropolitan city entirely located in the Po plain, landslides are not a concerning hazard (ISPRA 2021). As for the seismic risk, despite the area being characterized by averagely low seismicity, some municipalities of the metropolitan city and the city of Milan itself are classified within zone 3, meaning a low seismic hazard with possible moderate ground shaking (Regione Lombardia 2014).

Urban heat island is another relevant phenomenon that affects the urban center, both in winter and summer, primarily during clear sky nights (ClimaMI 2019). Lastly, like most of the cities across the Po plain, Milan suffers from air pollution, which may become a particularly concerning hazard during wintertime. To provide an example, in 2021 the PM10 concentration threshold of 50 µg/m^3 was exceeded in 61 days in the urban area (ARPA Lombardia. 2022).

Given these pieces of information, an investigation of openly available datasets that could help describe conditioning factors for the above-cited hazards was carried out. Data is mainly provided by the Municipality of Milan Geoportal (https://geoportale.comune.milano.it/sit/open-data/), the Lombardy Region Geoportal (https://www.geoportale.regione.lombardia.it/download-ricerca), the Italian National Institute of Statistics (ISTAT) (http://dati.istat.it/), the INGV (https://istituto.ingv.it/it/risorse-e-servizi/archivi-e-banche-dati.html), and the Lombardy Region Environmental Protection Agency (ARPA) (https://www.arpalombardia.it/Pages/Ricerca-Dati-ed-Indicatori.aspx).

These agencies provide researchers, local stakeholders, and private citizens with a consistent amount of authoritative geospatial data with open licenses in easily readable formats.

Table 2 provides insights into the characteristics of the datasets, including source, format, reference system, resolution (for raster data) and scale (for vector data), year of last revision, and license. Datasets are primarily provided in standard vector and raster formats (shapefile, GeoPackage, grid, and GeoTIFF) or tabular formats (CSV and XLS), within geographic (WGS84 or ETRF00) or projected (WGS84 UTM32N) coordinate reference systems. Datasets are characterized by different scales and spatial resolutions, but all of them may be reasonably considered adequate for the analysis of hazards at the urban level. As for the temporal availability, most datasets were updated a few years ago. Datasets are updated by the provider depending on their temporal variability. Specifically, meteoro-

GeoScape 16(2) — 2022: 93—107 doi: 10.2478/geosc-2022-0008 Available online at content.sciendo.com

Table 2 Open datasets available for Milan

Dataset	Format	Reference system	Resolution / Scale	Year of last revision	License
Digital Terrain Model	Grid	WGS84 UTM32N	5 m	2015	CC-BY 4.0
Topographic database	GeoPackage	WGS84 UTM32N	1:2000	2021	CC-BY 4.0
Land use and cover	Shapefile	WGS84 UTM32N	1:10000	2019	CC-BY 4.0
Underground geological database	Shapefile	WGS84 UTM32N	1:10000	2022	CC BY-NC-ND 4.0
Geological map	Shapefile	WGS84 UTM32N	1:50000	2017	CC-BY-NC-SA~3.0 IT
Groundwater piezometric levels	Shapefile	WGS84 UTM32N	1:25000	2014	CC-BY-NC-ND 4.0
River network	Shapefile	WGS84 UTM32N	1:10000	2022	CC-BY 4.0
Roads, railways, and underground	Shapefile	WGS84 UTM32N	1:10000	2022	CC-BY 4.0
Population census	CSV, XLS	WGS84 UTM32N	Not specified	2011	CC-BY 3.0
Parametric earthquake catalog	XLS	WGS84	Not specified	2021	CC-BY-SA 4.0
Macroseismic database	XLS	WGS84	Not specified	2021	CC-BY-SA 4.0
Seismic hazard	XLS, TXT	WGS84	Not specified	2004	CC-BY 4.0
Land Surface Temperature	GeoTIFF	ETRF00	85 m	2018	License terms*
Meteorological data	CSV	WGS84	Not specified	2022	CC0 1.0 Universal

* Non-exclusive, fully-paid up, royalty free, worldwide, non sublicensable, non-transferable right to access and use the Materials for academic, non-profit, or other similar noncommercial purposes only.
Source: **Lombardy Region Geoportal**- **ISTAT**- **INGV**- **Milan Geoportal**-**Lombardy Region ARPA**

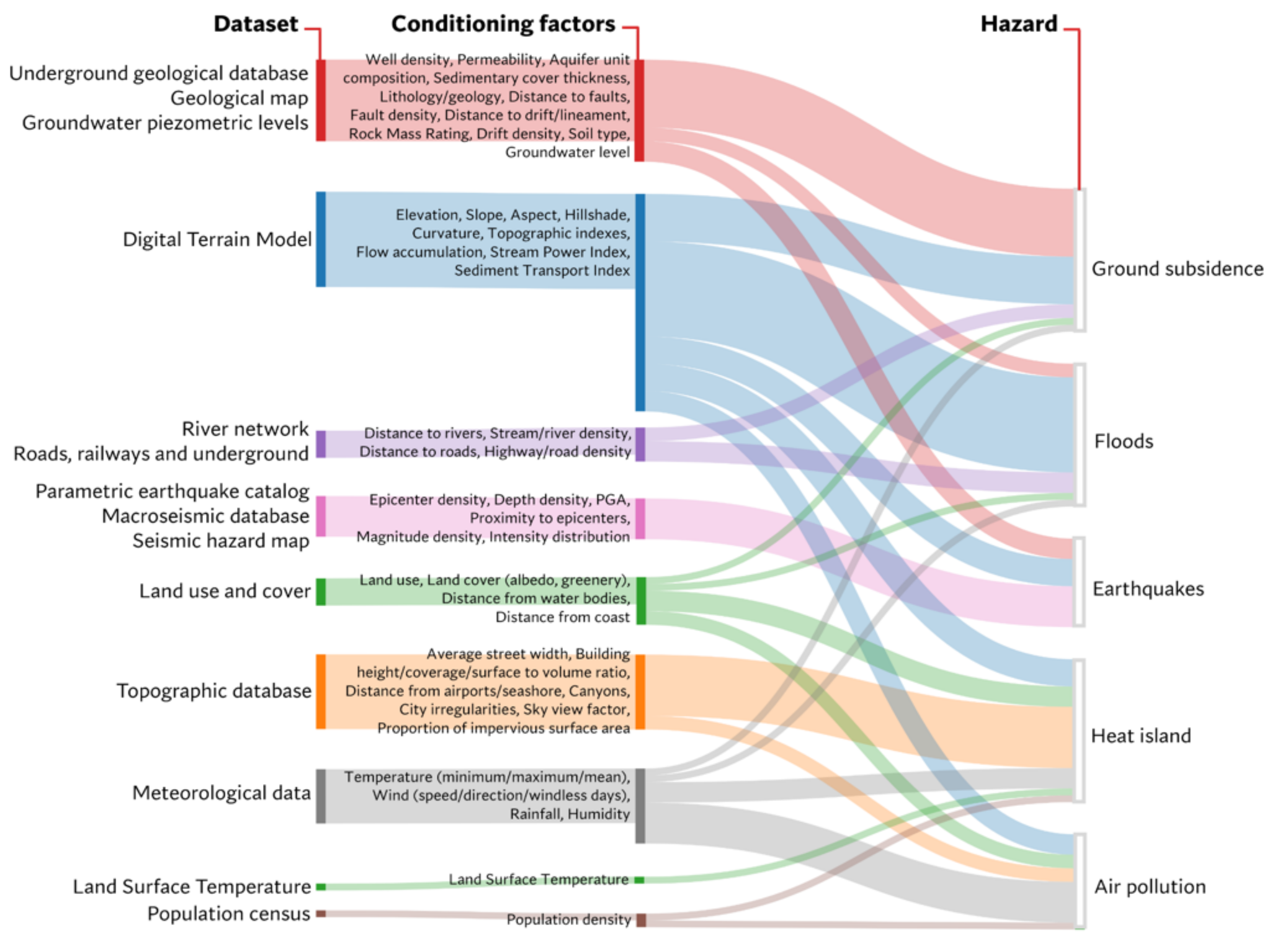

Fig. 3 Connections between datasets, conditioning factors, and hazards for Milan.

logical and seismic data are provided with a daily frequency, whereas quasi-static layers (e.g. geological map, river network, DTM) are occasionally updated.

The analysis of data pointed out that most conditioning factors may be described through open datasets. Fig. 3 represents through a Sankey diagram the relationships between the above-mentioned datasets, conditioning factors, and hazards. Specifically, the DTM and the TDB may be leveraged to derive a consistent number of conditioning factors that are pivotal for mapping most of the considered hazards. For the same reason, other relevant datasets include land cover, geological databases, and meteorological data.

Conditioning factors that cannot be described with the available open datasets include some meteorological variables that are relevant for air pollution (e.g., actual evapotranspiration, soil moisture, vapor pressure, and dew point). However, satellite imagery derived products made available by Copernicus may compensate for the lack of data.

3.2 Sofia case study

Sofia city is located in the Sofia Valley at the foot of the Vitosha mountain in the country of Bulgaria. The city was built in the west of the Iskar River and is surrounded by mineral springs. According to the literature review, the most concerning hazards and the ones that were considered in this paper for Sofia are floods, air pollution, earthquakes, and landslides.

Large floods and drought periods have increasingly taken place in the Upper Iskar Basin in the region of Sofia (Daniell 2011). PM10 concentrations in Bulgaria were one of the highest in 2009 and it continues to be the aforesaid nowadays (Dimitrova & Velizarova 2021). PM is particularly harmful during the winter period in big cities, such as Sofia, being domestic heating and transport emissions the main sources. The city is also exposed to a high seismic risk due to its location in the center of the Sofia seismic area (Paskaleva et al. 2004). Landslides are the most serious part of the geological hazards in Bulgaria, after earthquakes (Ivanov 2017), in fact, according to the susceptibility map proposed (Ivanov et al. 2020), the administrative region of Sofia has a moderate landslide susceptibility.

Some of the conditioning factors necessary to model the above-mentioned hazards can be retrieved from publicly available datasets. The main sources of data for the city of Sofia are Sofiaplan (https://sofiaplan.bg/) and Geographic Information System Sofia (GIS Sofia) (http://www.isofmap.bg/). Sofiaplan provides a catalog of datasets used for their own analysis or the result of their work which may be accessed via their API. GIS Sofia datasets, e.g., cadastral map, buildings, road network, may be accessed via WMS and WFS connections freely or paid.

The relevant open datasets available for this case study are described in Table 3, which indicates the format, reference system, resolution or scale, year of last revision, and license terms of the open datasets which correspond directly or indirectly to one or more conditioning factors. Sofiaplan is the only source of data utilized as it is the only platform in which data could be directly accessed and not only visualized in a WebGIS.

The relationship between the available datasets, the conditioning factors that can be derived from them, and the hazards of Sofia city is depicted in Fig. 4 as a Sankey diagram. The most important dataset is the DTM from Sofiaplan because of the list of conditioning factors that can be derived from it, e.g., slope, elevation, aspect, which are useful to model all the considered hazards. Daily rainfall and temperatures for the historical period of 1976 to 2005 datasets are considered to model floods, landslides, and air pollution but are constraint to the time range availability. Furthermore, some of the datasets included in Table 3 are not included in Fig. 4 as they refer to forecasted data under certain scenario which may be used after a hazard susceptibility assessment but not to model hazard occurrences.

Other relevant datasets include the landslide inventory and the areas with significant potential flood risk. The landslide inventory (Ministry of Regional Development & Public Works 2022) is a very useful dataset as it provides historical data of the landslide occurrences in Bulgaria, alongside the relevant details of each event. This information is of key importance because the occurrences of the hazardous events have to be correlated with the conditioning factors. On the other hand, the areas with significant potential flood risk (Sofiaplan 2018) dataset directly assesses the flooding susceptibility.

The lack of data necessary to model the concerning hazards in the city of Sofia can be observed in Fig. 4, and also, when compared with Fig. 3, the strong difference of data availability with respect to the city of Milan can be remarked. There is not enough data to cover most of the conditioning factors, in fact, most of the meteorological, land cover, geological, and seismic factors are missing. Only the topographical factors are mostly available. For this reason, global or continental datasets, e.g.,

sciendo

GeoScape 16(2) — 2022: 93—107 doi: 10.2478/geosc-2022-0008 Available online at content.sciendo.com

Table 3 Open datasets available for Sofia

Dataset	Format	Reference system	Resolution / Scale	Year of last revision	License
Digital Terrain Model	GeoTIFF	WGS84 UTM34N	5 m	2017	Sofiaplan license terms*
River network	GeoJson	WGS84	Not specified	2019	Sofiaplan license terms*
Daily rainfall for the historical period 1976 - 2005	GeoJson	WGS84	Not specified	2020	Sofiaplan license terms*
Maximum daily rainfall according to the RCP 4.5 (moderate) scenario for the future period 2021 - 2050	GeoJson	WGS84	Not specified	2020	Sofiaplan license terms*
Daily precipitation amounts according to the RCP 4.5 (moderate) scenario for the future period 2021 - 2050	GeoJson	WGS84	Not specified	2020	Sofiaplan license terms*
Temperatures for the historical period 1976 - 2005	GeoJson	WGS84	Not specified	2020	Sofiaplan license terms*
Temperatures under scenario RCP 4.5 (moderate) for the future period 2021 - 2050	GeoJson	WGS84	Not specified	2020	Sofiaplan license terms*

*It must be mentioned in a clear and visible way that these materials are a product of Sofiaplan Municipal Enterprise (Sofiaplan 2022)

Source: Sofiaplan

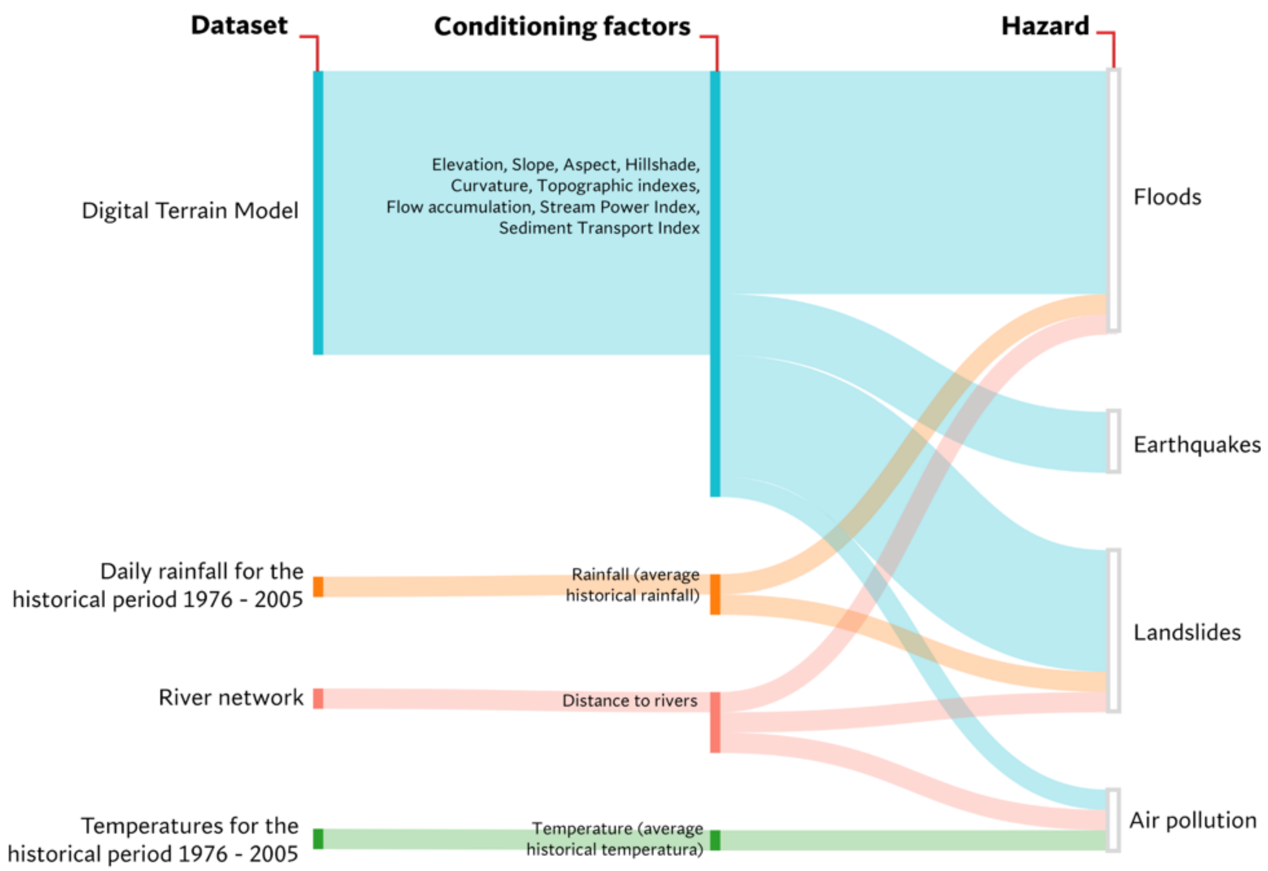

Fig. 4 Connections between datasets, conditioning factors, and hazards for Sofia.

Available online at content.sciendo.com GeoScape 16(2) — 2022: 93—107 doi: 10.2478/geosc-2022-0008

Copernicus Services derived products or Worldpop, may be utilized to cover the missing information.

3.3 Copernicus Services' products

Copernicus is the European Union's Earth Observation (EO) program which offers information services from satellite EO and in-situ data (Copernicus 2022). There are six thematic streams of Copernicus services: land, atmosphere, climate change, emergency, marine, and security. The first four streams are considered particularly relevant for this study. Each one of these four thematic streams provides useful information in Europe for the single and multi-hazard assessment by providing the conditioning factors necessary data when there is no local and/or higher resolution data.

Copernicus Land Monitoring Services (CLMS) provides geographical information on land cover, i.e., Corine Land Cover, the DEM, and the Urban Atlas. Corine Land Cover 2018, with 44 classes and 100 meters of spatial resolution (Copernicus Land Monitoring Service 2018). The DEM is also available at 25 meters resolution as a GeoTIFF (Copernicus Land Monitoring Service 2016a), from which slope, aspect, and hillshade are derived and provided (Copernicus Land Monitoring Service 2018, 2016b,c,d). Other conditioning factors can be derived from the DEM and shall be computed by the reader. Furthermore, the Urban Atlas 2012 provides land use and land cover data for Functional Urban Areas (FUA) (Copernicus Land Monitoring Service 2016e). This service was updated in 2019 to integrate the population data in the service polygons.

Copernicus Atmosphere Monitoring Service (CAMS) has a large product catalog including parameters like carbon dioxide and monoxide, methane, nitrogen oxides, ozone, PM1, PM2.5, PM10, sulfates, solar radiation, and others. This information can be used as part of the input for an air quality assessment (Copernicus Atmosphere Monitoring Service 2022).

Copernicus Climate Change Service (C3S) provides authoritative information and applications to analyze the past, present, and future climate. The relevant C3S datasets for this study include:

- Temperature and precipitation gridded data for global and regional domains derived from in-situ and satellite observations (Copernicus Climate Change Service 2022a): high resolution gridded dataset which integrates temperature and precipitation observations (remote sensed and in-situ) from selected sources. The variables for this dataset are precipitation, temperature (air temperature at 2 meters from the Earth's surface), and temperature anomaly.

- Essential climate variables for assessment of climate variability from 1979 to present (Copernicus Climate Change Service 2022b): contains climatic data, monthly anomalies and monthly mean fields of Essential Climate Variables (ECVs) at a 0.25° x 0.25° horizontal resolution and a monthly temporal resolution. The variables include 0-7cm volumetric soil moisture, precipitation, sea ice cover, surface air relative humidity, and surface air temperature.

- In situ temperature, relative humidity and wind profiles from 2006 to March 2020 from the GRUAN reference network (Copernicus Climate Change Service 2022c): GRUAN stands for The Global Climate Observing System (GCOS) Reference Upper-Air Network and it is an international reference observing network of sites measuring essential climate variables above Earth's surface. It is provided in a point format as a CSV file with 17 stations around the world. Some of the main variables of this dataset are air temperature, air pressure, relative humidity, shortwave radiation, wind from direction, and wind speed.

- In situ observations of meteorological variables from the Integrated Global Radiosounding Archive and the Radiosounding Harmonization dataset from 1978 onward (Copernicus Climate Change Service 2022d): The data is available as points (656 stations around the globe) in a sub-daily temporal resolution and can be downloaded as CSV file. The most relevant variables of this dataset include air dewpoint depression, air temperature, air pressure, ascent speed, eastward wind component, northward wind component, relative humidity, solar zenith angle, water vapor volume mixing ratio, wind from direction, and wind speed.

- E-OBS daily gridded meteorological data for Europe from 1950 to present derived from in-situ observations (Copernicus Climate Change Service 2022e): a regular latitude-longitude gridded dataset, providing data with 0.1° x 0.1° and 0.25° x 0.25° spatial resolution and a daily temporal resolution. The variables of this dataset are land surface elevation, maximum temperature, mean temperature, mini-

GeoScape 16(2) — 2022: 93—107 doi: 10.2478/geosc-2022-0008 Available online at content.sciendo.com

mum temperature, precipitation amount, relative humidity, sea level pressure, surface shortwave downwelling radiation, and wind speed.

- River discharge and related forecasted data by the European Flood Awareness System (Copernicus Climate Change Service 2022f): provides gridded modelled hydrological time series forced with medium-range meteorological forecasts. The dataset is available at a sub-daily high resolution (5km x 5km). The available variables include river discharge in the last 24 hours, river discharge in the last 6 hours, snow depth water equivalent, soil depth, volumetric soil moisture, and others.

The Copernicus Emergency Management Service (CEMS) is divided into two components, a mapping component and an early warning component (Copernicus Emergency Management Service 2022). The mapping component, which has a worldwide coverage, provides maps derived from satellite imagery to support emergency management activities and risk reduction activities. It has been active since 2012. The early warning component is composed of three different systems: The European Flood Awareness System (EFAS), The European Forest Fire Information System (EFFIS), and The European Drought Observatory (EDO). Each one of them has a global component to provide a global coverage. Considering this review, a hazard susceptibility assessment at the urban level, EFAS is relevant to model flooding as it provides overviews on ongoing and forecasted (up to 10 days) floods in Europe.

4 Discussion and conclusions

In this work, urban susceptibility to nature- and human-induced hazards was addressed. The first objective of this paper was the identification of the conditioning factors that affect the susceptibility of an urban area to a series of nature- and man-made hazards. Factors were identified by leveraging methodologies and disclosures of recent papers dealing with the problem of single and multi-hazard mapping through artificial intelligence models. An insightful and methodical literature review was carried out to point out the conditioning factors that have to be considered to properly feed machine and deep learning algorithms with the aim of producing single-hazard susceptibility maps. The scientific literature review permitted to define a list of factors per hazard. Two indicators were conceived and used to assign a first degree of relevance to each factor.

The lists of conditioning factors reported in this paper must be intended as a preliminary result that was achieved through a scientific literature review. However, reported tables will be discussed with experts from different domains (e.g., hydraulics, geotechnics, hydrogeology, climatology), partners of the project Harmonia, to keep the essential variables and include relevant missing factors. The discussion with domain experts will hopefully disclose additional details concerning the data characteristics requirements - such as the needed spatial resolution and temporal frequency of the information - as well as other sources of data and the best performing machine/deep learning techniques to be adopted for the susceptibility mapping.

As a second objective, the work aimed at investigating the availability of open datasets that could support the description of the conditioning factors and thus the hazards' modeling. As data availability depends on the particular context, two cases of study, namely the cities of Milan and Sofia, were here considered. The most concerning hazards for these two cities were identified based on existing documentations and reports, and national/local geoportals were explored to collect suitable datasets that may help describing or deriving the conditioning factors.

Local and regional open datasets found for the two case studies look promising, whilst not sufficient for a thorough description of conditioning factors. Furthermore, significant differences in terms of data availability and characteristics between the two cities were pointed out. Specifically, the datasets available for Milan enable most of the conditioning factors to be described, while way more limited information was found for Sofia, suggesting that the employment of different susceptibility models for the two cities is necessary. However, the large availability of worldwide coverage satellite imagery derived products, such as those provided by the Copernicus Services, partially enables to overcome this limitation. Accordingly, useful datasets made available by the Copernicus Services were identified and listed. Despite not being city-specific and often characterized by a coarser spatial resolution in comparison to local authoritative data, Copernicus datasets represent a pivotal resource to enable the susceptibility and hazard mapping in cities with poor data availability. Furthermore, the use of global-coverage Copernicus datasets would enable a transparent comparison among different case studies being the input variables for different cities coherent and unbiased.

A comparison between the two case studies and a city with no local data is briefly illustrated in Table 4. Based on data availability and characteristics in terms of spatial resolution, temporal frequency and year of last update, an evaluation of single hazard susceptibility modeling feasibility was carried out. For the city of Milan, local datasets may almost entirely describe the conditioning factors lists with reasonably adequate space-time resolution, and they are updated with satisfactory temporal frequency by the data providers. Air pollution may only partially be modeled with city-specific datasets, as some important related conditioning factors cannot be described with local data. For the city of Sofia, landslides and floods may be partially modeled considering the availability of topographical factors and the lack of land cover, geological, and hydrological ones. Meanwhile, it is not possible to model earthquakes and air pollution because the main factors necessary to model these hazards, which are geological/seismic and meteorological, respectively, could not be retrieved from local data.

Table 4 Evaluation of hazard modeling feasibility based on existing open data

Hazard	Milan	Sofia	City with no local data
Ground subsidence	🟢	Not considered	🟡
Landslide	Not considered	🟡	🟡
Flood	🟢	🟡	🟢
Earthquake	🟢	🔴	🔴
Heat island	🟢	Not considered	🟡
Air pollution	🟡	🔴	🟢

Hazard modelling feasibility:
🟢 Possible 🟡 Partially possible 🔴 Not possible

On the other hand, a city with no local data was also included for the sake of comparison. In this case, the hazards shall be modeled considering only Copernicus services which are available at a lower spatial resolution (when compared to local data). Copernicus land services cover topographical factors (by means of DEM), land cover factors (by means of CLC), and land use/cover and population (by means of Urban Atlas). Copernicus Climate Change Services cover the meteorological factors and river discharge data; Copernicus Atmosphere Monitoring Services covers the pollutants data. Therefore, in a city with no local data through the Copernicus services it is possible to completely model flooding and air pollution. It is only partially possible to model ground subsidence, landslides, and heat island due to the lack of key conditioning factors, i.e., groundwater drawdown, distance to faults, and city canyons. Finally, it is not possible to model earthquakes because the lack of geological/seismic conditioning factors.

The single-hazard susceptibility maps that will be obtained for the two case studies with the methodology described in this paper represent a starting point to produce single and multi-hazard/risk maps at the city level, that are foreseen for the future development of this work. The available open datasets will be integrated and pre-processed through existing open technologies, such as Data Cube or Earth Engine, to be easily leveraged as an input to machine and deep learning algorithms.

Acknowledgments

This work was carried out in the frame of the project Harmonia (identifier: H2020-LC-CLA-2018-2019-2020, GA No. 101003517), a EU Horizon 2020 project dedicated to the development of a support system for improved resilience and sustainable urban areas to cope with climate change and urban areas. The project is a collaboration of 22 partners over Europe, involving academia, national research centres, municipalities, non-governmental organizations, and industry. The authors would like to show their gratitude to Vasil Yordanov for his valuable support in the research of open data for the municipality of Sofia.

References

Academy of Disaster Reduction and Emergency Management (2021) Ministry of Emergency Management - Ministry of Education, National Disaster Reduction Center of China, Ministry of Emergency Management, International Federation of Red Cross and Red Crescent Societies. https://www.preventionweb.net/publication/2020-global-natural-disaster-assessment-report.

Adger N (2006) Vulnerability. *Global Environmental Change* 16: 268–281. doi: 10.1016/j.gloenvcha.2006.02.006.

Ahmad H, Ningsheng C, Rahman M, Islam M, Pourghasemi H, Hussain S, Habumugisha J, Liu E, Zheng H, Ni H, Dewan A (2021) Geohazards Susceptibility Assessment along the Upper Indus Basin Using Four Machine Learning and Statistical Models. *ISPRS International Journal of Geo-Information* 10, (5). doi: 10.3390/IJGI10050315.

Alexander D (1995) A survey of the field of natural hazards and disaster studies. In: Carrara A, Guzzetti F, (eds.) *Geographical information systems in assessing natural hazards. Advances in Natural and Technological Hazards Research 5*. Springer, Dordrecht, pp. 1 19. doi: 10.1007/978-94-015-8404-3_1.

GeoScape 16(2) — 2022: 93—107 doi: 10.2478/geosc-2022-0008 Available online at content.sciendo.com

ARPA Lombardia (2022) Qualità aria: i dati 2021 provincia per provincia. Available at: <https://www.arpalombardia.it/Pages/Aria/Qualita-aria.aspx>.

Bevere L, Remondi F (2022) Natural catastrophes in 2021: the floodgates are open. Swiss Re Institute. Available at: <https://www.swissre.com/institute/research/sigma-research/sigma-2022-01.html>.

Bianchini S, Solari L, Del Soldato M, Raspini F, Montalti R, Ciampalini A, Casagli N (2019) Ground Subsidence Susceptibility (GSS) Mapping in Grosseto Plain (Tuscany, Italy) Based on Satellite InSAR Data Using Frequency Ratio and Fuzzy Logic. *Remote Sensing* 11:2015. doi: 10.3390/rs11172015.

Chen S, Yang Y, Deng F, Zhang Y, Liu Y, Liu C, Gao Z (2022) A high-resolution monitoring approach of canopy urban heat island using a random forest model and multi-platform observations. *Atmospheric Measurement Techniques* 15, (3):735–756. doi: 10.5194/amt-15-735-2022.

Choubin B, Abdolshahnejad M, Moradi E, Querol X, Mosavi A, Shamshirband S, Ghamisi P (2020) Spatial hazard assessment of the PM10 using machine learning models in Barcelona, Spain. *Science of The Total Environment* 701:134474. doi: 10.1016/j.scitotenv.2019.134474.

ClimaMI (2019) ClimaMI project. Available at: <https://www.progettoclimami.it/>.

Copernicus (2022) About Copernicus. Available at: <https://www.copernicus.eu/en/about-copernicus>.

Copernicus Atmosphere Monitoring Service (2022) Catalogue. Available at: <https://atmosphere.copernicus.eu/catalogue>.

Copernicus Climate Change Service (2022a) Temperature and precipitation gridded data for global and regional domains derived from in-situ and satellite observations. Available at: <https://cds.climate.copernicus.eu/cdsapp!/dataset/insitu-gridded-observations-global-and-regional>.

Copernicus Climate Change Service (2022b) Essential climate variables for assessment of climate variability from 1979 to present. Available at: <https://cds.climate.copernicus.eu/cdsapp!/dataset/ecv-for-climate-change>.

Copernicus Climate Change Service (2022c) In situ temperature, relative humidity and wind profiles from 2006 to March 2020 from the GRUAN reference network. Available at: <https://cds.climate.copernicus.eu/cdsapp!/dataset/insitu-observations-gruan-reference-network.

Copernicus Climate Change Service (2022d) In situ observations of meteorological variables from the Integrated Global Radiosounding Archive and the Radiosounding Harmonization dataset from 1978 onward. Available at: <https://cds.climate.copernicus.eu/cdsapp!/dataset/insitu-observations-igra-baseline-network>.

Copernicus Climate Change Service (2022e) E-OBS daily gridded meteorological data for Europe from 1950 to present derived from in-situ observations. Available at: <https://cds.climate.copernicus.eu/cdsapp!/dataset/insitu-gridded-observations-europe>.

Copernicus Climate Change Service (2022f) River discharge and related forecasted data by the European Flood Awareness System. Available at: <https://cds.climate.copernicus.eu/cdsapp!/dataset/efas-forecast>.

Copernicus Emergency Management Service (2022) About. Available at: <https://emergency.copernicus.eu>.

Copernicus Land Monitoring Service (2016a) European Digital Elevation Model (EU-DEM), version 1.1. Available at: <https://land.copernicus.eu/imagery-in-situ/eu-dem/eu-dem-v1.1>.

Copernicus Land Monitoring Service (2016b) Slope derived from EU-DEM version 1.0. Available at: <https://land.copernicus.eu/imagery-in-situ/eu-dem/eu-dem-v1-0-and-derived-products/slope>.

Copernicus Land Monitoring Service (2016c) Aspect derived from EU-DEM version 1.0. Available at: <https://land.copernicus.eu/imagery-in-situ/eu-dem/eu-dem-v1-0-and-derived-products/aspect>.

Copernicus Land Monitoring Service (2016d) Hillshade derived from EU-DEM version 1.0. Available at: <https://land.copernicus.eu/imagery-in-situ/eu-dem/eu-dem-v1-0-and-derived-products/hillshade>.

Copernicus Land Monitoring Service (2016e) Urban Atlas LCLU 2012. Available at: <https://land.copernicus.eu/local/urban-atlas/urban-atlas-2012>.

Copernicus Land Monitoring Service (2018) Corine Land Cover. Available at: <https://land.copernicus.eu/pan-european/corine-land-cover/clc2018>.

Dang VH, Hoang ND, L-M-D N, Bui D, Samui P (2020) A Novel GIS-Based Random Forest Machine Algorithm for the Spatial Prediction of Shallow Landslide Susceptibility. *Forests* 11:118. doi: 10.3390/f11010118.

Daniell K (2011) Collaborative flood and drought risk management in the Upper Iskar Basin, Bulgaria. In: Grafton R, Hussey K, (eds.) *Water Resources Planning and Management*. Cambridge University Press, Cambridge, pp. 395–420.

Dimitrova R, Velizarova M (2021) Assessment of the Contribution of Different Particulate Matter Sources on Pollution in Sofia City. *MDPI Atmosphere* 12, (4):423. doi: 10.3390/atmos12040423.

Dos Santos R (2020) Estimating spatio-temporal air temperature in London (UK) using machine learning and earth observation satellite data. *International Journal of Applied Earth Observation and Geoinformation* 88:102066. doi: 10.1016/j.jag.2020.102066.

Ebrahimi-Khusfi Z, Taghizadeh-Mehrjardi R, Kazemi M, Nafarzadegan A (2021) Predicting the ground-level pollutants concentrations and identifying the influencing factors using machine learning, wavelet transformation, and remote sensing techniques. *Atmospheric Pollution Research* 12, (5): 101064. doi: 10.1016/j.apr.2021.101064.

Ebrahimy H, Feizizadeh B, Salmani S, Azadi H (2020) A comparative study of land subsidence susceptibility mapping of Tasuj plane, Iran, using boosted regression tree, random forest and classification and regression tree methods. *Environmental Earth Sciences* 79:223. doi: 10.1007/s12665-020-08953-0.

Elmahdy S, Ali T, Mohamed M (2020) Flash Flood Susceptibility Modeling and Magnitude Index Using Machine Learning and Geohydrological Models: A Modified Hybrid Approach. *Remote Sensing* 12, (17):2695. doi: 10.3390/RS12172695.

EM-DAT (2020) OFDA/CRED The International Disaster Database, Université catholique de Louvain – Brussels – Belgium. Available at: <https://www.emdat.be/>.

Emami S, Yousefi S, Pourghasemi H, Tavangar S, Santosh M (2020) A comparative study on machine learning modeling for mass movement susceptibility mapping (a case study of Iran. *Bulletin of Engineering Geology and the Environment* 79, (10): 5291–5308. doi: 10.1007/S10064-020-01915-7.

Ghorbanzadeh O, Blaschke T, Aryal J, Gholaminia K (2020) A new GIS-based technique using an adaptive neuro-fuzzy inference system for land subsidence susceptibility mapping. *Journal of Spatial Science* 65, (3):401–418. doi: 10.1080/14498596.2018.1505564.

Hakim W, Achmad A, Lee CW (2020) Land Subsidence Susceptibility Mapping in Jakarta Using Functional and Meta-Ensemble Machine Learning Algorithm Based on Time-Series InSAR Data. *Remote Sensing* 12, (21):3627. doi: 10.3390/rs12213627.

HARMONIA (2022) About Harmonia. Available at: <http://harmonia-project.eu/about_us>.

Hong H, Shahabi H, Shirzadi A, Chen W, Chapi K, Ahmad B, Roodposhti M, Yari Hesar A, Tian Y, Tien Bui D (2019) Landslide susceptibility assessment at the Wuning area, China: a comparison between multi-criteria decision making, bivariate statistical and machine learning methods. *Natural Hazards* 96, (1):173–212. doi: 10.1007/S11069-018-3536-0.

Hu H, Wang C, Liang Z, Gao R, Li B (2021) Exploring Complementary Models Consisting of Machine Learning Algorithms for Landslide Susceptibility Mapping. *ISPRS International Journal of Geo-Information* 10, (10). doi: 10.3390/IJGI10100639.

ISPRA (2020) Comuni interessati da subsidenza. Available at: <https://annuario.isprambiente.it/sys_ind/733>.

ISPRA (2021) Progetto IFFI (Inventario dei Fenomeni Franosi in Italia. Available at: <https://www.progettoiffi.isprambiente.it/cartografia-on-line/>.

Ivanov P (2017) Analysis and Mapping the Landslide Hazard in Bulgaria. In: Mikoš M, Tiwari B, Yin Y, Sassa K, (eds.) *Advancing Culture of Livingwith Landslides, Volume 2, Advances in Landslide Science*. Springer Nature, pp. 1111–1118. doi: 10.1007/978-3-319-53498-5_126.

Ivanov P, Dobrev N, Berov B, Krastanov M, Nankin (2020) Assessment of landslide hazard in Bulgaria using GIS. *Proceedings International Conference on Cartography and GIS* 1, (8). doi: 10.1007/978-3-319-53498-5.

Jena R, B P, Alamri A (2020a) Susceptibility to Seismic Amplification and Earthquake Probability Estimation Using Recurrent Neural Network (RNN) Model in Odisha, India. *Applied Sciences* 10, (15):5355. doi: 10.3390/app10155355.

Jena R, Pradhan B, Al-Amri A, Lee C, H-j P (2020b) Earthquake Probability Assessment for the Indian Subcontinent Using Deep Learning. *Sensors* 20, (16):4369. doi: 10.3390/s20164369.

Jena R, Pradhan B, Beydoun G, Alamri A, Nizamuddin A, Sofyan H (2020c) Earthquake hazard and risk assessment using machine learning approaches at Palu, Indonesia. *Science of the Total Environment* 749:141582. doi: 10.1016/j.scitotenv.2020.141582.

Jena R, Pradhan B, Beydoun G, Ardiansyah N, Sofyan H, Affan M (2020d) Integrated model for earthquake risk assessment using neural network and analytic hierarchy process: Aceh province, Indonesia. *Geoscience Frontiers* 11, (2):613–634,. doi: 10.1016/j.gsf.2019.07.006.

Kelman I, Gaillard J, Mercer J (2015) Climate change's role in disaster risk reduction's future: beyond vulnerability and resilience. *International Journal of Disaster Risk Science* 6: 21–27. doi: 10.1007/s13753-015-0038-5.

Lei X, Chen W, Panahi M, Falah F, Rahmati O, Uuemaa E, Kalantari Z, Ferreira C, Rezaie F, Tiefenbacher J, Lee S, Bian H (2021) Urban flood modeling using deep-learning approaches in Seoul, South Korea. *Journal of Hydrology* 601:126684. doi: 10.1016/J.JHYDROL.2021.126684.

Ministry of Regional Development and Public Works (2022) Landslide inventory. Available at: <http://gz-pernik.mrrb.government.bg/landslide/>.

Mohammady M, Pourghasemi H, Amiri M (2019) Land subsidence susceptibility assessment using random forest machine learning algorithm. *Environmental Earth Sciences* 78:503. doi: 10.1007/s12665-019-8518-3.

Na T, Kawamura Y, seung Kang S, Utsuki S (2021) Hazard mapping of ground subsidence in east area of Sapporo using frequency ratio model and GIS. *Geomatics, Natural Hazards and Risk* 12, (1):347–362. doi: 10.1080/19475705.2021.1873198.

Nachappa T, Ghorbanzadeh O, Gholamnia K, Blaschke T (2020) Multi-Hazard Exposure Mapping Using Machine Learning for the State of Salzburg, Austria. *Remote Sensing* 12, (17):2757. doi: 10.3390/RS12172757.

Oh HJ, Syifa M, Lee CW, Lee S (2019) Land Subsidence Susceptibility Mapping Using Bayesian, Functional, and Meta-Ensemble Machine Learning Models. *Applied Sciences* 9, (6): 1248. doi: 10.3390/app9061248.

Okumus D, Terzi F (2021) Evaluating the role of urban fabric on surface urban heat island: The case of Istanbul. *Sustainable Cities and Society* 73:103128. doi: 10.1016/j.scs.2021.103128.

Park S, Im J, Park S, Rhee J (2017) Drought monitoring using high resolution soil moisture through multi-sensor satellite data fusion over the Korean peninsula. *Agricultural and Forest Meteorology* 237–238:257–269. doi: 10.1016/J.AGRFORMET.2017.02.022.

Paskaleva I, Panza G, Vaccari F, Ivanov P (2004) Deterministic modelling for microzonation of Sofia - An expected earthquake scenario. *Acta Geodaetica et Geophysica Hungarica* 39: 275–295. doi: 10.1556/AGeod.39.2004.2-3.10.

Pham B, Luu C, Dao D, Phong T, Nguyen H, Le H, Meding J, Prakash I (2021) Flood risk assessment using deep learning integrated with multi-criteria decision analysis. *Knowledge-Based Systems* 219:106899. doi: 10.1016/J.KNOSYS.2021.106899.

Ranjgar B, Razavi-Termeh S, Foroughnia F, Sadeghi-Niaraki A, Perissin D (2021) Land Subsidence Susceptibility Mapping Using Persistent Scatterer SAR Interferometry Technique and Optimized Hybrid Machine Learning Algorithms. *Remote Sensing* 13, (7):1326. doi: 10.3390/rs13071326.

Raška P, Dolejš M, Pacina J, Popelka J, Píša J, Rybová K (2020) Review of current approaches to spatially explicit urban vulnerability assessments: hazard complexity, data sources, and cartographic representations. *Geoscape* 14, (1):47–61. doi: 10.2478/geosc-2020-0005.

Regione Lombardia (2014) Seismic Zonation. Available at: <https://www.regione.lombardia.it/wps/portal/istituzionale/HP/DettaglioRedazionale/servizi-e-informazioni/Enti-e-Operatori/protezione-civile/Rischio-sismico/nuova-zonazione-sismica/zonazione-sismica>.

Resilient Cities Network (2022) What is urban resilience. Available at: <https://resilientcitiesnetwork.org/what-is-resilience/>.

Sangiorgio V, Fiorito F, Santamouris M (2020) Development of a holistic urban heat island evaluation methodology. *Scientific Reports* 10:17913. doi: 10.1038/s41598-020-75018-4.

Satarzadeh E, Sarraf A, Hajikandi H, Sadeghian M (2022) Flood hazard mapping in western Iran: assessment of deep learning vis-à-vis machine learning models. *Natural Hazards* 111: 1355–1373. doi: 10.1007/s11069-021-05098-6.

Schneider R, Vicedo-Cabrera A, Sera F, Masselot P, Stafoggia M, Hoogh K, Kloog I, Reis S, Vieno M, Gasparrini A (2020) A Satellite-Based Spatio-Temporal Machine Learning Model to Reconstruct Daily PM2.5 Concentrations across Great Britain. *Remote Sensing* 12:3803. doi: 10.3390/rs12223803.

Shogrkhodaei S, Razavi-Termeh SV, Fathnia A (2021) Spatio-temporal modeling of PM2.5 risk mapping using three machine learning algorithms. *Environmental Pollution* 289: 117859. doi: 10.1016/j.envpol.2021.117859.

Skilodimou H, Bathrellos G (2021) Natural and Technological Hazards in Urban Areas: Assessment, Planning and Solutions. *Sustainability* 13, (15):8301. doi: 10.3390/su13158301.

Skilodimou H, Bathrellos G, Chousianitis K, A Y, Pradhan B (2019) Multi-hazard assessment modeling via multi-criteria analysis and GIS: a case study. *Environmental Earth Sciences* 78:47. doi: 10.1007/s12665-018-8003-4.

Sofiaplan (2018) Areas with significant potential flood risk. Available at: <https://api.sofiaplan.bg/datasets/51>.

Sofiaplan (2022) Rights to use the published materials. Available at: <https://sofiaplan.bg/>.

Tien Bui D, Shahabi H, Shirzadi A, Chapi K, Pradhan B, Chen W, Khosravi K, Panahi M, Bin Ahmad B, Saro L (2018) Land Subsidence Susceptibility Mapping in South Korea Using Machine Learning Algorithms. *Sensors* 18, (8). ISSN 1424-8220. doi: 10.3390/s18082464.

United States Environmental Protection Agency (EPA) (2021) Climate Change Indicators: U.S. and Global Precipitation. Available at: <https://www.epa.gov/climate-indicators/climate-change-indicators-us-and-global-precipitation>.

Yao Y, Chang C, Ndayisaba F, Wang S (2020) A New Approach for Surface Urban Heat Island Monitoring Based on Machine Learning Algorithm and Spatiotemporal Fusion Model. *IEEE Access* 8:164268–164281. doi: 10.1109/ACCESS.2020.3022047.

Zheng X, He G, Wang S, Wang Y, Wang G, Yang Z, Yu J, Wang N (2021) Comparison of Machine Learning Methods for Potential Active Landslide Hazards Identification with Multi-Source Data. *ISPRS International Journal of Geo-Information* 10, (4): 253. doi: 10.3390/IJGI10040253.

GeoScape 16(2) — 2022: 93—107 doi: 10.2478/geosc-2022-0008 Available online at content.sciendo.com

Supplementary material

Table 5 Supplementary table

Ground subsidence [1]			
Type of factor	Physical variable	Unit	Papers
Land cover	Land use/land cover	-	[2-10]
Topographical	Slope	degrees	[3-10]
Geological/seismic	Lithology/geology	-	[2-9]
Hydrological	*Groundwater drawdown/level*	m below Ground Level (GL)	[3] [5-10]
Hydrological	*Distance to the river network*	m	[5] [6] [8-10]
Topographical	*Elevation*	m	[5] [6] [8-10]
Topographical	*Aspect*	-	[5] [6] [8-10]
Hydrological	*Topographic Wetness Index*	-	[6] [8-10]
Geomorphological	*Plan curvature*	-	[6] [8-10]
Land cover	*Distance to the road network*	m	[3] [6] [9] [10]
Geological/seismic	*Distance to the fault*	m	[2] [6] [9]
Hydrological	*Rainfall*	mm/year	[3] [5] [6]
Geomorphological	*Profile curvature*	-	[6] [9] [10]
Geological/seismic	*Distance to drift/lineament*	m	[4] [7]
Geological/seismic	*Rock Mass Rating*	-	[4] [7]
Topographical	*Stream Power Index/stream density*	-	[8] [10]
Geological/seismic	*Drift density*	m/m^2	[4]
Hydrological	*Well density*	m/m^2	[5]
Geological/seismic	*Permeability*	m/s	[7]
Geological/seismic	*Aquifer unit (composition)*	-	[2]
Geological/seismic	*(Sedimentary) cover thickness*	m	[2]
Geological/seismic	*Earthquake intensity*	MMI (Modified Mercalli Intensity)	[3]

Definition: [1] https://oceanservice.noaa.gov/facts/subsidence.html (National Oceanic and Atmospheric Administration - NOAA)

References: [2] Bianchini et al. (2019) [3] Na et al. (2021) [4] Tien Bui et al. (2018) [5] Ghorbanzadeh et al. (2020) [6] Hakim et al. (2020) [7] Oh et al. (2019) [8] Ebrahimy et al. (2020) [9] Mohammady et al. (2019) [10] Ranjgar et al. (2021)

Available online at content.sciendo.com GeoScape 16(2) — 2022: 93—107 doi: 10.2478/geosc-2022-0008

Table 6 Supplementary table

Landslides [1]			
Type of factor	Physical variable	Unit	Papers
Geological	*Distance to faults*	km	[2-8]
Topographical	*Slope*	degrees	[2-8]
Topographical	*Elevation*	m	[2-8]
Topographical	*Aspect*	-	[3-8]
Land cover	*Land use/land cover*	-	[3-8]
Geological	*Lithology/geology*	-	[3-8]
Hydrological	*Topographic Wetness Index*	-	[2] [4-8]
Land cover	*Distance to the road network*	km	[2-4] [6-8]
Hydrological	*Distance to streams*	km	[2-4] [6-8]
Hydrological	*Rainfall*	mm/year	[2] [3] [7] [8]
Geomorphological	*Profile curvature*	-	[2] [5] [6] [7]
Geomorphological	*Plan curvature*	-	[2] [4] [5] [6]
Hydrological	*Stream Power Index*	-	[5] [7] [8]
Land cover	*NDVI*	-	[6] [7] [8]
Geological	*Soil type*	-	[5] [8]
Geological	*Sediment Transport Index*	-	[5] [8]
Topographical	*Slope length*	-	[5] [6]
Hydrological	*Drainage density*	-	[4] [6]
Geological	*Valley depth*	m	[5]
Geological	*Convergence Index*	-	[4]
Topographical	*Surface roughness*	-	[6]
Topographical	*Terrain relief*	-	[6]
Topographical	*Roundness*	-	[2]

Definition: [1] https://www.usgs.gov/faqs/what-landslide-and-what-causes-one (United States Geological Survey - USGS) References: [2] Hu et al. (2021) [3] Nachappa et al. (2020) [4] Emami et al. (2020) [5] Dang et al. (2020)) [6] Zheng et al. (2021) [7] Ahmad et al. (2021) [8] Hong et al. (2019)

GeoScape 16(2) — 2022: 93—107 doi: 10.2478/geosc-2022-0008 Available online at content.sciendo.com

Table 7 Supplementary table

Floods [1]			
Type of factor	Physical variable	Unit	Papers
Topographical	*Slope*	degrees	[2-7]
Topographical	*Elevation/Digital Elevation Model*	m	[2-7]
Land cover	*Land use/land cover*	-	[2] [4-7]
Hydrological	*Distance to streams/rivers*	km	[2-5] [7]
Hydrological	*Stream power index*	-	[2] [4] [5] [7]
Hydrological	*Topographic wetness index*	-	[2] [4] [5] [7]
Hydrological	*Rainfall*	mm/year	[2] [5] [6]
Geological	*Lithology/geology*	-	[2] [3] [5]
Hydrological	*Stream/river density*	-	[3-5]
Topographical	*Profile curvature*	-	[5] [7]
Topographical	*Plan curvature*	-	[5] [7]
Geological	*Sediment transport index*	-	[4] [5]
Topolographical	*Topographic position index*	-	[4] [7]
Topolographical	*Topographic ruggedness index*	-	[4] [7]
Hydrological	*Flow accumulation*	-	[5]
Topographical	*Slope length factor*	-	[7]
Land cover	*Distance to the road network*	km	[2]
Topographical	*Aspect*	-	[2]
Land cover	*NDVI*	-	[2]
Topographical	*Topographic minimum curvature*	-	[3]
Topographical	*Topographic relief*	-	[3]
Geological	*Soil type*	-	[4]

Definition: [1] https://www.nssl.noaa.gov/education/svrwx101/floods/ (National Oceanic and Atmospheric Administration - NOAA) References: [2] Nachappa et al. (2020) [3] Elmahdy et al. (2020) [4] Satarzadeh et al. (2022) [5] Pham et al. (2021) [6] Park et al. (2017) [7] Lei et al. (2021)

Table 8 Supplementary table

Earthquakes [1]			
Type of factor	Physical variable	Unit	Papers
Topographical	*Slope*	degrees	[2-5]
Topographical	*Elevation*	m	[2-5]
Geological/seismic	*Epicenter density*	no./km^2	[2-5]
Geological/seismic	*Proximity to fault*	km	[2-5]
Geological/seismic	*Geology*	Amplification factor	[2-5]
Geological/seismic	*Depth density*	no./km^2	[2] [3] [5]
Topographical	*Curvature*	-	[2] [3] [5]
Geological/seismic	*Fault density*	no./km^2	[3-5]
Geological/seismic	*Proximity to epicenters*	km	[3-5]
Geological/seismic	*Magnitude density*	Mw/km^2 (Mw: moment magnitude)	[2-4]
Geological/seismic	*Peak Ground Acceleration density*	g/km^2 (g: gravity)	[4] [5]
Geological/seismic	*Amplification factor*	-	[2] [3]
Topographical	*Aspect*	-	[5]
Geological/seismic	*Intensity distribution*	Magnitude	[5]
Geological/seismic	*Peak Ground Acceleration*	g	[2]
Geological/seismic	*Intensity variation*	-	[3]

Definition: [1] https://www.usgs.gov/faqs/what-earthquake-and-what-causes-them-happen (United States Geological Survey - USGS) References: [2] Jena et al. (2020b) [3] Jena et al. (2020c) [4] Jena et al. (2020d) [5] Jena et al. (2020a)

Available online at content.sciendo.com GeoScape 16(2) — 2022: 93—107 doi: 10.2478/geosc-2022-0008

Table 9 Supplementary table

Heat island [1]			
Type of factor	Physical variable	Unit	Papers
Land cover	*Land cover types (albedo)*	%	[2-6]
Land cover	*NDVI (and other land cover indexes)*	-	[2] [3] [5] [6]
Topographical	*Elevation*	m	[2] [5] [6]
Topographical	*Slope*	degrees	[2] [5] [6]
Topographical	*Aspect*	-	[2] [5] [6]
Land cover	*Land cover (greenery)*	%	[4] [5]
Topographical	*Longitude and latitude*	degrees	[2] [6]
Meteorological	*Windless days*	%	[4]
Meteorological	*Average max summer temperature*	°C	[4]
Meteorological	*Average summer thermal excursion*	°C	[4]
Meteorological	*Clear sky days*	%	[4]
Anthropogenic heat	*Population density*	no./km^2	[4]
City canyons	*Building height*	UCZ (Urban Climate Zones)	[4]
City canyons	*Average width of streets*	m	[4]
City canyons	*Canyons orientation*	-	[4]
City canyons	*Irregularities of the city*	-	[4]
Topographical	*Proportion of land use/cover area*	-	[6]
Topographical	*Distance from city center*	km	[6]
Topographical	*Proportion of impervious surface area*	%	[6]
Anthropogenic heat	*Anthropogenic heat flux*	W/m^2	[6]
Topographical	*Distance from the coast*	km	[2]
Climatic	*Land surface temperature*	°C	[2]
Climatic	*Sun zenith angle*	degrees	[2]
Topographical	*Hillshade*	-	[5]
City canyons	*Building Coverage Ratio*	-	[3]
City canyons	*Surface/Volume ratio*	-	[3]
City canyons	*Sky View Factor*	-	[3]
City canyons	*Canyon Geometry Factor*	-	[3]

Definition: [1] https://www.epa.gov/heatislands/learn-about-heat-islands (United States Environmental Protection Agency - USEPA) References: [2] Dos Santos (2020) [3] Okumus and Terzi (2021) [4] Sangiorgio et al. (2020) [5] Yao et al. (2020) [6] Chen et al. (2022)

GeoScape 16(2) — 2022: 93—107 doi: 10.2478/geosc-2022-0008 Available online at content.sciendo.com

Table 10 Supplementary table

Air pollution [1]			
Type of factor	Physical variable	Unit	Papers
Meteorological	*Wind speed*	m/s	[2-5]
Meteorological	*Annual average precipitation*	mm	[2-5]
Socio-economic	*Highway/road density*	m/km^2	[2-4]
Land cover	*NDVI*	-	[2-5]
Meteorological	*Humidity index*	-	[3-5]
Meteorological	*Wind direction distribution*	%	[2] [4] [5]
Topographic	*Elevation*	m	[2] [4]
Socio-economic	*Population density*	$no./km^2$	[3] [4]
Meteorological	*Annual average temperature*	°C	[4] [5]
Meteorological	*Minimum/maximum temperature*	°C	[2] [3]
Land cover	*Land use/land cover*	-	[2] [4]
Topographical	*Topographic Wetness Index*	-	[2]
Topographical	*Terrain Roughness Index*	-	[2]
Land cover	*Distance from water body*	km	[2]
Socio-economic	*Distance from airports and seashore*	km	[4]
Meteorological	*Actual evapotranspiration*	mm	[5]
Meteorological	*Meteorological drought*	-	[5]
Meteorological	*Soil moisture*	%	[5]
Meteorological	*Vapor pressure*	Pascal	[5]
Meteorological	*Soil heat flux*	W/m^2	[5]
Meteorological	*Dew point*	°C	[5]

Definition: [1] https://www.environmentalpollutioncenters.org/air/ (Environmental Pollution Centers) References: [2] Choubin et al. (2020) [3] Shogrkhodaei et al. (2021) [4] Schneider et al. (2020) [5] Ebrahimi-Khusfi et al. (2021)