Asian Journal of Information Technology 17 (5): 287-297, 2018
ISSN: 1682-3915

A Rough Method for the Detection of Hotspots along Railway Lines

[1]Paolino Di Felice and [2]Antonello Di Felice
[1]Department of Industrial and Information Engineering and Economics,
[2]Department of Information Engineering, University of L'Aquila, L'Aquila, Italy

Abstract: Safety is a relevant issue in many domains. Unfortunately, in the real life the implementation of programs devoted to achieve the safety is impeded by the budget of public administrations which is in constant contraction, especially in periods of crisis as those we are living, since, many years. The study proposes a scientifically robust method for the identification of the top-N list of railway Hotspots that can be used as input for the definition of a strategy of selective monitoring of the state of safety the railway network of an administrative unit (e.g., a region) with respect to the exposure to the landslide hazard. The knowledge of the Hotspots as meant in this study is a conceptual tool for providing a rigorous analytical basis for narrowing down a global problem train derailments to smaller, highest risk, geographic areas where the management of the disaster risk is most crucial. The method we propose is simple to understand and to implement. As counterpart, it may return false positives from which originates the denomination of rough method. Nevertheless, our method is suitable to implement a "Do more with less" strategy with respect to the case where the railway lines have to be inspected in full.

Key words: Railway, landslide, Hotspot, exposure, topological relation, 9-intersection model

INTRODUCTION

Safety is the primary requirement for railway transportation. A recent regulation (RSSB, 2014) released by the Rail Safety and Standards Board Ltd. (http://www.rssb.co.uk/) defines "A hazard as a condition that could lead to an accident". Train derailments are the primary risk of accidents. Terrain instability is among the most severe causes of derailments.

This study proposes a method that consists of two steps. The first step, identifies the portions of railway lines (hereinafter, called Hotspots) where the danger to have terrain instability is high because of the geomorphology of the underlying terrain. If the ground where the tracks are leaned or that in their immediate vicinity, undergoes an altitude variation (for example, due to prolonged rains) then modifications in the bed of the tracks are possible and this in turn becomes a motivating factor of railway derailments.

The second step returns the identified Hotspots in order of decreasing propensity to the aforementioned terrain instability. This output allows the scheduling of selective inspections, of tretches of railway lines limiting, thus, the use of resources without lowering their level of safety.

The method we propose is simple to understand and to implement. Often, to keep things simple entails a price to be paid. In our case simplicity entails that the method may return false positives. However, as it will be explained forward, the detection of the false positives is rapid and simple.

Many methods have been developed to predict, prevent and mitigate accidents in the railway context. He *et al*. (2011) for example, resort to the principal component analysis to identify a limited number of factors to be taken into account in order to study the rock slope stability. Theoretical studies like that just mentioned are very interesting because they help to understand the complexity of the problem by providing global information about the variables that come into play and how they are related each other but they have the limitation of not being directly translatable into a monitoring work plan to be carried out by those who have the responsibility to ensure the safety of railway networks.

In the study we adopt a different approach based on methods proposed in the field of geographical information science to study the topological relationships between geographic features. The method we are going to describe enables the identification of the Hotspots along a train line which are most exposed to terrain instability and therefore, a source of high potential danger of derailments. These Hotspots are those that primarily, require periodic inspections. Such a strategy exceeds the limits of the periodic checks on the entire railway path

Corresponding Author: Paolino Di Felice, Department of Industrial and Information Engineering and Economics, University of L'Aquila, L'Aquila, Italy

which are long and consequently, expensive and not always useful as stressed by Sadler *et al.* (2016). They developed a geospatial safety risk model which can help to identify "Low-risk locations or routes and areas where control measures can be safely reduced, offering the potential for the railway industry to achieve significant cost and efficiency savings through the removal of unnecessary/disproportionate control measures."

The aim of our method is identical to that of Sadler *et al.* (2016). In fact, by returning the location of the portions of railway lines where the exposure to the landslide hazard is highest, actually it drives the people responsible for the railway safety in the definition of monitoring plans and consequently, the portions where such supervision can be thinned without lowering the level of safety for the people and for the transported assets. In summary, the goal of the study described in this study is to develop a scientifically robust methodology for the identification of the railway Hotspots which in turn is the input for the definition of a strategy of selective monitoring of the safety status of a category of assets strategic for a country that is its railway network with respect to the exposure to landslides. The exposure we refer to in this study is called "Spatial" exposure by (SafeLand, 2011).

MATERIALS AND METHODS

Notations

Here in after we use the following notations: GeoArea is the portion of land of interest for the study. GeoArea is defined as the pair<description, geometry of the boundary of GeoArea> where description is a string.

$Z = \{z_k (k = 1, 2, \ldots ,)\}$, where z_k is a "Zone". Accord ing to the existing literature (Guzzetti *et al.*, 2006; Fell *et al.*, 2008) by zone we mean a portion of land characterized by a set of ground conditions. The generic element of Z can be modelled as a "Simple polygon". According to the OpenGIS abstract specification (Ryden, 2005) a simple polygon consists of a single "Patch" that is associated with 1 exterior boundary and 0 or more interior boundaries. Each interior boundary defines a hole in the polygon. The boundary of a simple polygon is the set of closed curves corresponding to its exterior and interior boundaries. Card(Z) denotes the cardinality of set Z, i.e., the number of its elements. Δz_k denotes the boundary of zone z_k while z_k° denotes its interior.

The elements in Z make a full partition of GeoArea. In other words, knowing set Z for a given area is equivalent to have built a map like that prepared by Guzzetti *et al.* (2006) for the collazzone area (central Italy). The generic element of Z (i.e., z_k) is defined as the tuple <ID, Sz_k, boundary of z_k>, being ID an identifying code. Sz_k is a numerical value that quantifies the (spatial) probability that z_k produces landslides. The value of Sz_k ranges from 0-1. Brabb and Harrod (1989) introduced the term susceptibility to denote such a quantitative estimate. Assessing and mapping landslide susceptibility has been a relevant issue (Guzzetti *et al.*, 2006; Fell *et al.*, 2008; Magliulo *et al.*, 2009). In the study, z_k is an overloaded notation, since, it represents the ID of a zone ad its geometry as well.

$RL = \{l_i (I = 1, 2, \ldots ,)\}$, where l_i denotes a railway line that crosses the GeoArea. Each line is delimited by two end points that are identified with the railway stations of departure and arrival. Each train line in RL is defined as the tuple <ID, name, geometry of l_i>, being ID an identifying code and geometry of l_i the set of pairs of coordinates describing the shape of l_i. Card (RL) denotes the cardinality of set RL that is the number of train lines in GeoArea. A train line may be modeled as a simple line that is as a curve with two disconnected boundaries that does not pass through the same point more than once (Ryden, 2005). Δl_i denotes the boundary of line l_i, namely both its end points while l_i° denotes the interior of l_i. Hereafter, l_i is an overloaded notation, since, it represents the ID of a railway line ad its geometry as well.

GC is a two dimensional matrix of sets with card (RL) rows and card (Z) columns. Element $GC [i, k] = \{h_{ik,j} (j = 1, 2, \ldots ,)\}$ where $h_{ik,j}$ denotes a Hotspot, i.e., the n-th component of the geometry collection result of the intersection between the railway line l_i (of RL) and the zone z_k (in Z). Card (GC [i, k]) with k = 1, 2, ... , card (Z), denotes the number of Hotspots along the railway line l_i. Hotspot $h_{ik,j}$ is fully defined by the tuple <ID, geometry of $h_{ik,j}$, $Exp\_h_{ik,j}$>, being ID an identifying code, geometry of $h_{ik,j}$ the geometry of the location of Hotspot $h_{ik,j}$ and $Exp\_h_{ik,j}$ a positive numeric value denoting the degree of (spatial) exposure of $h_{ik,j}$ to the landaslide hazard. Hereafter, $h_{ik,j}$ is an overloaded notation, since, it represents the ID of a Hotspot ad its geometry as well.

Figure 1 shows a scene about a GeoArea (the big rectangle), a generic railway line (l_i) crossing it and four zones without holes partitioning the GeoArea (i.e., $Z = \{z_k (k = 1, \ldots , 4)\}$). The zones are depicted as rectangles. Darker rectangles denote zones with a bigger value of Sz_k. In the figure, $Sz_1 < Sz_3 < Sz_4 < Sz_2$. The intersection between l_i and the zones in Z gives rise to five Hotspots along line l_i: $\{h_{i1,1}, h_{i1,2}, h_{i2,1}, h_{i2,2}, h_{i4,1}\}$.

Detection of the railway line's Hohspots and computation of their exposure: In order to implement a policy of selective monitoring about the state of health of the land on which are leaned the tracks of the railway lines that

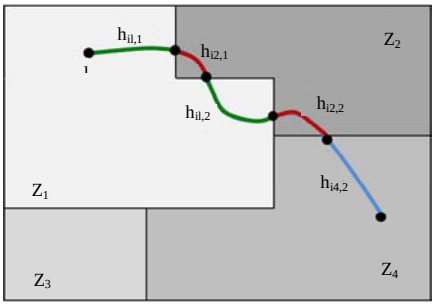

Fig. 1: A scene depicting previous definitions

cross the GeoArea, it is required that for each l_i in RL the geometry of the Hotspots ($h_{ik,j}$) internal to the line is returned, together with the value of their exposure (i.e., $Exp\_h_{ik,j}$). The higher is the value of parameter $Exp\_h_{ik,j}$ and bigger is the l andaslide hazard faced by the ground underneath the stretch of the railway line which coincides with the Hotspot $h_{ik,j}$ and in the final analysis, the greater is the risk of derailment of the train when it crosses such a portion of territory. The value of the exposure of a generic Hotspot ($h_{ik,j}$) along the railway line l_i is given by Eq. 1:

$$Exp\_h_{ik,j} = Size \times Sz_k \quad (1)$$

where, size is given by Eq. 2, being N = card (Z):

$$Size = \frac{\text{Area of } z_k}{\left(\sum_{j=1}^{N} \text{area of } z_j\right)/N} \quad (2)$$

Both equations come from a method proposed by Felice (2014) to compute the ranking of buildings according to their exposure to the landslide hazard. To identify the Hotspots in the Railway Line l_i (belonging to RL) it is necessary to calculate the intersection of l_i with the Zones in Z. Given the pair (l_i, z_k), two cases are possible:

$$l_i \cap z_k \neq \varnothing (\text{case1}) \text{or} l_i \cap z_k = \varnothing (\text{case2})$$

Below case 1 is investigated, since, it is the only one that can give rise to Hotspots. The goal is to introduce a way to refine the "Base" value of the exposure attached by Eq. 1 to the Hotspots located along the railway lines. The method that will be proposed makes use of results about the topological relations between a pair of geographic features. In the study, the geographic features taken into account are pairs (l_i, z_k) that is a line and an area. Topological relations between a pair of geographic features are characterized by the topological invariants of the, so-called 9-intersection model (Egenhofer and Herring, 1991) which is concisely represented as a 3 times 3 matrix (Eq. 3) where °Δ and denote in order, the interior, the boundary and the exterior of a geographic feature:

$$\begin{pmatrix} l_i^{\circ} \cap z_k^{\circ} & l_i^{\circ} \cap \Delta z_k & l_i^{\circ} \cap z_k^{-} \\ \Delta l_i \cap z_k^{\circ} & \Delta l_i \cap \Delta z_k & \Delta l_i \cap z_k^{-} \\ l_i^{-} \cap z_k^{-} & l_i^{-} \cap \Delta z_k & l_i^{-} \cap z_k^{-} \end{pmatrix} \quad (3)$$

Relevant topological invariants applicable to the 9-intersection model are the content (i.e., emptiness or non-emptiness) of a set, the number of separations in the intersection (i.e., the number of disconnected components) and the dimension of each "separation". By separation it is meant the geometry returned by each of the nine intersections of Eq. 3. Hereafter, ($l_i \cap z_k$) denotes the geometry collection result of the intersection between l_i and z_k while No. ($l_i \cap z_k$) denotes the number of separations of such an intersection. In our study, the dimension of an intersection may be either 0 or 1.

Case 1 will be treated by taking into account the three invariants mentioned above between l_i and z_k. In addition when possible, we will take into account two metrics, the Inner Area Splitting (IAS) and the Entrance Splitting (ENS) (introduced by Egenhofer and Shariff by (Egenhofer and Shariff, 1998)) to capture relevant geometric details from the topological relations between l_i and z_k.

The content invariant: The relations/configurations between a simple line and a simple polygon without holes are 19 (Egenhofer and Herring, 1991) (Fig. 2). But since, a zone can have holes then it is immediate to draw as many configurations starting from those of Fig. 2 simply by introducing (at least) a hole inside the polygon.

However, in the problem at hand, the meaningful configurations between l_i and z_k are a restricted subset because the area of z_k is modest when compared to the extension of l_i that conversely is several 10 km. It follows that it is impossible that a railway line is wholly contained inside a zone (therefore, configurations 8-12 and 13 of Fig. 2 are to be excluded) or that it has both end points in the same zone (which excludes configurations 2, 4, 6, 15, 17 and 19 of Fig. 2). Therefore, when we examine the content invariant, the configurations of Fig. 2 that are significant for our study are: 3, 5, 7, 14, 16 and 18. Configuration 1 must be ignored because it does not determine Hotspots. As for configurations 3, 5, 7, 14, 16 and 18 of Fig. 2 the matrix of the 9-intersection has the structure of Eq. 4:

$$\begin{pmatrix} ? & ? & \neg\varnothing \\ ? & ? & \neg\varnothing \\ \neg\varnothing & \neg\varnothing & \neg\varnothing \end{pmatrix} \quad (4)$$

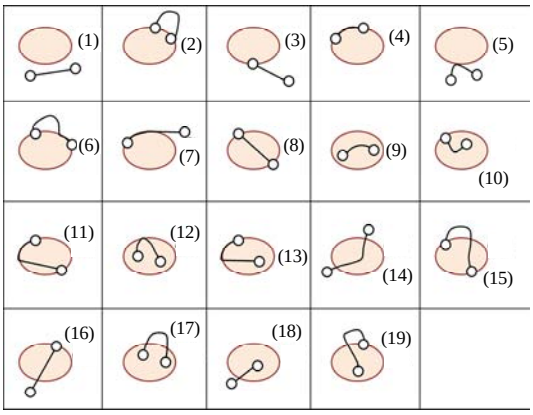

Fig. 2: The geometric configurations realizable between a simple line and a simple polygon without holes in 2D according to the 9-intersection model

It can be concluded that those intersections are distinguishable by the four intersection's values of Eq. 5:

$$\begin{pmatrix} ^\circ\cap{}^\circ & ^\circ\cap\Delta \\ \Delta\cap{}^\circ & \Delta\cap\Delta \end{pmatrix} \quad (5)$$

Number of separations: The configurations of Fig. 3 are characterized by the same content invariant (Eq. 3) of the corresponding scenes of Fig. 2 but they have a different number of disconnected components. Note that the scenes of Fig. 3 are not the only one possible in fact the number of components can be as large as desired.

Dimension of the separations: The five scenes of Fig. 4 show geometric configurations having the same content invariant and the same number of separations according to the 9-intersection model of the correspondent scenes of Fig. 3 but they differ from the latter for the dimension of the partitions. Table 1 shows an example.

Given the objective of our study, it is essential to grasp the details of the geometry that involves an entire railway line (l_i) and each zone (z_k) in Z. In fact, to be able to return to end-users detailed information about the Hotspots present along the railway line l_i it is necessary the knowledge of: the number of separations, their dimension and their exact location. Among the six configurations of Fig. 5 taken into account in this study (i.e., 3, 5, 7, 14, 16) and 18, 14, 16 and 18 meet in order, the three patterns of Eq. 6:

Therefore, it is possible to operate a closer examination about those geometric scenes by calculating the value of the IAS metric whose definition is as follows:

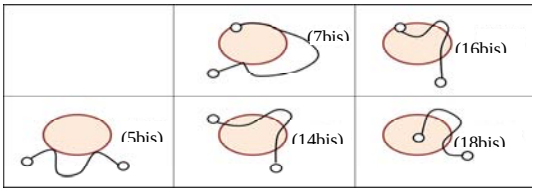

Fig. 3: A refinement of the corresponding scenes of Fig. 2

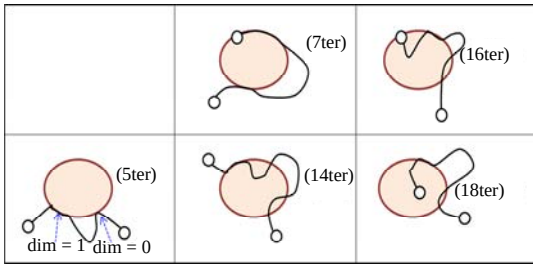

Fig. 4: Geometric variants of the scenes of Fig. 3

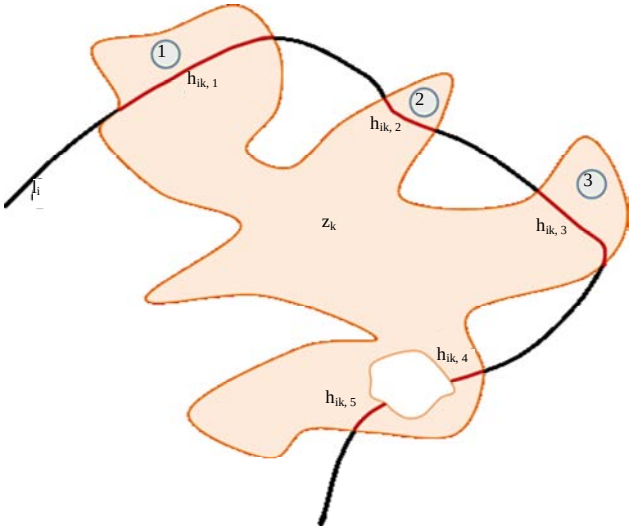

Fig. 5: A zone (z_k) with a hole (the white region). $IAS_{ik,1}$, $IAS_{ik,2}$ and $IAS_{ik,3}$ are >0 while $IAS_{ik,4} = IAS_{ik,5} = 0$

Table 1: A comparison between the scenes 5 (Fig. 2) 5bis (Fig. 3) and 5ter (Fig. 4)

Variables	(5)	(5 bis)	(5 ter)
$\Delta z_k \cap l_i^\circ$	$\neg\phi$	$\neg\phi$	$\neg\phi$
No. $(\Delta z_k \cap l_i^\circ)$	1	2	2
Separation's dimension	0	{0, 0}	{1, 0}

$$\begin{pmatrix} ^\circ\cap{}^\circ=\neg\varnothing & ^\circ\cap\Delta=? \\ \Delta\cap{}^\circ=\varnothing & \Delta\cap\Delta=\varnothing \end{pmatrix} \begin{pmatrix} ^\circ\cap{}^\circ=\neg\varnothing & ^\circ\cap\Delta=\neg\varnothing \\ \Delta\cap{}^\circ=\varnothing & \Delta\cap\Delta=\neg\varnothing \end{pmatrix}$$

$$\begin{pmatrix} ^\circ\cap{}^\circ=\neg\varnothing & ^\circ\cap\Delta=\neg\varnothing \\ \Delta\cap{}^\circ=\neg\varnothing & \Delta\cap\Delta=\varnothing \end{pmatrix}$$

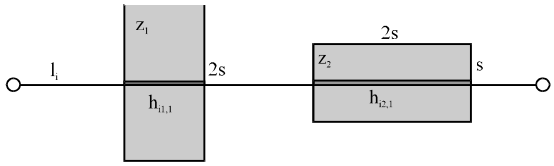

Fig. 6: Three simple geometries

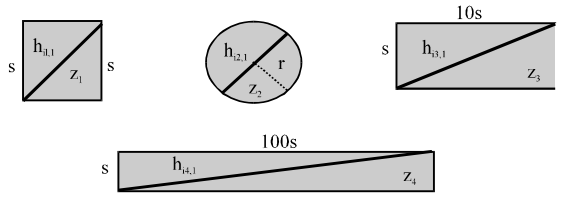

Fig. 7: Two possible instances of configuration 14 of Fig. 2

$$IAS = \frac{\min(\text{area}(\text{left separation}(l_i^\circ \cap z_k^\circ)), \text{area}(\text{right separation}(l_i^\circ \cap z_k^\circ)))}{\text{area}(Z_k)} \quad (7)$$

IAS describes how the line's interior (in our case the interior of a railway line) divides the interior of a region (in our case the interior of a zone). With this separation, line l_i splits area z_k into two or more parts such that parts of the area's interior are on one side of the line and others are located on the opposite side of the line. From Eq. 7 it follows that $0<IAS\leq 0.5$. In the general case in which l_i splits area z_k in several parts (Fig. 5 shows an example), each intersection identifies a hotspot. To assign them an appropriate value of exposure, it is necessary to attach to them a value of the IAS metric. This is achieved by applying Eq. 7 to each portion of land identified by the intersection between l_i and z_k whose area has to be compared to the total area of region z_k. With regard to the example of Fig. 5, we have four values of IAS as many as are the Hotspots.

Note that, Eq. 7 may not be applicable to some H otspots result of the intersection between a railway line and a zone with holes. This is the case of Hotspots $h_{ik,4}$ and $h_{ik,5}$ of Fig. 5 to which we assign, conventionally, IAS = 0.

Downstream of the calculation of the value of the metric IAS, the correction to be made to the basic value of the exposure of the corresponding Hotspot is expressed by Eq. 8. Such an equation retains the value of parameter Exp_h_{ikj} if $IAS_{ikj} = 0.5$ while such a value is decreased up to the 50% if IAS_{ikj} tends to zero:

$$Exp\_h_{ik,j} = Exp\_h_{ikj} \times [1-(0.5-IAS_{ik,j})] \quad (8)$$

For configurations 14, 16 and 18 we can operate a further close examination of the geometry of the scene by calcu Hotspots lating the value of metric ENS whose definition is given by Eq. 9:

$$ENS = \frac{\text{length}(l_i^\circ \cap Z_k^\circ)}{\text{length}(\Delta Z_k)} \quad (9)$$

ENS applies to relations in which the line's interior crosses the region's interior. The presence of holes inside a zone does not make Eq. 9 inapplicable to the resulting Hotspots. Therefore, for example, ENS>0 for the 5 H otspots in Fig. 5. Equation 9 normalizes the common interiors of the line and the area with respect to the length of the region's boundary. ENS>0 while the upper bound is undefined from a theoretical point of view. However, for a given case study (i.e., given the pair of sets RL and Z) it is possible to compute the value of the upper bound by applying Eq. 9 to all the pairs (l_i, z_k). For the time being we quantify the upper bound of ENS through simple arguments. Specifically, let us refer to three different simple geometries: a square, a rectangle and a circle, through which we can make a rough abstraction of the shape of the zones. The square and the circle approximate zones with compact form while the rectangle approximates zones with long and narrow shape.

Figure 6 presents four different geometric scenes. Each of them is composed of a zone and its hotspot, result of the intersection between the zone and the railway line l_i. The values of metric ENS for the four geometric scenes of Fig. 6 are the following:

$$ENS_{i1,1} = \frac{1.41\times s}{4\times s}\times 0.352; \quad ENS_{i2,1} = \frac{2\times r}{2\times \pi \times r} = 0.318$$

$$ENS_{i3,1} = \frac{10.05\times s}{22\times S} = 0.457; \quad ENS_{i4,1} = \frac{100\times s}{202\times s} = 0.495$$

It can be concluded that for the problem at hand is correct to assume that $0<ENS<0.5$. Downstream of the computation of the value of ENS, the correction to be made to the value of the exposure of the corresponding hotspot is expressed by Eq. 10. It retains the value of parameter Exp_h_{ikj} if $ENS_{ikj} = 0.5$ while it decreases such a value up to the 50% if ENS_{ikj} tends to zero:

$$Exp\_h_{ik,j} = Exp\_h_{ik,j} \times [1-(0.5-ENS_{ik,j})] \quad (10)$$

The use of ENS captures metric information sufficient to distinguish geometric configurations that on the opposite are identical with respect to the three topological invariants taken into account in this study as well as with

respect to the value of metric IAS. Let us consider, for example, the scene of Fig. 7. It proposes a railway line (l_i) that intersects in the middle two zones (z_1 and z_2).

The assumptions that describe the scene of Fig. 7 are the following: the zones z_1 and z_2 have identical value of probability of triggering lansdslides (i.e., $Sz_1 = Sz_2$), moreover "area of z_1 = area di $z_2 = 2x\ s^2$". The intersection between the railway line l_i and the zones z_1 and z_2 gives rise to two Hotspots ($h_{i1,1}$ and $h_{i2,1}$). Applying Eq. 1 to them, it is obtained that $Exp\_h_{li,1} = Exp\_h_{i2,1} = E$. Since, $IAS_{i1,1} = IAS_{i2,1}$ such a metric does not allow to differentiate the ranking of the two Hotspots. Vice versa, $ENS_{i1,1} <> ENS_{i2,1}$, in fact:

$$ENS_{i1,1} = \frac{lenght(l_i^{\circ} \cap z_1^{\circ})}{lenght(\Delta z_1)} = \frac{s}{6s} = 0.1666$$

$$ENS_{i2,1} = \frac{lenght(l_i^{\circ} \cap z_2^{\circ})}{lenght(\Delta z_2)} = \frac{2s}{6s} = 0.333$$

By applying Eq. 10 it follows that $Exp\_h_{i1,1} = E \times 0.666$ while $Exp\_h_{i2,1} = E \times 0.833$ that is in the final analysis, we have that $Exp\_h_{i2,1} > Exp\_h_{i1,1}$. The fact that the exposure of the Hotspot $h_{i2,1}$ exceeds the value of the exposure of the Hotspot $h_{i1,1}$ determines that the first precedes the second in the final ranking and this is correct given that for the same geometric configuration, length of $h_{i2,1}$ = 2x length di $h_{i1,1}$. Eq. 11 merges Eq. 8 and 10:

$$Exp\_h_{ik,j} = Exp\_h_{ik,j} \times [1-(0.5-IAS_{ik,j})] \times [1-(0.5-ENS_{ik,j})] \quad (11)$$

From the ranking of the Hotspots to the selective monitoring of the railway lines: The intersection between the railway lines in RL and the zones in Z returns card (RL)×card (Z) Hotspots. Usually, this number is very high. To identify the Hotspots that primarily require monitoring it is necessary to build the 2+card (RL) tables described in this section.

The values of the exposure assigned to the Hotspots by Eq. 11 oscillate between a maximum and a minimum, the latter being close to zero. These values can be aggregated into an arbitrary number of ranges in order to structure the first of the 2+card (RL) tables. We call it "frame of reference" table. This name originates from the fact that to decide, for a given railway line in RL, the top-N list of Hotspots to be monitored with the highest priority it is necessary to refer to this table. We will return to this point shortly. The data in the frame of reference table also, allow to gain an idea about the overall situation of the bed of the railway lines that cross the GeoArea.

Table 2: The frame of reference table

Hazard	Values of exposure	#Hotspots
Very high	Exp>3×AVG (total exposure)	v_1
High	1.5×AVG (total exposure) <Exp = 3×AVG (total exposure)	v_2
Moderate	0.5×AVG (total exposure) <Exp< = 1.5×AVG (total exposure)	v_3
Low	0.25×AVG (total exposure) <Exp<= 0.5×AVG (total exposure)	v_4
Null	Exp< = 0.25×AVG (total exposure)	v_5

Our proposal is to adopt five ranges of hazard (very high, high, moderate, low and null) as defined in Table 2. In it AVG (total exposure) denotes the arithmetic mean of the exposure values of the card (RL)×card (Z) identified Hotspots while #Hotspots denotes the total number of Hotspots whose exposure value (v_i, i = , 1, … , 5) falls in the corresponding 5 ranges.

The second table to be constructed consists of card (RL) rows and 6 columns (Table 3). The i-th row of such a table shows in order, the total number of Hotspots located along railway line l_i and to follow, the total number of Hotspots whose exposure value falls in each of the 5 ranges of Table 2. The remaining card (RL) tables have all the same structure. The i-th of them is composed of v_{li} rows and 4 columns (Table 4). These latter denote in sequence: the ID of the hotspot, its exposure value (according to Eq. 11) and the coordinates of its two end points (called end point A and end point B in the table). The coordinates are fundamental to locate on the field the Hotspots to be monitored.

A crucial issue is how to identify the top-N Hotspots (along line l_i in RL) to be flagged as priorities for mitigating the risks of derailments. Since, the value of "N" is not known in advance, nor it is constant for all the railway lines, to decide whether a Hotspot is to be monitored or not we suggest that it is ascertained in which of the five ranges of Table 2 drops its value of exposure. In this way there is an objective assessment of the level of hazard which weighs on the Hotspot under consideration based on the comparison of its exposure value with all the computed values. The value of N is defined as follows: N is equal to the total number of Hotspots along line l_i such that $Exp\_h_{ik,j}$ falls either in the very high range or in the high range (Table 2). Referring to the notations in Table 3, we can write that for line: l_i, $N = v_{li,1}+v_{li,2}$. Definition above emphasizes the centrality of Table 2 as a frame of reference. Because Eq. 11 does not take into account the elevation of the terrain inside the GeoArea, the proposed method may return false positives. In other words, we can say that each of the Hotspots in the top-N list may be a false positive until it

Table 3: The distribution of the Hotspots within the five ranges of the frame of reference table for the railway lines in RL

Railway lines	No. of Hotspots	No. of Hotspots in the very high range	No. of Hotspots in the high range	No. of Hotspots in the moderate range	No. of Hotspots in the low range	No. of Hotspots in the null range
l_1	v_{l1}	$v_{l1,1}$	$v_{l1,2}$	$v_{l1,3}$	$v_{l1,4}$	$v_{l1,5}$
l_2	v_{l2}	$v_{l2,1}$	$v_{l2,2}$	$v_{l2,3}$	$v_{l2,4}$	$v_{l2,5}$
, … ,	, … ,	, … ,	, … ,	, … ,	, … ,	, … ,
$l_{card(RL)}$	$v_{lcard(RL)}$	$v_{lcard(RL),1}$	$v_{lcard(RL),2}$	$v_{lcard(RL),3}$	$v_{lcard(RL),4}$	$v_{lcard(RL),5}$

Table 4: The structure of the table about all the Hotspots along line l_i

Hotspot's ID	Hotspot's exposure	End point A	End point B
1		Latitude longitude	Latitude longitude
…	…	…	…
v_{li}		Latitude longitude	Latitude longitude

Table 5: The railway lines that cross the Abruzzo Region and their length

Railway name	Length (km)
Bologna-Bari	124
Ortona-Crocetta	36
Marina di San Vito-Castel di Sangro	103
Roma-Pescara	170
Avezzano-Roccasecca	46
Archi stazione-Atessa	15
Sulmona-Carpinone	85
Rieti-L, Aquila-Sulmona	81
Teramo-Giulianova	25

is ascertained the opposite. The easiest and fastest way to detect the false positive Hotspots (belonging to a given railway line) and remove them from the top-N list consists in looking at each of them against the DEM of the underlying terrain by making recourse to a geo-viewer.

A case study

Input datasets

GeoArea: GeoArea coincides with the boundary of the Abruzzo region; An area of about 11,000 km^2, structured as four provinces, 305 municipalities and a population of about 1,330,000.

Set Z: For the Abruzzo Region, it is not available a dataset with the characteristics of set Z. What we have found is a shapefile about the landslide inventory of the region (a landslide inventory is a register about the distribution and characteristics of past landslides (Bobrowsky, 2013). Landslide inventories are often used by scholars. For instance (Mandal and Maiti, 2013) used landslide inventory statistics to investigate the relationship between rainfall and landslip events.). The limit of this dataset is that it does not achieve a complete partition of the region. This (real) dataset coincides with the theoretical one by setting $Sz_k = 0$ for the portions of land not surveyed.

Within the landslide inventory, landslides are classified according to the type of movement, the estimated age, the state of activity, the depth of failure surface and the velocity. The categories of landslides making part of the inventory are fall/topple, rotational/transational slide, slow eart flow, rapid debris flow, sinkhole, complex landslide. Moreover, landslides are classified as active, quiescent/dormant and inactive.

The elements contained in the Abruzzo landslide inventory are grouped into three susceptibility classes called S1 (low susceptibility), S2 (high susceptibility) and S3 (very high susceptibility). Overall, the inventory is composed of 4,425 elements in S1, 8,886 elements in S2 and 3,959 elements in S3. With few exceptions, it can be said the following: S3 includes active landslides, quiescent landslides are in S2 while inactive landslides are in S1. The area of the landslides in Z ranges from 161-6,302,190 m^2. The average area measures 93,653 m^2.

Values of the parameters used in the experiments: As just said, the case study partitions the set of zones Z into three classes each characterized by a single value for Sz_k for all the z_k belonging to it. This is a simplification of the general case which does not exclude that each z_k has a specific value of Sz_k. In order to carry out the experiments, we have associated to the zones in the aforementioned three classes the probability values 0.25, 0.5 and 1, respectively. Obiously those values may be changed but the following constraint must be fulfilled: the value of Sz_k must be positive (because it expresses a probability value) and such that Sz_k (class S1)<Sz_k (class S2)<Sz_k (class S3).

Set RL: In Abruzzo there are nine railway lines (Table 5), some completely internal to the region while others cross it (having origin and/or end outside the borders of region). Regarding the latter in the study only the portions of railways internal to the Abruzzo region were considered. The 9 lines all together add up to 685 km.

The geographical database: We implemented a PostgreSQL/PostGIS Geographical DataBase (shortly, Geo-DB) to store the records of the input shapefiles and the results of the experiments to be carried out. The relevance of Geo-DBs as a tool for the management of natural hazards is well-known in the literature, see for instance (Blahut *et al*., 2012; Rawat *et al*., 2012).

The Geo-DB is composed of four tables: GeoArea, zones, railway routes and Hotspots. The appendix shows the entity-relationship diagram of the Geo-DB, moreover it lists the Postgres/PostGIS SQL/DDL commands that define those tables. Tables GeoArea, zones and railway routes store data about sets GeoArea, Z and RL, respectively. Table Hotspots collects the results of the experiments to be carried out (the two dimensional matrix GC).

The Geo-DB has been enriched with a large number of PL/pgSQL user defined functions and SQL views through which it was possible to carry out complex processing by writing simple SQL queries.

RESULTS AND DISCUSSION

This study reports about the results of the experiments we carried out, at three different levels of analysis. The first level concerns the construction of Table 2 and 3. The second level is about the top-N list of the Hotspots along one of the railway lines of the A-Bruzzo Region. Lastly, we focus on a geometric configuration, similar to many other that emerged from the experiments, suitable to prove the effectiveness of the addition of the metrics IAS and ENS in the computation of the value of the exposure.

First level of analysis: The intersection between the 9 Abruzzo railway lines and the 17,312 zones in Z returned 210 Hotspots with exposure value>0. They add up to 63.7 km. Hotspots with exposure value equal to zero are ignored because they are not to be monitored as not exposed to the landslide danger. The exposure value assigned to each Hotspot by Eq. 11 ranges between 4.76 and 0. The exposure is null when $Sz_k = 0$.

Table 6 shows the frame of reference table for the case study. Compared to Table 2 and 6 also shows, for each range, the percentage of the totality of the Hotspots belonging to it (forth column) the total length (in km) of all the Hotspots that are part of the same range (column length) and the percentage value of their extension (% length column) referred to the total length of the 9 railway lines of the Abruzzo Region.

Table 7 shows the distribution of the Hotspots within the 5 ranges of the frame of reference table for the railway lines of the Abruzzo Region. The railway lines "Teramo-Giulianova" and "Avezzano-Roccasecca" do not appear because they do not give rise to any hotspot with exposure value >0. Table 6 and 7 show that in the case study, the majority of the Hotspots have a low or null level of exposure to the landslide risk while 42 of them (20% of the totality of the Hotspots with value of exposure >0) present a level of exposure that brings out the need of more frequent controls (very high and high exposure level). The total extent of the just mentioned 42 Hotspots (24.5 km) is equal to just the 3.57% of the total length of the 9 railway lines of the Abruzzo Region and to the 38.46% of the total extension of the 210 Hotspots. These numbers describe a state of the land on which run the tracks of the 9 railway lines that cross the Abruzzo region broadly stable and this results in the conclusion that the monitoring actions to be taken on the field are numerically limited.

Second level of analysis: Table 8 shows the top-N Hotspots along the railway line "Marina di San Vito-Castel di Sangro". Such a line is long 103 km and has 77 Hotspots. The color is used to link the Hotspots to the range they belong to, according to the schema of Table 2. For such a railway line N = 14. The value of N is determined by linking the data in the second column of Table 8 to those in the second column of Table 6 that as mentioned, acts as a frame of reference. The top-N list of the Hotspots for the other lines can be determined similarly.

Third level of analysis: We refer to the geometric configuration of Fig. 8, since, it is suitable to highlight the effects introduced by the metrics IAS and ENS on the calculation of the value of the exposure parameter.

Table 9 lists the Hotspots result of the intersection of the railway line "Roma-Pescara" with the zone having id = 479 and $Sz_k = 0.25$. The exposure values in Table 9 were obtained by applying Eq. 1. These values are not satisfactory because they are identical despite the four Hotspots have different length (it ranges from 110-480 m) and different degree of crossing of the zone (Fig. 8).

Table 10 shows the effects caused by the application of Eq. 11 in the calculation of the exposure of these Hotspots. As can be seen putting in comparison the Hotspots with id = 144 and id = 145, their exposure values changed from 2.5 (when Eq. 1 is used) to 0.83 and 0.65, respectively (when Eq. 11 is used). The summary is that through the introduction of the metrics IAS and ENS we got a double corrective effect of the "base" value of the exposure returned by Eq. 1. The first corrective effect (attributable to the ENS metric, Eq. 10) is represented by

Table 6: An enhanced version of the frame of reference table for the case study

Hazard	Exposure	No. of Hotspots	Percentage of the Hotspots	Length (km)	Length (%)
Very high	[1.83..4.76]	19	9.05	14.9	2.17
High	[0.88..1.83)	23	10.95	9.6	1.4
Moderate	[0.32..0.88)	53	25.24	20.5	3.0
Low	[0.15..0.32)	47	22.38	10.8	1.58
Null	(0..0.15)	68	32.38	7.9	1.15
		210		63.7	

Table 7: The distribution of the hotspots within the ranges of the frame of reference table for the railway lines in RL

Railway lines	No. of Hotspots	No. of Hotspots in the very high range	No. of Hotspots in the high range	No. of Hotspots in the moderate range	No. of Hotspots in thelow range	No. of Hotspots in the null range
Archi Stazione-Atessa	5	0	0	0	1	4
Bologna-Bari	38	2	1	7	10	18
Marina di San Vito-Castel di Sangro	77	9	5	24	16	23
Ortona-Crocetta	3	1	0	2	0	0
Rieti-lAquila-Sulmona	12	0	0	2	3	7
Roma-Pescara	27	1	7	11	2	6
Sulmona-Carpinone	48	5	9	8	12	14

Table 8: The top-N Hotspots along line "Marina di San Vito-Castel di Sangro"

Hotspot	Exposure	Point A	Point B
1	4.76	42°11'24.598"N 14°21'23.816"E	42°11'46.928"N 14°20'58.002"E
2	3.73	41°55'23.097"N 14°17'22.451"E	41°55'6.691"N 14°17'22.150"E
3	3.56	41°52'57.975"N 14°14'39.255"E	41°52'39.796"N 14°13'48.324"E
4	3.47	41°54'44.714"N 14°17'11.585"E	41°54'39.120"N 14°17'0.727"E
5	3.41	42°2'27.055"N 14°20'52.980"E	42°1'51.642"N 14°20'54.252"E
6	2.75	41°58'50.261"N 14°22'18.812"E	41°58'43.730"N 14°22'18.515"E
7	2.73	42°1'20.307"N 14°21'17.746"E	42°0'58.483"N 14°21'34.270"E
8	2.62	41°58'55.655"N 14°22'18.294"E	41°58'52.316"N 14°22'18.410"E
9	2.48	41°58'10.316"N 14°22'14.014"E	41°57'54.260"N 14°22'2.956"E
10	1.11	41°55'52.901"N 14°18'49.777"E	41°55'58.620"N 14°18'32.945"E
11	1.10	42°11'32.235"N 14°19'47.938"E	42°11'16.294"N 14°19'56.344"E
12	0.99	42°12'39.863"N 14°23'2.996"E	42°12'38.618"N 14°23'2.477"E
13	0.96	42°1'43.719"N 14°21'5.255"E	"42°1'35.123"N 14°21'8.763"E
14	0.94	41°56'46.294"N 14°21'25.414"E	41°56'40.327"N 14°21'21.419"E

Table 9: The Hotspots result of the intersection of zone 479 and the railway line "Roma-Pescara"

Hotspot's ID	Hotspot's exposure	Hotspot's length (m)
144	2.5	480
145	2.5	110
146	2.5	158
147	2.5	236

Table 10: Eq.11 vs. Eq.1

Hotspot's ID	Exposure without metrics (Eq. 1)	Exposure with metrics (Eq. 11)	IAS	ENS
144	2.5	0.83	0.103	0.049
145	2.5	0.65	0.006	0.012
146	2.5	0.66	0.015	0.017
147	2.5	0.79	0.102	0.025

the distinction that Eq. 10 does about the values of the exposure to be attributed to the two Hotspots according to their extension (480 vs. 110 m). The second corrective effect (attributable to both metrics, Eq. 11) is represented by the strong reduction of the exposure value with respect to the "base" value returned by Eq. 1 as a consequence of the low values of the metrics IAS and ENS.

Fig. 8: The map about the Hotspots result of the intersection of railway line "Roma-Pescara" and the zone with id = 479

Results validation: As already said, Eq. 11 does not take into account the terrain elevation inside the GeoArea that is why some of the Hotspots in the top-N list of a given railway line may be a false positive. Through Quantum GIS, we found for the case study, 9 false positives

Table 11: The number of false positive Hotspots in the top-N list, for the railway lines of the Abruzzo Region

Railway lines	No. of Hotspots in the very high range	No. of Hotspots in the high range	No. of Hotspots in the moderate range	No. of Hotspots in thelow range	No. of Hotspots in the null range
Archi stazione Atessa	0	0	0	0	0
Bologna-Bari	2	0	1	0	0
Marina di San Vito-Castel di Sangro	9	1	5	1	2
Ortona-Crocetta	1	0	0	0	0
Rieti-L'Aquila-Sulmona	0	0	0	00	
Roma-Pescara	1	1	7	2	3
Sulmona-Carpinone	5	1	9	3	4
Total	18	3	22	6	9

(Table 11) out of 40 top-N Hotspots present in the 9 railway lines crossing the Abruzzo Region (Table 7). The 3 false positives fall in the very high range while the remaining 6 fall in the high range. For all the 9 false positives, out of 40 Hotspots returned by our method, QGIS has shown that the terrain around such Hotspots is flat so, it is correct to exclude that they may be give rise to landslides even in the case of long periods of rain. Figure 9 shows the Entity-Relationship diagram of the Geo-DB of Sec.2.5. Hereafter, the Postgres/PostGIS SQL commands that define the four table of the Geo-DB are listed:

Algorithm 1; Geo-DB algorithm:

```
CREATE TABLE geo_area (
    description   character varying (30)        PRIMARY KEY
    geom          geometry(MultiPolygon, 3004)
)
CREATE TABLE zones (
    id            integer serial                PRIMARY KEY
    Szk           integer
    geom          geometry(Polygon, 3004)
)
CREATE TABLE railway_routes (
    id            integer serial                PRIMARY KEY,
    name          character varying(50),
    geom          geometry(MultiLineString, 3004)
)
CREATE TABLE Hotspots (
    id            integer serial                PRIMARY KEY
    id_route      integer
    id_dz         integer
    geom          geometry(Geometry, 3004)
    ias           double precision
    ens           double precision
    exposure      double precision
CONSTRAINT foreign_dz FOREIGN KEY (id_dz)
          REFERENCES zones (id) MATCH SIMPLE
    ON UPDATE NO ACTION        ON DELETE NO ACTION,
CONSTRAINT foreign_route FOREIGN KEY (id_route)
          REFERENCES railway_routes (id)  MATCH SIMPLE
    ON UPDATE NO ACTION        ON DELETE NO ACTION
)
```

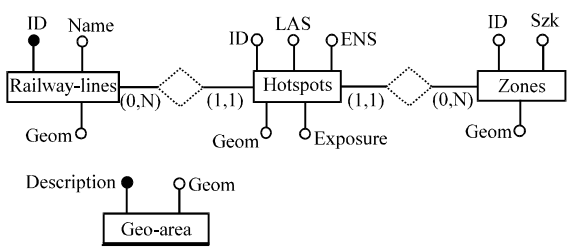

Fig. 9: The conceptual schema of the implemented Geo-DB

As Spatial Reference System we used EPSG:3004 (Monte Mario/Italy zone 2) which best models the Italian territory

CONCLUSION

This study gives two contributions. Firstly, it proposes a scientifically robust method for the identification of the top-N list of railway Hotspots with respect to their exposure to soil mass movements. Then, we sketched a strategy for implementing a selective monitoring of the soil nearby the identified Hotspots, thus, reducing the maintenance costs without lowering the railway safety level with respect to the case where the full length of the railway line has to be inspected. The method proposed for the identification of the Hotspots has its basis in consolidated results in the geographic information science domain. To our knowledge this is the first time that such a point of view is adopted for the problem at hand. The identification of the Hotspots along railway lines besides allowing the implementation of a selective monitoring of the lines, moreover allows the quantification of the risk which the railway lines are exposed, considering as "Elements at risk" not the whole lines but just the top-N Hotspots along them. The results returned by Eq. 11 (for a given case study) usually remain stable over periods of several years, unless either it does not change the geometry of some strecth of the railway lines or updates of the data set Z are made available. Both these circumstances are not very frequent in the reality.

In summary, we can say the following about the proposed method. Despite it is simple to be understood and implemented, it provides useful information to limit the inspection of the whole railway lines to a limited number of stretches along them.

The detection of the potential false positives present in the top-N is simple and rapid. Anyway, it is better to have false positives than false negatives.

RECOMMENDATIONS

The current proposal is just the preliminary step towards the release of a software tool to be meant for the people responsible of the control of the safety of the railway beds within any territory (province, region, nation). The next step in our agenda will be devoted to divise an algorithm that takes into account the terrain elevation in the study area in order to eliminate the possibility of detecting false positives.

REFERENCES

Blahut, J., I. Poretti, M.D. Amicis and S. Sterlacchini, 2012. Database of geo-hydrological disasters for civil protection purposes. Nat. Hazards, 60: 1065-1083.

Bobrowsky, P.T., 2013. Encyclopedia of Natural Hazards. Springer, Netherlands, ISBN:798-90-481-8699-0, Pages: 1135.

Brabb, E.E. and B.L. Harrod, 1989. Landslides: Extent and Economic Significance. Balkema Publisher, Rotterdam, Netherlands, Pages: 385.

Egenhofer, M.J. and A.R. Shariff, 1998. Metric details for natural-language spatial relations. ACM Trans. Inform. Syst., 16: 295-321.

Egenhofer, M.J. and J.R. Herring, 1991. Categorizing binary topological relations between regions, lines and points in geographic databases. Technical Report, Department of Surveying Engineering, University of Maine, Orono, ME.

Felice, P.D., 2014. On the development of a software suitable to support natural area hikers in the selection of paths. Asian J. Inf. Technol., 13: 770-776.

Fell, R., J. Corominas, C. Bonnard, L. Cascini and E. Leroi *et al.*, 2008. Guidelines for landslide susceptibility, hazard and risk zoning for land-use planning. Eng. Geol., 102: 99-111.

Guzzetti, F., P. Reichenbach, F. Ardizzone, M. Cardinali and M. Galli, 2006. Estimating the quality of landslide susceptibility models. Geomorphol., 81: 166-184.

He, H., S. Li, H. Sun and T. Yang, 2011. Environmental factors of road slope stability in mountain area using principal component analysis and hierarchy cluster. Environ. Earth Sci., 62: 55-59.

Magliulo, P., A.D. Lisio and F. Russo, 2009. Comparison of GIS-based methodologies for the landslide susceptibility assessment. Geoinformatics, 13: 253-265.

Mandal, S. and R. Maiti, 2013. Assessing the triggering rainfall-induced landslip events in the shivkhola watershed of darjiling Himalaya, West Bengal. Eur. J. Geography, 4: 21-37.

RSSB, 2014. Guidance on hazard identification and classification. Rail Safety and Standards Board Limited, London, England. h ttps: // www. rssb. co. uk/ rgs/standards/GEGN8642%20Iss%202.pdf.

Rawat, P.K., P.C. Tiwari and C.C. Pant, 2012. Geo-hydrological database modeling for integrated multiple hazards and risk assessment in Lesser Himalaya: A GIS-based case study. Nat. Hazards, 62: 1233-1260.

Ryden, K., 2005. Open GISs implementation specification for geographic information-simple feature access part2: SQL option. Open Geospatial Consortium Inc, New York, USA.

Sadler, J., D. Griffin, A. Gilchrist, J. Austin and O. Kit *et al.*, 2016. GeoSRM: Online geospatial safety risk model for the GB rail network. IET. Intell. Transp. Syst., 10: 17-24.

SafeLand, 2011. Guidelines for landslide susceptibility, hazard and risk assessment and zoning. SafeLand, London, England.