Applied Geography 82 (2017) 59–65

Contents lists available at ScienceDirect

Applied Geography

journal homepage: www.elsevier.com/locate/apgeog

Validation of landslide hazard models using a semantic engine on online news

Alessandro Battistini, Ascanio Rosi, Samuele Segoni*, Daniela Lagomarsino [1], Filippo Catani, Nicola Casagli

Department of Earth Sciences, University of Firenze, Via La Pira n°4, 50121 Firenze, Italy

ARTICLE INFO

Article history:
Received 8 July 2015
Received in revised form
10 March 2017
Accepted 12 March 2017

Keywords:
Warning systems
Automatic validation
Semantic engine
Geohazards
Inventories
Landslides
Geotagging

ABSTRACT

The objective of this work is twofold: (i) automatically setting up a landslide inventory using a state-of-the art semantic engine based on data mining on online news and (ii) evaluating if the automatically generated inventory can be used to validate a regional scale landslide warning system based on rainfall-thresholds.

The semantic engine scanned internet news in real time in a 50 months test period. At the end of the process, an inventory of approximately 900 landslides was automatically set up for the Tuscany region (23,000 km^2, Italy). Using a completely automated procedure, the inventory was compared with the outputs of the regional landslide early warning system and a good correspondence was found, e.g. 84% of the events reported in the news is correctly identified by the warning system.

On the basis of the obtained results, we conclude that automatic validation of landslide models using geolocalized landslide events feedback is possible. The source of data for validation can be obtained directly from the Internet channel using an appropriate semantic engine dedicated to perform a monitoring of the Google News aggregator.

Moreover, validation statistics can be used to evaluate the effectiveness of the predictive model and, if deemed necessary, an update of the rainfall thresholds could be performed to obtain an improvement of the forecasting effectiveness of the warning system.

In the near future, the proposed procedure could operate in continuous time and could allow for a periodic update of landslide hazard models and landslide early warning systems with minimum or none human intervention.

1. Introduction

Landslides triggered by rainfall are one of the major natural hazards to human life and infrastructure worldwide (Petley, 2012).

One of the key elements in landslide hazard assessment, especially at regional scale, are landslide inventories. As an instance, when the location of past landslides is known with sufficient accuracy, landslide inventories can be used for landslide susceptibility assessments (Aleotti & Chowdhury, 1999; Manzo, Tofani, Segoni, Battistini & Catani, 2013; Van Den Eeckhaut & Hervás, 2012). When both location and time of triggering are known in detail, landslide inventories can be used to correctly calibrate physically based models (Baum, Godt, & Savage, 2010; Crosta & Frattini, 2003; Mercogliano et al., 2013; Rossi, Catani, Leoni, Segoni, & Tofani, 2013). Landslide inventories, even with lower spatial accuracy, can be analyzed and correlated with rainfall measurements to establish empirical rainfall thresholds for the triggering of landslides (Caine, 1980; Guzzetti, Peruccacci, Rossi, & Stark, 2008; Peruccacci, Brunetti, Luciani, Vennari, & Guzzetti, 2012; Segoni, Rossi, Rosi, & Catani, 2014a; Vennari et al., 2014). More in general, landslide inventories can be used to calibrate and validate warning systems (Calvello & Piciullo, 2016; Capparelli & Tiranti, 2010; Piciullo et al., 2016; Segoni et al., 2015a) or landslide hazard models (Chung & Fabbri, 2003; Segoni, Lagomarsino, Fanti, Moretti, & Casagli, 2015b), in risk assessment (Cardinali et al., 2006; Lepuschitz, 2015), to characterize the impact of extreme events on the environment and the society at all scales

* Corresponding author.
E-mail addresses: alessandro.battistini@gmail.com (A. Battistini), ascanio.rosi@unifi.it (A. Rosi), samuele.segoni@unifi.it (S. Segoni), daniela.lagomarsino80@gmail.com (D. Lagomarsino), filippo.catani@unifi.it (F. Catani), nicola.casagli@unifi.it (N. Casagli).
[1] Now at Eni S.p.A, S. Donato Milanese, Milano, Italy.

http://dx.doi.org/10.1016/j.apgeog.2017.03.003
0143-6228/

(Anderson et al., 2011; Petley, 2012), and to investigate landslide physical and statistical properties (Malamud, Turcotte, Guzzetti, & Reichenbach, 2004).

For these reasons, landslide inventories at different scales (from the local to the national or global one) have become an important objective for the research community and for governmental agencies (Damm & Klose, 2015; Kirschbaum, Adler, Hong, Hill, & Lerner-Lam, 2010; Petley, 2012; Van Den Eeckhaut & Hervás, 2012; Trigila, Iadanza, & Spizzichino, 2010).

Traditionally, the most used methods to set up landslide inventories are remote sensing methodologies (Guzzetti et al., 2012; Tofani, Segoni, Agostini, Catani, & Casagli, 2013), field surveys (Brunsden, 1985), retrieval of data from technical reports and/or local newspapers (Guzzetti et al., 2008; Lagomarsino, Segoni, Fanti, & Catani, 2013; Vennari et al., 2014) or a combination of them (Dikau, Cavallin, & Jäger, 1996; Rosi, Segoni, Catani, & Casagli, 2012). In any case, the setting up of landslide inventories is a time consuming activity, which could be very costly, and could often be the objective of an entire research project or technical activity (Damm & Klose, 2015; Hilker, Badoux, & Hegg, 2009; Petley, 2012).

This is true for both post-event inventories (i.e. landslides occurred after a particular triggering event like an earthquake or a rainstorm) and historical inventories (i.e. systematic collection of all the landslides that in times occurred in a given area).

This brings the disadvantage that the application of the landslide inventory for the validation or calibration of landslide hazard models, or for the characterization of the hazardous event itself, can occur only after a long time after the landslide events took place.

The objective of this work is to explore the feasibility of reducing this gap and automatically set up a landslide inventory using a state-of-the art semantic engine based on data mining on online news (Battistini, Segoni, Manzo, Catani, & Casagli, 2013) and to evaluate if the automatically generated inventory can be used to validate a regional scale landslide warning system based on rainfall-thresholds or even to automatically update the thresholds.

We adopted the methodology of Battistini et al. (2013), which has already been tested at national scale and is based on recent technologies that publish news in internet. This source of information allows a continuous feedback from real world and news concerning landslide events can be collected very rapidly and used in extremely shorter times compared to traditional techniques.

In this manuscript, we report the outcomes of the operational employ in support of the Tuscany Region landslide early warning system (LEWS henceforth) (Segoni et al., 2015a). We also automated the validation procedure, which is based on a comparison between forecasts and reported events. We verified that our approach can be automatically used for a near real time validation of the warning system and for a semi-automatic update of the rainfall thresholds, which could lead to an improvement of the forecasting effectiveness of the warning system (Rosi et al., 2015).

2. Materials and methods

2.1. Study area

Tuscany is a region located in central Italy, with a mainly hilly and mountainous territory (Fig. 1).

The main reliefs of the region are the Apuan Alps (to the North) and the Apennine mountain belt (to the East). While the former are constituted by metamorphic rocks, the latter are made up by flysch. The Apennine reliefs alternate with intermontane basins filled with lacustrine and fluvial granular and cohesive deposits. The rest of the territory is mainly hilly, with limited plain areas (mainly in the coast or along the main river courses) and isolated mountains (Mt Amiata).

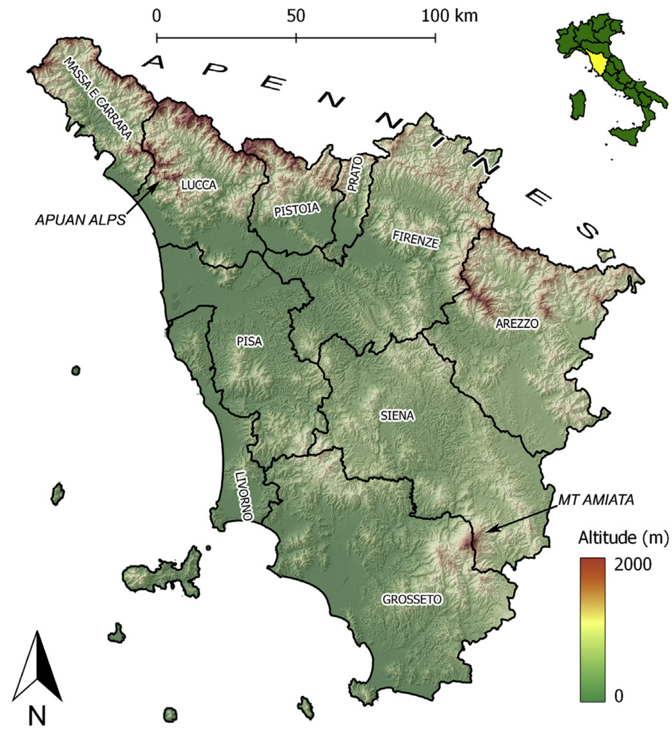

Fig. 1. Test site description and location.

The rainfall regime is typically Mediterranean, with dry summers, a main peak of precipitation during autumn and a secondary one in spring or winter. The spatial distribution of the reliefs influences the spatial distribution of rainfall: the most rainy areas are located in the mountains, especially in the Apuan Alps (mean annual precipitation above 2000 mm/y), while the flat southern part of the region has a mean annual precipitation lower than 600 mm/y (Rosi et al., 2012).

Landsliding is one of the main geomorphic processes of Tuscany. According to rainfall characteristics, bedrock lithology and landscape morphology, different landslide typologies affect Tuscany. Shallow landslides prevail in the northern steep mountainsides made up of flysch or schist rocks, while reactivations of rotational slides usually take place in the hills made up of soils or soft rocks.

2.2. Regional landslide early warning system

A regional scale landslide early warning system (LEWS) based on intensity-duration rainfall thresholds is operating in Tuscany (Segoni et al., 2015a). Due to the high meteorological and geomorphological heterogeneity of Tuscany, a single regional threshold would be ineffective (Rosi et al., 2012). Therefore, the region was divided into 25 homogeneous Alert Zones (AZ) and a distinct threshold was defined for each of them (Segoni, Rosi, Rossi, Catani, & Casagli, 2014b). The threshold analysis was performed using a standardized and highly automated methodology that ensures a high degree of objectivity and a fully consistent reproduction by the automated LEWS (Segoni et al., 2014a).

LEWS monitors in continuous time the rainfall measurements collected by a network of more than 300 automated rain gauges. The measurements of each rain gauge are compared with the rainfall thresholds of the reference AZ and, when a threshold is exceeded, LEWS issues an alert spatially limited to that alert zone. In addition, the system automatically stores in a DBMS (Data Base