

**MONITORAGGIO SELETTIVO DI STAZIONI E TRATTE
FERROVIARIE IN BASE ALLA LORO ESPOSIZIONE ALLA
PERICOLOSITÀ DI FRANA: UN APPROCCIO GIS**

**CARUSO FEDERICA
DI SABATINO FABIO
GIANCOLA DAVIDE**

INDICE

1	INTRODUZIONE	1
2	MATERIALS AND METHODS	4
2.1	Notazioni	4
2.2	Un metodo per classificare le stazioni ferroviarie	16
2.2.1	Metodo di calcolo originario	16
2.2.2	Metodo di calcolo ex novo	18
2.2.3	L'algoritmo	21
2.3	Un metodo per classificare le linee ferroviarie	28
2.3.1	Metodo di calcolo	28
2.3.2	L'algoritmo	30
3	UN CASO DI STUDIO: L'ABRUZZO	31
3.1	Territorio	31
3.2	Rete Ferroviaria	32
3.2.1	Stazioni Ferroviarie	32
3.2.2	Linee Ferroviarie	33
3.3	Zone	34
3.4	Basi di dati spaziale	35
3.4.1	Progettazione concettuale	35
3.4.2	Progettazione logica	39
4	RISULTATI E LORO DISCUSSIONE	40
4.1	Ranking e Valutazione delle stazioni ferroviarie	40
4.1.1	Analisi critica del Nuovo Metodo Proposto	41
4.1.2	Discussione dei risultati ottenuti	48
4.2	Ranking e Valutazione delle linee ferroviarie	49
4.2.1	Discussione dei risultati ottenuti	49
5	CONSIDERAZIONI FINALI E PROPOSTE PER LAVORI FUTURI	51
5.1	Proposte per lavori futuri sul metodo di ranking delle stazioni ferroviarie	51
5.2	Proposte per lavori futuri sul metodo di ranking delle tratte ferroviarie	51
5.3	Considerazioni sui dataset	51
5.4	Considerazioni finali	52
I	APPENDICI	54
A	APPENDICE A	55
A.1	Proposta alternativa di Dataset delle Zone	55
A.1.1	Processamento dei poligoni interni	56
A.1.2	Processamento dei poligoni lungo il bordo	57
A.1.3	Conclusioni	58
B	RISULTATI DI LABORATORIO	59
B.1	Classificazione completa delle stazioni tramite il NMC	59

B.2	Classificazione completa delle linee ferroviarie tramite il NMC	59
B.3	Classificazione ufficiale delle stazioni utilizzata per la validazione	59

ELENCO DELLE FIGURE

Figura 1	La nostra <i>GeoArea</i> : La regione Abruzzo	4
Figura 2	Generico elemento z_k dell'insieme Z	5
Figura 3	Generico elemento b_i dell'insieme B	6
Figura 4	(a) Generico elemento e_j dell'insieme; (b)Esempio di curve di livello per un certo territorio	7
Figura 5	Esempio di buffer_i costruito su una generica b_i	7
Figura 6	$V=\{v_1, v_2, v_3, v_4, v_5\}$	8
Figura 7	$P=\{p_2, p_3\}$	8
Figura 8	$T=\{t_1, t_2, t_3, t_4, t_5\}$	9
Figura 9	Esempio di una v_γ con tre curve di livello	9
Figura 13	Esempio di angolo azimuth per un certo punto	12
Figura 14	Esempio di vettore direzionale per un certo edificio b_i	13
Figura 15	$Q = \{q_1, q_2, q_3\}$	13
Figura 16	Esempio di un <i>trend=salita</i> nel caso in cui la stazione si trova ad una quota minore rispetto la curva di livello q_1	14
Figura 17	Esempio di un <i>trend = salita</i> nel caso in cui la stazione si trova alla stessa quota della curva di livello q_1	14
Figura 18	Esempio di un <i>trend=discesa</i> nel caso in cui la stazione si trova a quota maggiore rispetto la della curva di livello q_1	15
Figura 19	Esempio di un <i>trend=discesa</i> nel caso in cui la stazione si trova a quota maggiore rispetto la della curva di livello q_1	15
Figura 20	dove	16
Figura 21	Esempio di un edificio b_i all'interno di una zona z_k	17
Figura 22	Esempio di un edificio b_i esterno alla zona z_k	18
Figura 23	Esempio di un edificio all'interno di una zona v_γ	19
Figura 24	Esempio di una v_γ con $h_{\text{media}}_{v_\gamma}$ elevata la cui intersezione con il buffer_i risulta essere un'area piccola	20
Figura 25	Esempio di rappresentazione geometrica di una linea ferroviaria	28
Figura 26	Esempio di campionamento di una linea ferroviaria	29
Figura 27	dove	29
Figura 28	dove	30
Figura 29	La regione Abruzzo	32
Figura 30	Stazioni ferroviarie presenti in Abruzzo	33
Figura 31	Linee ferroviarie presenti in Abruzzo	34
Figura 32	Partizionamento completo della regione Abruzzo	34
Figura 33	Diagramma E-R della base di dati	36
Figura 34	Schema logico	39

Figura 37	dove	50
Figura 38	Stazione di Archi	52
Figura 39	Stazione di Perano	52
Figura 40	Stazione di Pettorano	52
Figura 41	Stazione di Civitaluparella	52
Figura 42	Visione d'insieme delle 24142 zone generate	55
Figura 43	Particolare delle zone in prossimità del <i>boundary</i>	56

ABSTRACT

Short summary of the contents...a great guide by Kent Beck how to write good abstracts can be found here:

<https://plg.uwaterloo.ca/~migod/research/beck00PSLA.html>

INTRODUZIONE

La mobilità delle persone e delle merci sono una componente essenziale del mercato interno dell'Unione Europea (UE) ed è di fondamentale importanza garantire la sua fattibilità al fine di salvaguardare la crescita economica. La rete ferroviaria ha un ruolo strategico in questo contesto almeno sotto due punti di vista:

1. Offre un impulso rilevante all'integrazione e all'efficienza dell'economia dell'UE.
2. Facilita il libero movimento delle persone e delle merci garantendo una modalità di trasporto efficiente e sostenibile dal punto di vista ambientale. Secondo l'Agenzia europea dell'ambiente, le emissioni di CO₂ provenienti dal trasporto ferroviario sono 3,5 volte inferiori, per tonnellata-chilometro, a quelle prodotte dal trasporto su strada. Di conseguenza, promuovere modalità di trasporto come quello su rotaia, piuttosto che su strada, permetterebbe all'UE di essere meno dipendente dall'importazione di petrolio e di ridurre l'inquinamento.

Alla luce di queste considerazioni preliminari non sorprende che la sicurezza assuma un rilievo crescente in questo contesto; Una testimonianza di ciò è il regolamento su un metodo comune di sicurezza (CSM) per la valutazione del rischio e l'accertamento, rilasciato negli ultimi anni dall'agenzia ferroviaria europea (ERA) e dalla Commissione Europea [Regulation EU, No.402/2013, 2015]. Un recente regolamento [GE/GN8642, 2014] pubblicato dalla Rail Safety and Standards Board Ltd (<http://www.rssb.co.uk/>) definisce "*un pericolo, come una condizione che potrebbe portare a un incidente*". A sua volta, la direttiva sulla sicurezza [2004/49/CE] del Parlamento europeo definisce un incidente come: "*per 'incidente' si intende un evento improvviso indesiderato e non intenzionale o specifica catena di siffatti eventi aventi conseguenze dannose. Gli incidenti sono suddivisi nelle seguenti categorie: collisioni, deragliamenti, incidenti ai passaggi a livello, incidenti a persone causati da materiale rotolante in movimento, incendi ed altri.*"

Nel tempo sono stati sviluppati molti metodi per predire, prevenire e ridurre il numero di incidenti e per aumentare la sicurezza nel contesto ferroviario. Inoltre è fondamentale sottolineare il fatto che le tratte e le relative stazioni insistono su un territorio esposto, tra gli altri, al rischio idrogeologico. Di riflesso anche la rete ferroviaria e le relative stazioni che su di esso insistono lo sono. I deragliamenti dei treni sono il rischio principale di incidenti. [Lloyd et al., 2001], [Manning et al., 2008], [He et al., 2011] sono esempi di studi che hanno affrontato la questione. In tali studi si fa uso di metodi matematici classici. Ad esempio, [He et al. (2011)] fanno ricorso alla

Principal Component Analysis per identificare un numero limitato di fattori che devono essere presi in considerazione per studiare la stabilità dei pendii rocciosi. Studi teorici come quelli appena citati sono molto interessanti perché aiutano a comprendere la complessità del problema, fornendo delle informazioni globali circa le variabili in gioco e come esse sono correlate, ma hanno il limite di non essere direttamente traducibili in piani di lavoro attuabili da parte di coloro che hanno responsabilità di garantire la sicurezza delle tratte ferroviarie e le relative stazioni.

Nel nostro documento si adotta un approccio differente, basato su metodi proposti nell'ambito della Geographical Information Science e riferiti allo studio di relazioni topologiche tra caratteristiche geografiche. Come vedremo, facendo ricorso a tali metodi, possibile identificare quali sono le potenziali pericolosità delle tratte ferroviarie e delle relative stazioni determinate dalle caratteristiche morfologiche e litologiche del terreno dove sono situate, dalla sua pendenza nonché dall'uso dello stesso (prevalentemente abitativo, prevalentemente ad uso coltivazione, ecc).

Le frane costituiscono uno dei più importanti rischi naturali in molti paesi in tutto il mondo, [Brabb e Harrod, 1989]. Questo è il caso dell'Italia, dove le frane sono diffuse e provocano notevoli danni e decessi ogni anno, [Guzzetti, Stark, e Salvati, 2005], [Trigila et al., 2010]. Secondo il risultato di un recente studio effettuato da [Jaedicke et al. (2014)], “*l'Italia ha il più alto numero di persone esposte al pericolo di frana tra i paesi europei.*”

Mantenendosi ad un alto livello di astrazione quello che si può anticipare del metodo che ci si accinge a presentare è che esso consente di fare delle valutazioni circa le potenziali pericolosità delle tratte ferroviarie e delle relative stazioni che insistono sul territorio oggetto di studio. Ciò consente di restituire agli utenti finali informazioni dettagliate per la localizzazione delle porzioni di tratte ferroviarie e delle stazioni ove prioritariamente è opportuno fare dei controlli periodici. La strategia proposta quindi supera i limiti posti dai controlli periodici sull'intero percorso ferroviario e su tutte le stazioni situate sul territorio, i quali risultano essere lunghi e, di conseguenza, costosi e non sempre utili come sostenuto da [Sadler et al., (2016)]. Utilizzando tale metodo sarà possibile individuare quali sono le porzioni di tratte ferroviarie e le stazioni le cui misure di controllo possono essere ridotte in modo sicuro, permettendo al settore ferroviario significativi risparmi di costi a favore del potenziamento dei controlli di porzioni di tratte ferroviarie e stazioni che hanno una pericolosità potenziale maggiore.

In sintesi, obiettivo dello studio metodologico-sperimentale descritto in questo documento è sviluppare una metodologia scientificamente robusta per l'identificazione degli hotspot ferroviari ad alto rischio e che sia di supporto nella definizione di una strategia di monitoraggio selettiva dello stato di sicurezza per una categoria di beni strategica per un Paese, ovvero la sua rete ferroviaria, rispetto all'esposizione della

stessa ai movimenti franosi.

Le tratte ferroviarie e le relative stazioni sono dunque i due target da proteggere e data la loro diversa connotazione spaziale che le descrive vanno trattate ed analizzate separatamente. Infatti, le tratte sono modellabili con la geometria *linea*, mentre le stazioni sono modellabili con la geometria *punto*.

Lo strumento software da noi sviluppato è basato sull'uso di un Geo-DB arricchito con un ampio numero di User Defined Functions (UDF), tramite le quali è possibile effettuare delle elaborazioni complesse scrivendo delle queries SQL semplici alla portata della maggior parte dei tecnici che lavorano nel settore della sicurezza ferroviaria. In dettaglio, il sistema informatico è in grado di fornire quanto segue:

1. Circa le stazioni, una classifica delle stazioni ferroviarie costruita in relazione al loro valore di esposizione al rischio di frane e una relativa classificazione in fasce di rischio delle stesse;
2. Circa le linee ferroviarie che sono presenti sul territorio, una classifica delle tratte ferroviarie costruita in relazione al loro valore di esposizione al rischio di frane e una relativa classificazione in fasce di rischio delle stesse.

Il documento è organizzato come segue: nel Capitolo 2 verrà mostrata una panoramica degli strumenti e gli algoritmi utilizzati per valutare l'esposizione al rischio di frana delle stazioni ferroviarie dislocate sul territorio. Nel Capitolo 3 verrà discussa la situazione attuale del caso di studio, ove il metodo proposto sarà sperimentato con riferimento alla rete ferroviaria italiana dislocata all'interno del territorio della regione Abruzzo e quindi, verranno presentati i datasets di input. Nel Capitolo 4 verranno presentati e discussi i risultati ottenuti dall'applicazione dei metodi discussi nel Capitolo 2, in particolare verranno presentate anche i punti di forza e le criticità di tali applicazioni. Il Capitolo 5 invece presenterà quelle che sono le considerazioni finali circa il lavoro realizzato e presenterà alcuni possibili spunti per eventuali lavori futuri.

Come già preannunciato, il metodo proposto sarà sperimentato con riferimento alla rete ferroviaria Italiana dislocata all'interno del territorio della regione Abruzzo (Italia Centrale).

MATERIALS AND METHODS

2.1 NOTAZIONI

Da qui in poi useremo le seguenti notazioni:

1. *GeoArea*: Porzione di territorio di interesse per lo studio. La *GeoArea* può coincidere con un comune, una provincia, una regione, una nazione o un intero continente. Nel nostro caso di studio si tratta della regione Abruzzo (fig. 1).

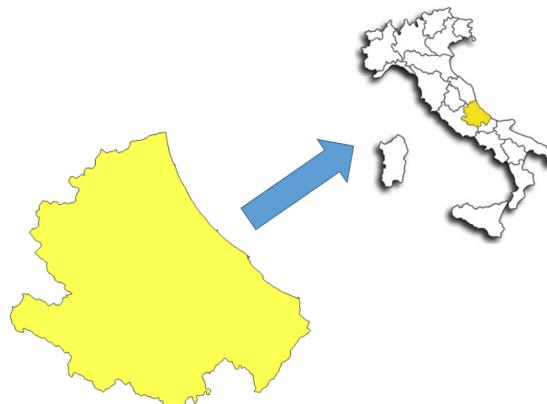


Figura 1: La nostra *GeoArea*: La regione Abruzzo

La *GeoArea* viene descritta dalla seguente tupla: <gid, geom>.

- *gid*: Identificativo numerico univoco del territorio.
 - *geom*: Rappresentazione geometrica del territorio (*MultiPolygon*).
2. $Z = \{z_k, k=1,2,\dots\}$ dove z_k denota una zona k-esima della *GeoArea*, ognuna caratterizzata da un insieme di informazioni territoriali. L'insieme Z è una partizione della *GeoArea*, ovvero è una collezione di elementi z_k , non nulli, tali da avere intersezione vuota a due a due e tali che la loro unione coincida con l'intera *GeoArea*, per maggiore chiarezza si rimanda alla fig. 2.



Figura 2: Generico elemento z_k dell'insieme Z

Il generico elemento di Z (i.e. z_k) è definito mediante la tupla: $\langle \text{gid}, S_{z_k}, \text{geom}, \text{avgElevation} \rangle$.

- gid : Identificativo numerico univoco dell'elemento z_k .
- S_{z_k} : Valore numerico che quantifica la pericolosità di frana dell'elemento z_k .
- geom : Rappresentazione geometrica del *boundary* dell'elemento z_k .
- avgElevation : Altitudine media dell'elemento z_k , ottenuta mediante una media ponderata delle altitudini delle curve di livello che intersecano, almeno in parte, l'elemento z_k . Il modo in cui verrà calcolato questo valore verrà descritto nel dettaglio nella notazione successiva di Δh poiché bisogna prima introdurre ulteriori concetti.

Si noti che z_k è una notazione sovraccaricata dato che rappresenta sia il *gid* della zona che la sua geometria.

3. $B = \{ b_i, i=1,2,\dots \}$ dove b_i indica l'edificio della stazione (B sta per *Station Building*) contenuta nel *boundary* della *GeoArea*. Di questo edificio non teniamo conto della sua conformazione architettonica, della sua altezza o del numero di piani della struttura poiché spesso sono delle informazioni che difficilmente sono reperibili ma bensì consideriamo la sua posizione geografica, espressa da una coppia di coordinate, informazione spesso nota. Un esempio di un generico elemento b_i è rappresentato in fig. 3.

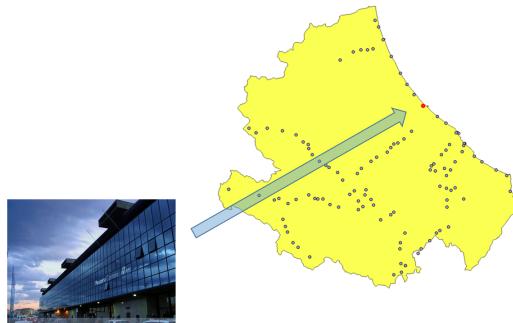


Figura 3: Generico elemento b_i dell'insieme B

Il generico elemento di B (i.e. b_i) è definito mediante la tupla: $\langle gid, name, geom, Exp_b_i, elevation \rangle$.

- *gid*: Identificativo numerico univoco dell'elemento b_i .
- *name*: Nome testuale dell'i-esimo elemento dell'insieme B , espresso tramite una stringa.
- *geom*: Rappresentazione geometrica (*Point*) dell'elemento b_i .
- *Exp_b_i*: Grado di esposizione al pericolo di frana del generico elemento b_i , determinato da tutte le z_k in Z .
- *elevation*: Altitudine espressa in metri a cui si trova il generico elemento b_i .

Si noti che b_i è una notazione sovraccaricata dato che rappresenta sia il *gid* della stazione ferroviaria che la sua geometria.

4. $E = \{ e_j, j=1,2,\dots \}$ dove e_j denota una curva di livello contenuta almeno in parte nel *boundary* della *GeoArea*. In geografia, con particolare riguardo alla cartografia, la curva di livello è quella curva che unisce punti con uguale quota (figura 4(b)), ovvero uguale distanza verticale dal piano di riferimento al quale è stato attribuito quota zero; se sono sopra il livello del mare si chiameranno *isoipse* (dal greco *íisos* = "uguale" e *hýpsos* = "altezza") mentre al contrario *isobate* (dal greco *íisos* = "uguale" e *báthos* = "profondità"). Nella figura 4(a) è possibile vedere un esempio di e_j .

Ogni elemento dell'insieme E (i.e. e_j) è definito mediante la tupla: $\langle gid, elevation, geom \rangle$.

- *gid*: Identificativo numerico univoco dell'elemento e_j .
- *elevation*: Valore numerico che rappresenta la quota espressa in metri dell'elemento e_j .
- *geom*: Rappresentazione geometrica (*MultiLineString*) dell'elemento e_j .

Si noti che e_j è una notazione sovraccaricata dato che rappresenta sia il *gid* della curva di livello che la sua geometria.

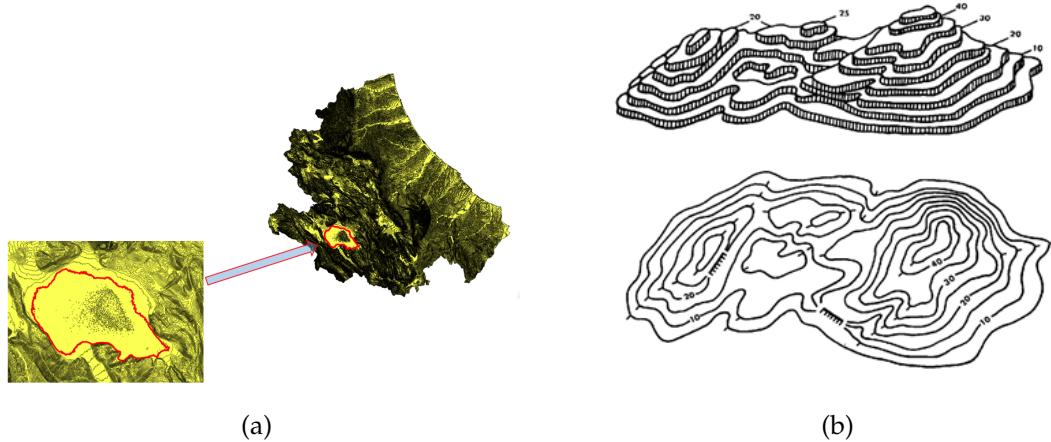


Figura 4: (a) Generico elemento e_j dell'insieme; (b)Esempio di curve di livello per un certo territorio

5. $buffer_i = buffer(b_i, r)$ rappresenta una geometria costituita da tutti i punti la cui distanza dal punto b_i risulta essere minore o uguale del raggio r . Nella figura 5 è possibile vedere un esempio di $buffer_i$.

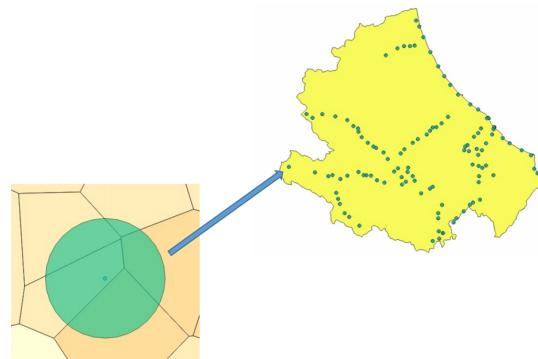


Figura 5: Esempio di buffer_i costruito su una generica b_i

6. $V = \{v_\gamma \mid (\gamma=1,2,\dots) := z_k \in Z \mid \text{buffer}_i \cap z_k \neq \emptyset\}$ è l'insieme di tutte le z_k contenute nella *GeoArea* che intersecano almeno in parte buffer_i . La cardinalità di tale insieme è pari a N . Un esempio di insieme V è osservabile nella fig. 6.



Figura 6: $V = \{v_1, v_2, v_3, v_4, v_5\}$

7. $P = \{p_m \mid (m=1,2,\dots) := \text{buffer}_i \cap e_j \neq \emptyset, \forall e_j \in E\}$ è l'insieme dei segmenti delle curve di livello, determinati dall'intersezione tra le curve di livello contenute nell'insieme E e il buffer_i . In figura 7 è riportato un esempio di insieme P .

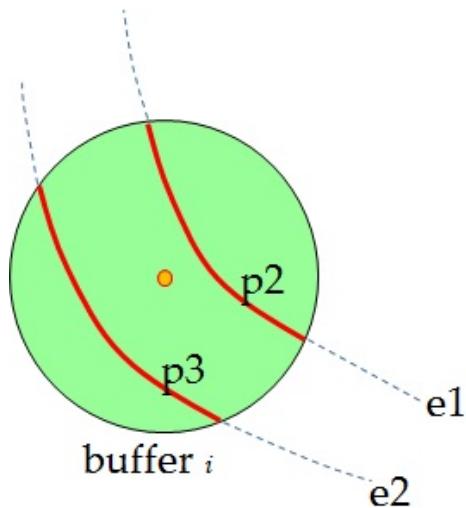


Figura 7: $P = \{p_2, p_3\}$

8. $T = \{t_f \mid f=1,2,\dots\} := v_\gamma \cap e_j \neq \emptyset, \forall e_j \in E, \forall v_\gamma \in V\}$ è l'insieme dei segmenti delle curve di livello, determinati dall'intersezione tra le curve di livello contenute nell'insieme E e le zone v_γ contenute nell'insieme V . In fig. 8 è possibile osservare un esempio di insieme T .

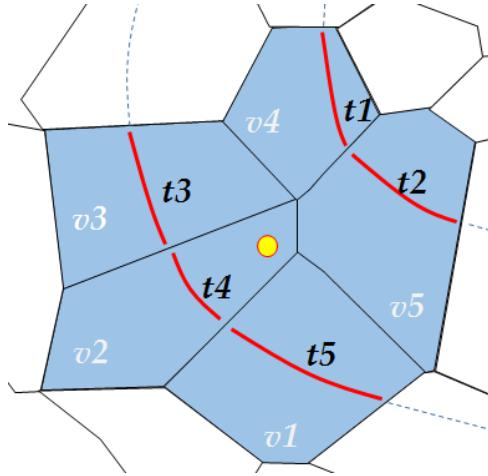


Figura 8: $T = \{t_1, t_2, t_3, t_4, t_5\}$

9. Δh è così definito:

$$\Delta h = |h_{stazione} - h_{media}_{v_\gamma}| \quad (1)$$

Dove:

- $h_{stazione}$ indica l'*elevation* dell'edificio b_i
- $h_{media}_{v_\gamma}$ indica l'*avgElevation* media di v_γ

Il metodo di calcolo dell'*elevation* per gli edifici e le zone è diverso. Infatti qui di seguito descriveremo le diverse modalità di calcolo:

- a) $h_{media}_{v_\gamma}$: l'*elevation* delle v_γ viene calcolata effettuando la media aritmetica dell'*elevation* delle curve di livello che intersecano v_γ .

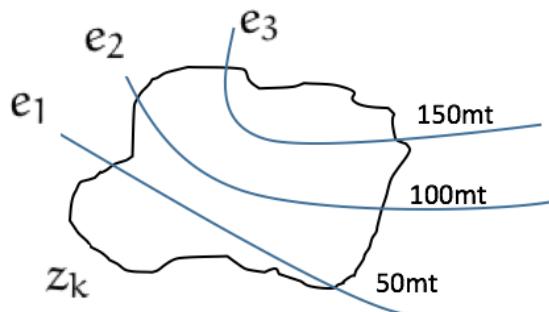


Figura 9: Esempio di una v_γ con tre curve di livello

Ad esempio, in fig.9 la quota media di v_γ risulta essere:
 $\frac{50 + 100 + 150}{3} = 100\text{mt}$

- b) h_stazione: il calcolo dell'*elevation* per gli edifici è più sofisticato ma permette di ottenere valori molto accurati. Le possibili situazioni che si possono presentare sono tre e verranno analizzate qui di seguito:
- edificio racchiuso da un'unica curva di livello (Fig.10) oppure in prossimità del confine della *GeoArea* (Fig.11). Si definisce un edificio di questo tipo come edificio in posizione "atypica".

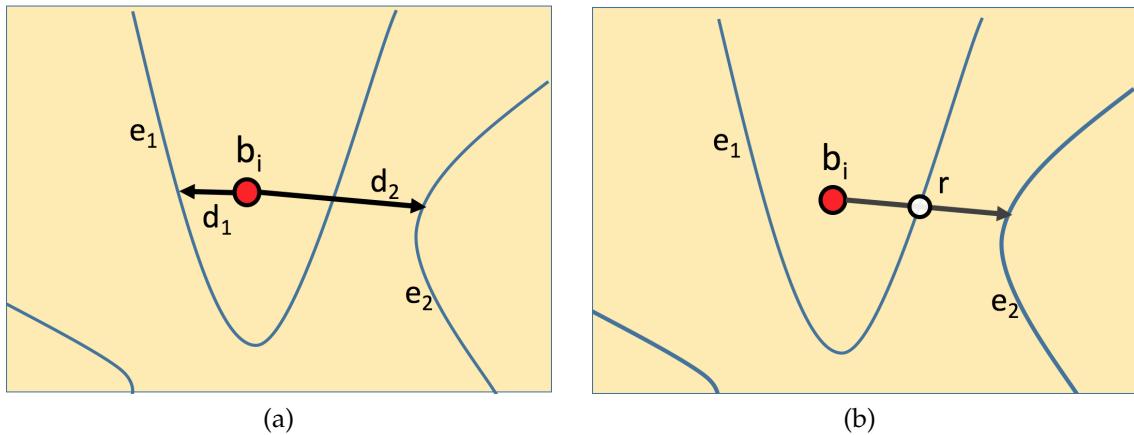


Figura 10: Edificio racchiuso da un'unica curva di livello.

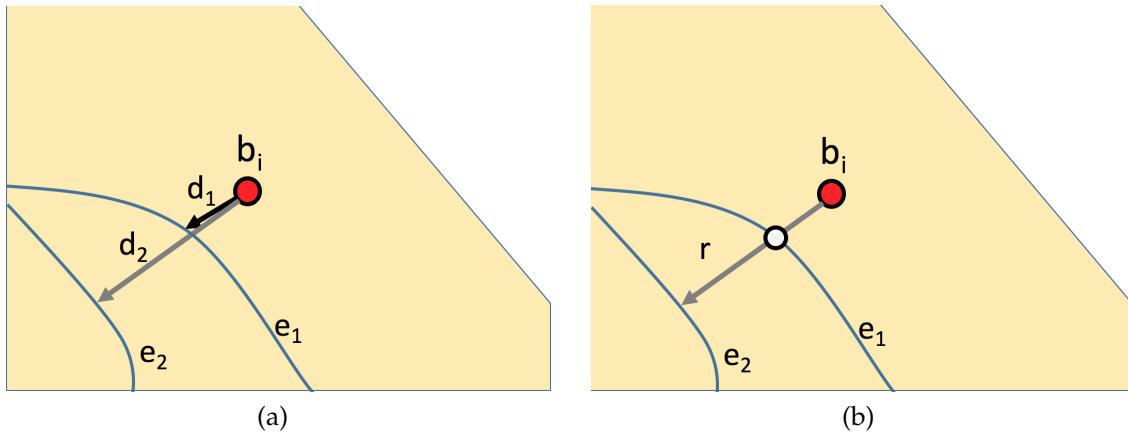


Figura 11: Edificio in prossimità del confine della *GeoArea*.

Entrambi i casi vengono rilevati nel medesimo modo, ovvero tracciando la semiretta avente come origine l'edificio b_i (pallino rosso) e passante per la curva e_2 . Se la semiretta interseca anche e_1 , allora ci troviamo in una delle situazioni esplicate in Fig.10 (b) e in Fig. 11 (b). Quindi h_stazione equivale all'*elevation* della curva di livello e_1 , supponendo che $d_1 < d_2$.

- Edificio circondato da due curve di livello come in Fig.12. Si definisce un edificio di questo tipo come edificio in posizione "tipica".

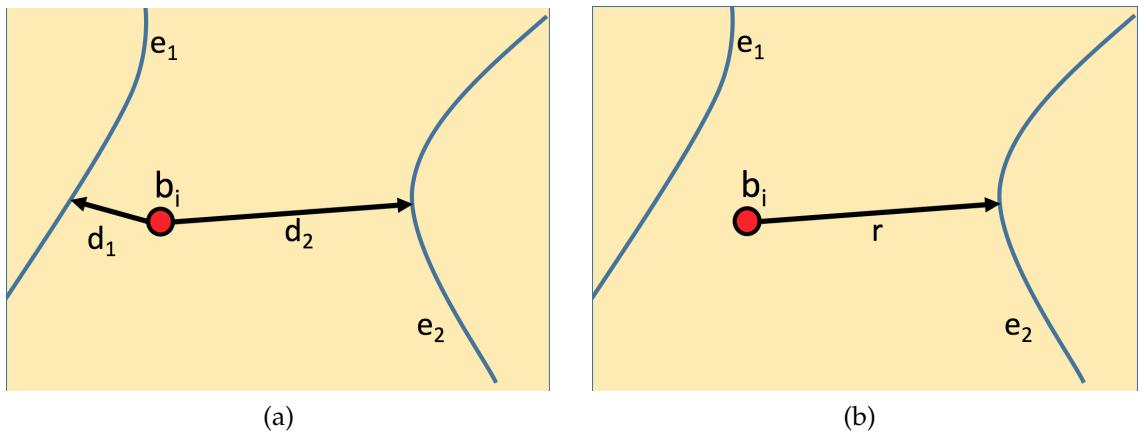


Figura 12: Edificio in prossimità del confine della *GeoArea*.

Dalla Fig.12 (b) si evince che, a differenza delle situazioni esplicate dalle Fig.10 (b) e Fig.11 (b), la semiretta avente origine in b_i e passante per e_2 non interseca la curva di livello e_1 . Dunque $h_{stazione}$ è così definito:

$$h_{stazione} = \frac{(h_{e1} \times d_2) + (h_{e2} \times d_1)}{d_1 + d_2} \quad (2)$$

Dove:

- h_{e_j} indica l'*elevation* della curva di livello e_j , con $j = 1, 2$
- d_j indica la distanza tra l'edificio b_i e la curva di livello e_j con $j = 1, 2$

L'Eq.2 è una media ponderata delle quote delle due curve di livello più vicine all'edificio b_i . Al numeratore le distanze vengono utilizzate come pesi della media, affinché la quota della curva e_1 , più vicina all'edificio b_i , risulti più determinante nel calcolo della $h_{stazione}$ rispetto alla quota della curva e_2 .

10. *Azimuth*: L'azimuth è la coordinata orizzontale angolare espressa dall'arco d'ortodromia della sfera celeste che si forma partendo convenzionalmente dal punto cardinale nord fino all'oggetto di osservazione, muovendosi in senso orario verso est, quindi a sud e a ovest, fino a tornare al punto di inizio a nord (cioè un angolo giro, 360° sessagesimali); la coordinata azimutale quindi, verrà espressa in gradi angolari sessagesimali/minuti/secondi, oppure in radianti, e avrà sempre un valore numerico positivo. Un punto nel cielo che si trovi esattamente a nord (il polo celeste dell'emisfero boreale) avrà una coordinata azimutale di 0° (0° 360°), se invece si trova esattamente a est sarà di 90° , esattamente a sud di 180° , esattamente a ovest di 270° , ecc. Si propone un esempio in Fig.13.



Figura 13: Esempio di angolo azimuth per un certo punto

11. *vettoreDirezionale*: il vettore direzionale viene così costruito:

- trova l'*Azimuth* tra b_i e il punto più vicino della p_m più vicina.
 - Crea il punto A , le cui coordinate (x_a, y_a) sono le coordinate di b_i traslate di Δx e Δy .
- Dove:

- $\Delta x = \cos(\text{Azimuth}) \times (r + 1)$
- $\Delta y = \sin(\text{Azimuth}) \times (r + 1)$

con r pari al raggio del $buffer_i$.

- crea infine il *vettoreDirezionale*, ovvero il vettore avente come origine b_i e come punto finale il punto A sopra definito.

Si propone un esempio esplicativo in Fig.14.



Figura 14: Esempio di vettore direzionale per un certo edificio b_i

12. $Q = \{ q_w \ (w=1,2,\dots,\beta) \} := p_m \in P \mid (p_m \cap \text{vettoreDirezionale}) \neq \emptyset \mid d_1 < d_2 < \dots < d_\beta \}$.

Dove:

- d_w è la distanza tra b_i e il punto $(p_m \cap \text{vettoreDirezionale})$

Per costruzione i pedici stabiliscono un ordinamento crescente degli elementi dell'insieme Q rispetto la distanza appena definita.

La Fig.15 propone un esempio esplicativo basato sulla Fig.14. Come si vede sono state escluse dall'insieme Q le $p_m \mid (p_m \cap \text{vettoreDirezionale} = \emptyset)$

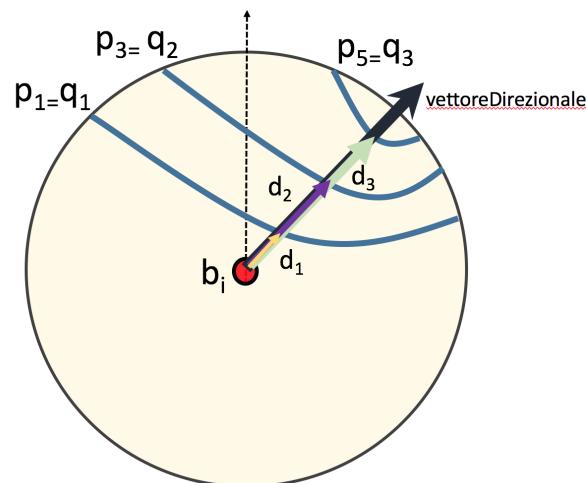


Figura 15: $Q = \{q_1, q_2, q_3\}$

13. *trend*: Il trend rappresenta l'andamento del territorio nelle immediate vicinanze di b_i in direzione del vettore *Direzionale* e può assumere uno dei seguenti valori {*salita*, *discesa*}, in base alle seguenti condizioni:

- $trend = \text{salita}$ se $\text{elevation}_{q_1} > h_{\text{stazione}}$.

Esempio in Fig.16



Figura 16: Esempio di un *trend=salita* nel caso in cui la stazione si trova ad una quota minore rispetto la curva di livello q_1

- $trend = \text{salita}$ se $\text{elevation}_{q_1} = h_{\text{stazione}} \wedge \text{elevation}_{q_2} > \text{elevation}_{q_1}$.
Esempio in Fig.17.



Figura 17: Esempio di un *trend = salita* nel caso in cui la stazione si trova alla stessa quota della curva di livello q_1

- $trend = \text{discesa}$ se $\text{elevation}_{q_1} < h_{\text{stazione}}$.
Esempio in Fig.18

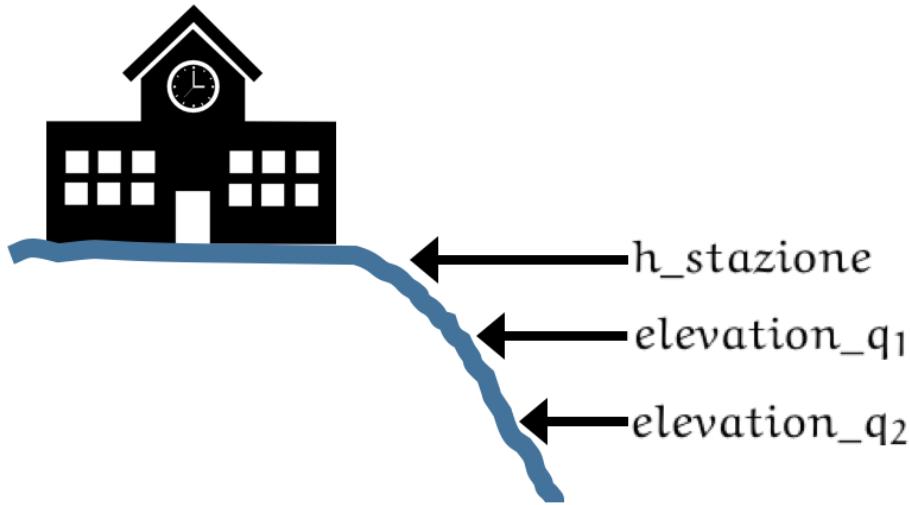


Figura 18: Esempio di un $trend=\text{discesa}$ nel caso in cui la stazione si trova a quota maggiore rispetto la della curva di livello q_1

- $trend = \text{discesa}$ se $\text{elevation}_{q_1} = h_{\text{stazione}} \wedge \text{elevation}_{q_2} < \text{elevation}_{q_1}$.
Esempio in Fig.19



Figura 19: Esempio di un $trend=\text{discesa}$ nel caso in cui la stazione si trova a quota maggiore rispetto la della curva di livello q_1

14. *situazione Monotonica*: una situazione monotonica è garantita fino ad un certo $q_{\bar{w}}$ se:

- è garantita $\forall w < \bar{w}$
- la differenza di elevation tra $q_{\bar{w}}$ e $q_{\bar{w}-1}$ conferma l'andamento del $trend$.
Ovvero:

$$\begin{cases} \text{elevation}_{q_{\bar{w}}} < \text{elevation}_{q_{\bar{w}-1}}, & \text{se } trend = \text{discesa} \\ \text{elevation}_{q_{\bar{w}}} > \text{elevation}_{q_{\bar{w}-1}}, & \text{se } trend = \text{salita} \end{cases} \quad (3)$$

15. *pendenza* è una stima dell'andamento del territorio vicino l'edificio b_i lungo il suo vettore *Direzionale* associata a q_w . Dal punto di vista geometrico la *pendenza* rappresenta il coefficiente angolare tra due punti. In un caso tra una stazione b_i e il punto ottenuto dall'intersezione tra il *vettoreDirezionale* e la curva di livello q_1 . Nell'altro caso tra i punti ottenuti dall'intersezione tra il *vettoreDirezionale* e le curve q_w e q_{w+1} . In entrambi i casi il coefficiente viene calcolato rispetto al sistema di riferimento della *GeoArea*. Come mostrato nella figura 20



Figura 20: Esempio esplicativo di *pendenza*. Δy corrisponde al numeratore e Δx al denominatore dell'Eq.4, per $w = 1$ subFig.20 (a) e per $w \geq 2$ subFig.20 (b)

Dal punto di vista matematico l'equazione per determinare la *pendenza* è la seguente:

$$\begin{cases} \frac{|h_{stazione} - elevation_{q_w}|}{d(b_i, (vettoreDirezionale \cap q_w))} & \text{se } w=1 \\ \frac{|elevation_{q_w} - elevation_{q_{w-1}}|}{d((vettoreDirezionale \cap q_w), (vettoreDirezionale \cap q_{w-1}))} & \text{se } w \geq 2 \end{cases} \quad (4)$$

16. *slope_i*: E' la somma aritmetica delle *pendenze* di tutti gli elementi di Q che garantiscono una *situazione Monotonica*.

2.2 UN METODO PER CLASSIFICARE LE STAZIONI FERROVIARIE

2.2.1 *Metodo di calcolo originario*

Prima di procedere con l'ideazione di un metodo di calcolo, si è pensato di effettuare un'analisi critica e dettagliata del metodo originario, frutto del lavoro di altri studenti negli anni passati. Supponiamo di avere quindi una *GeoArea* partizionata in zone diverse e indicate con z_k , ognuna delle quali ha associato un valore s_{zk} rappresentante il pericolo di frana. Si ricorda che tali notazioni sono già state introdotte in 2.1. Tale

metodo assegna un valore di *exposure* (Exp_{bi}) ad un generico edificio (b_i) all'interno della *GeoArea* nel seguente modo:

1. L'*Exposure* per l'edificio b_i è ottenuto come somma dei contributi di tutte le zone nella *Geoarea* per tale edificio.

$$\text{Exp}_{bi} = \sum_{k=1}^N \text{Exp}_{b_{i,k}} \quad (5)$$

2. Il contributo della generica zona z_k all'*exposure* dell'edificio b_i viene calcolato utilizzando la seguente equazione:

$$\text{Exp}_{b_{i,k}} = \overline{\text{Size}} \times S_{z_k} \times \begin{cases} 1, & \text{se } b_i \text{ è contenuta in } z_k \\ \frac{1}{d^3}, & \text{altrimenti} \end{cases} \quad (6)$$

Dove:

- $\overline{\text{Size}}$ è il rapporto tra l'area di z_k e l'area media di tutte le zone che costituiscono Z
- d denota la distanza minima tra b_i e il *boundary* della zona z_k

Al prodotto di questi due parametri viene poi moltiplicato un certo valore in base alla posizione relativa tra la zona z_k e l'edificio b_i . Tale valore corrisponderà ad 1 se vi è una situazione analoga alla figura seguente:

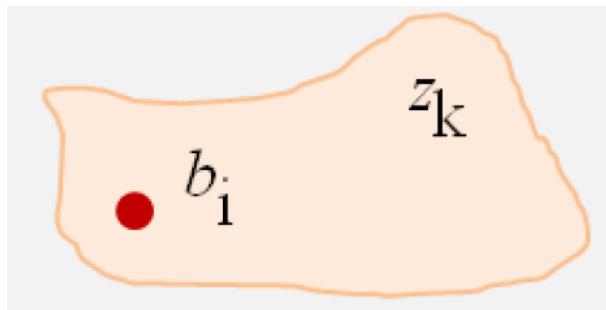


Figura 21: Esempio di un edificio b_i all'interno di una zona z_k

Se invece l'edificio b_i si trova esternamente rispetto alla zona z_k , il valore sarà $\frac{1}{d^3}$.

Detto ciò possiamo concludere questa breve panoramica sul metodo di calcolo originario con alcune critiche e considerazioni, che hanno rappresentato la base per la realizzazione di un nuovo metodo:

- Il metodo non tiene conto delle curve di livello che cadono nella *GeoArea*;
- Il valore del parametro *Size* è vicino allo zero per zone molto piccole, mentre è maggiore di 1 per tutte le zone che hanno area superiore al valore dell'area media delle zone della *GeoArea*. Inoltre *Size* assume valori molto alti per zone con area molto estesa.



Figura 22: Esempio di un edificio b_i esterno alla zona z_k

2.2.2 *Metodo di calcolo ex novo*

Il metodo proposto vuole sostituirsi a quello precedente, vedi 2.2.1, con lo scopo di migliorare il calcolo dell'*exposure*. Ovvero si vogliono assegnare valori che rappresentino con maggiore fedeltà il pericolo reale di frana per gli edifici. Il punto di forza del nuovo metodo, d'ora in avanti indicato con *NMC*, consiste nello sfruttare il "potenziale informativo" aggiuntivo fornito con l'introduzione delle **curve di livello**.

Viene da sé che un uso "intelligente" di queste curve, permette di ricavare importanti informazioni ai fini del nostro obiettivo, tra le quali:

- La quota dell'edificio
- La morfologia del territorio

Supponiamo per un attimo di essere riusciti a ricavare queste informazioni e di averle utilizzate in modo corretto, il *NMC* avrebbe così superato la prima debolezza del metodo originario, vedi 2.2.1. Tuttavia rimarrebbero due questioni in sospeso, ovvero:

1. Quali zone z_k della *GeoArea* influiscono realmente sul calcolo dell'*exposure* del generico edificio b_i ?
2. Supposto di aver trovato risposta al punto precedente, in che proporzione tali zone influiscono? tutte allo stesso modo?

Il metodo originario, rispondeva in parte a queste domande (vedi equazione 6). Utilizzando il parametro $(\frac{1}{d^3})$ si riusciva a rendere nullo il contributo di zone z_k abbastanza lontane, in quanto la funzione al cubo inversa smorza rapidamente, rispondendo così alla prima domanda.

Per quanto riguarda il secondo punto, si utilizzava il parametro *Size*, questo però porta alla generazione di falsi positivi/negativi nel calcolo dell'*exposure* dell'edificio, come già sottolineato nel paragrafo precedente.

Dunque vediamo come il *NMC* ricava le informazioni dalle curve di livello e come le utilizza nel calcolo delle *exposure* degli edifici b_i .

L'equazione del *NMC* che determina il valore finale di *exposure* per la generica b_i è la seguente:

$$\text{Exp}_{b_i} = \left(\sum_{\gamma=1}^N \text{Exp}_{b_i, \gamma} \right) + (\text{slope}_i \times \alpha_{\text{slope}}) \quad (7)$$

Tale valore dipende dalla somma di due addendi.

Il primo addendo rappresenta la sommatoria dei contributi delle zone $v_\gamma \in V$, ovvero la sommatoria degli $Exp_{-b_i, \gamma}$. Si ricorda che il limite superiore N della sommatoria nell'Eq.7 si riferisce al numero di zone $v_\gamma \in V$, ovvero alla cardinalità dell'insieme V .

La formula per calcolare il contributo della zona v_γ alla determinazione del pericolo di frana dell'edificio b_i è la seguente:

$$Exp_{-b_i, \gamma} = Size \times S_{v_\gamma} \times \Delta h \times \begin{cases} 1, & \text{se } b_i \text{ è contenuta in } v_\gamma \\ \frac{1}{d}, & \text{altrimenti} \end{cases} \quad (8)$$

Si notino le differenze rispetto l'equazione del metodo originario (vedi Eq.5).

Nella versione proposta dal NMC il parametro $(\frac{1}{d})$ non è elevato al cubo. Questo per evitare di smorzare troppo il contributo delle v_γ vicine. Il parametro *Size* è così definito:

$$Size = \frac{\text{area}(\text{buffer}_i \cap v_\gamma)}{\text{area}(v_\gamma)} \quad (9)$$

Ovvero l'area della geometria restituita dall'intersezione tra il poligono (v_γ) e il buffer_i che come abbiamo definito nelle notazioni ha raggio r ed è costruito attorno all'edificio b_i . Per un dato valore del raggio r , il valore del numeratore dell'Eq. 8 cresce se la posizione di b_i si avvicina al centroide della zona che lo contiene (ovvero v_γ), di conseguenza, in base all'Eq.7, anche il corrispondente valore dell'exposure sarà più alto.

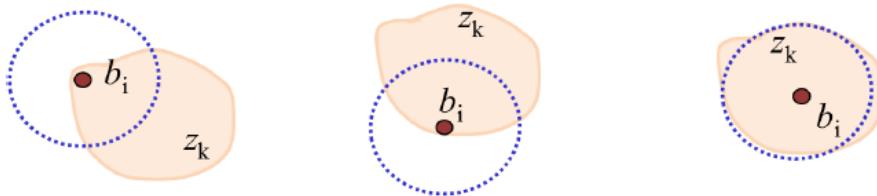


Figura 23: Esempio di un edificio all'interno di una zona v_γ

Il parametro *Size* risponde alle domande che ci siamo posti precedentemente (vedi 2.2.2), ovvero:

- Quali zone z_k della *GeoArea* influiscono realmente sul calcolo dell'*exposure* di b_i ? È facile ora rispondere a questa domanda, tutte e sole le zone v_γ , ovvero come abbiamo già detto le zone z_k per le quali vale:

$$\text{area}(\text{buffer}(b_i, r) \cap z_k) \neq \emptyset \quad (10)$$

Bisogna stabilire quanto deve essere il valore del raggio, infatti se tale valore è sufficientemente grande da garantire che eventi franosi a distanza maggiore da

b_i non interessino quest'ultimo, allora le zone z_k che non intersecano il Buffer non devono contribuire al calcolo dell'*exposure*.

- In che proporzione tali zone v_γ influiscono? Tutte allo stesso modo? Ovviamente no. Grazie all'espressione 9, il parametro *Size* tiene conto anche dell'area effettivamente occupata da v_γ all'interno del buffer. In questo modo si riduce notevolmente la probabilità che il *NMC* generi falsi positivi/negativi.

Tale parametro rappresenta, in modo quantitativo, un ulteriore legame tra l'Eq.8 e la morfologia del territorio. Vediamo un esempio concreto della sua utilità.

La figura 24 è stata estrapolata mediante il software di visualizzazione di dati territoriali QGIS e mostra una particolare situazione.

La sfera rossa rappresenta un edificio mentre la circonferenza in blu il buffer, di cui abbiamo ampiamente parlato, di raggio r pari a 500mt.

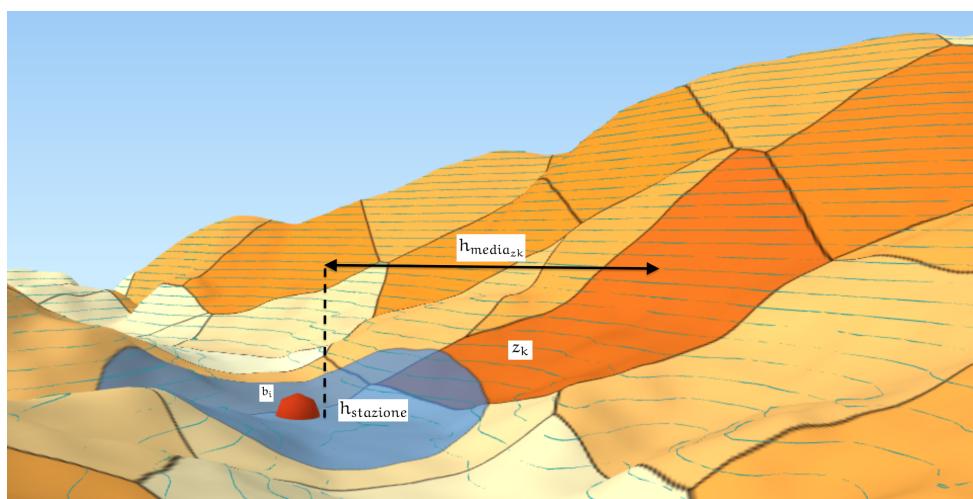


Figura 24: Esempio di una v_γ con $h_{\text{media}} v_\gamma$ elevata la cui intersezione con il buffer_i risulta essere un'area piccola

Notiamo come la v_γ in questione occupi una parte molto piccola del buffer_i . Questo sicuramente porterà ad avere un valore di *Size* molto basso e di conseguenza anche il contributo di v_γ all'*exposure*, generando un falso negativo. Il fattore Δh interviene in soccorso di *Size*, permettendo all' $\text{Exp}_{b_i, \gamma}$ di assumere in valore non eccessivamente basso, tenendo conto dell'elevata differenza tra l'altezza della stazione b_i e l'altezza media di v_γ .

Il secondo addendo dell'Eq.7 va a rappresentare un fattore di rischio dovuto al dislivello che si trova più vicino a b_i e che quindi va considerato con maggior peso.

α_{slope} è un fattore moltiplicativo intero utilizzato per incrementare o diminuire il peso di slope_i nel calcolo di Exp_{-b_i} rispetto al primo addendo.

2.2.3 L'algoritmo

Lo scopo di questo paragrafo è formalizzare gli algoritmi che implementa NMC, descritto nella sezione 2.2.2 .

Algoritmo 1

Il primo algoritmo denominato averageElevationNearZones ha come obiettivo quello di calcolare il valore di avgElevation associato ai v_γ in V . Prende come parametri di input gli insiemi Z , E e l'elemento $buffer_i$ mentre restituisce come output l'insieme V .

Algorithm 1: averageElevationNearZones

Input : $Z, E, buffer_i$

Output: V

Result: Calcola la quota media delle zone che intersecano il buffer

```
1   $V = \emptyset;$ 
2  avgElevation = 0;
3  contatore = 0;
4  for each  $z_k \in Z$  do
5      if  $z_k \cap buffer_i$  then
6          for each  $e_j \in E$  do
7              if  $z_k \cap e_j$  then
8                  avgElevation = avgElevation + (elevation di  $e_j$ );
9                  contatore = contatore + 1;
10             end
11         end
12         if contatore > 0 then
13             avgElevation = avgElevation / contatore;
14         end
15         aggiorna il valore avgElevation di  $z_k$ ;
16          $v_\gamma = z_k;$ 
17          $V = V \cup v_\gamma;$ 
18         avgElevation = 0;
19         contatore = 0;
20     end
21 end
22 return  $V;$ 
```

Commenti:

Dalla riga 1 alla 3 sono presenti alcune inizializzazioni: l'insieme V inizialmente è vuoto, mentre sono fissati a 0 i valori di avgElevation e contatore. Dalla riga 4 alle

21 è presente un ciclo che scandisce gli elementi z_k di Z . Immediatamente all'interno del ciclo (righe 5-20) viene effettuato un controllo che verifica se la zona in esame ha un'intersezione con il buffer.

In caso di esito negativo si passa subito al successivo elemento di Z . In caso di esito positivo, per ogni elemento e_j in E , si verifica se interseca con la zona z_k in esame. Se anche questo controllo viene superato si incrementa il valore di avgElevation con l'elevation di e_j e il valore di contatore di 1.

Una volta completato il ciclo sugli elementi in E , si controlla (riga 12) se contatore ha un valore maggiore di zero, quindi se almeno per una linea e_j è verificata l'intersezione con la z_k esaminata. In questo caso il valore di avgElevation viene sovrascritto a avgElevation/contatore.

Si passa quindi (riga 15) ad aggiornare il valore avgElevation nella tupla z_k con il valore appena calcolato.

Alla riga 16 la tupla z_k viene copiata nella tupla v_γ che viene inserita nell'insieme V . Questa operazione viene eseguita per mantenere nell'insieme V solo gli elementi z_k di interesse, ossia quelli per cui è verificata l'intersezione con $buffer_i$.

Alle righe 18 e 19 le variabili avgElevation e contatore sono resettate a 0, per poter essere appropriatamente riutilizzate nella successiva iterazione del ciclo.

L'algoritmo infine (riga 22) restituisce l'insieme V .

Riassumendo l'algoritmo prende tutte e sole le z_k che intersecano il buffer e per ognuna calcola la quota media. Inserisce infine queste zone nell'insieme V che restituisce come risultato.

La complessità computazionale dell'algoritmo è data dal ciclo in riga 4 e da quello in riga 6. È quindi pari a $\mathcal{O}(\text{card}(Z) \times \text{card}(E))$ anche se mediamente è molto inferiore in quanto è molto improbabile che ogni elemento in Z intersechi con $buffer_i$ e con tutte le linee in E .

Algoritmo 2

L'algoritmo denominato elevationBuild ha come obiettivo quello di stimare l'elevation di b_i . Prende come parametri di input gli insiemi V (l'insieme delle zone che intersecano $buffer_i$), E (l'insieme delle curve di livello) e l'elemento b_i . In output restituisce lo stesso elemento b_i di input il cui campo elevation è stato aggiornato con la stima calcolata.

Algorithm 2: elevationBuild

Input : V, E, b_i
Output: b_i
Result: Calcola una stima dell'elevation di b_i

```
1  $T = \emptyset;$ 
2 for each  $v_\gamma \in V$  do
3   for each  $e_j \in E$  do
4     if  $v_\gamma \cap e_j$  then
5       geom di  $t_f$  = st_intersection (  $v_\gamma, e_j$  );
6       elevation di  $t_f$  = elevation di  $e_j$ ;
7        $T = T \cup t_f$ ;
8     end
9   end
10 end
11 for each  $t_f \in T$  do
12   dist = st_distance (  $b_i, t_f$  );
13   Trova  $t_{f\min}$  t.c. dist è minima;
14   Trova  $t_{f\min2}$  t.c. dist è la successiva al minimo;
15 end
16 if  $b_i$  in posizione atipica then
17    $t_{f\min2} = t_{f\min}$ ;
18 end
19 Calcola  $h_{stazione}$ 
20 Aggiorna elevation di  $b_i$  con  $h_{stazione}$ ;
21 return  $b_i$ ;
```

Commenti:

Nella prima riga dell'algoritmo l'insieme T è inizializzato come insieme vuoto.

Nelle righe 2-10 è presente un ciclo annidato. Il ciclo più esterno itera sugli elementi v_γ in V (l'insieme delle zone che intersecano il buffer), mentre quello più interno sugli elementi e_j in E (l'insieme delle linee di livello). Se l'intersezione tra gli elementi v_γ e e_j esiste, allora questa intersezione viene assegnata come geometria di t_f mentre come sua elevation viene assegnata l'elevation di e_j . L'ennupla t_f così definita è inserita nell'insieme T (riga 7).

All'interno del ciclo in riga 11 si vanno a scandire i t_f in T . All'interno del ciclo si va a calcolare la distanza dist tra l'elemento t_f e b_i . Si vanno quindi a individuare (nelle righe 13-14) $t_{f\min}$ e $t_{f\min2}$.

$t_{f\min}$ è la $t_f \in T$ tale che la distanza dist tra t_f e b_i è la distanza minima tra tutte le distanze calcolate tra ogni altra $t_f \in T$ e b_i .

In modo analogo $t_{f\min2}$ è la $t_f \in T$ la cui distanza dist tra t_f e b_i ha un valore immediatamente successivo alla distanza minima.

Alla riga 16 si effettua un controllo per verificare se b_i si trova in una posizione "atipi-

ca" come definito nella sezione 2.1. Nel caso di esito positivo si copia l'ennupla $t_{f\min}$ in $t_{f\min2}$.

Si calcola quindi alla riga 19 il valore di h_{stazione} come descritto precedentemente (in sezione 2.1), dove $t_{f\min}$ rappresenta da e_1 , mentre $t_{f\min2}$ rappresenta e_2 .

Nella riga 20 si aggiorna l'elevation di b_i con il valore appena calcolato h_{stazione} . Viene infine restituiti b_i .

Riassumendo l'algoritmo individua i segmenti di linea frutto dall'intersezione tra tutte le linee in E e le zone in V e li inserisce nell'insieme T . Individua poi i 2 elementi in T che sono più vicini a b_i . Sfruttando questi 2 elementi calcola h_{stazione} cioè una stima dell'elevation di b_i .

La complessità computazionale è data dal ciclo annidato in riga 2 ed è pari a $\mathcal{O}(\text{card}(V) \times \text{card}(E))$ anche se mediamente è molto inferiore in quanto è molto improbabile che ogni elemento in V intersechi tutte le linee in E .

Algoritmo 3

L'algoritmo denominato slopeFactor ha come obiettivo quello di calcolare lo slope associato a b_i . Prende come parametri di input l'insieme E delle curve livello e l'elemento $buffer_i$, mentre restituisce $slope_i$ ossia lo slope associato a b_i .

Algorithm 3: slopeFactor

Input : E , $buffer_i$
Output: $slope_i$
Result: Calcola lo slope associato a b_i

```
1  $P = \emptyset;$ 
2  $Q = \emptyset;$ 
3  $slope_i = 0;$ 
4 for each  $e_j \in E$  do
5   if  $buffer_i \cap e_j$  then
6     geom di  $p_m = st\_intersection ( buffer_i, e_j );$ 
7     elevation di  $p_m = elevation di e_j;$ 
8      $P = P \cup p_m;$ 
9   end
10 end
11 Calcola il vettoreDirezionale;
12 for each  $p_m \in P$  do
13   if  $p_m \cap vettoreDirezionale$  then
14     Calcola la distanza tra  $b_i$  e  $(p_m \cap vettoreDirezionale);$ 
15     Aggiorna distance di  $p_m$  con la distanza calcolata;
16      $q_w = p_m;$ 
17      $Q = Q \cup q_w ;$ 
18   end
19 end
20 for each  $q_w \in Q$  ordinato rispetto a distance do
21    $slope_i = slope_i + (pendenza associata a q_w);$ 
22 end
23 return  $slope_i;$ 
```

Commenti:

Nelle prime righe (1-3) vengono inizializzati gli insiemi P , Q come insiemi vuoti e la variabile $slope_i$ a zero.

Dalla riga 4 inizia un ciclo che scandisce gli elementi e_j in E e per ognuno controlla se interseca con $buffer_i$. Se questa intersezione esiste, questa viene assegnata come geometria di p_m mentre come sua elevation viene assegnata l'elevation di e_j . La tupla p_m così definita, viene inserita nell'insieme P . Queste operazioni permettono di lavorare su un insieme che è molto più piccolo di E e che contiene solo le porzioni di linee di interesse.

Alla riga 11 viene calcolato il vettoreDirezionale come definito nella sezione 2.1. Nella righe 12-19 è presente un ciclo che scandisce gli elementi p_m in P e per ognuno verifica se interseca con il vettoreDirezionale. Se questa intersezione esiste viene calcolata la distanza tra b_i e l'intersezione. Il valore di distanza calcolato va ad aggiornare il valore distance in p_m . La tupla p_m così costruita viene quindi copiata in q_w . Alla riga 17, q_w viene inserito nell'insieme Q . In questo modo in Q si vanno a inserire tutti e soli gli elementi p_m che intersecano il vettoreDirezionale.

Nella riga 20 inizia un ciclo che scandisce gli elementi q_w in Q ordinati rispetto al campo distance. All'interno del ciclo viene calcolata la pendenza associata a q_w come definita nelle sezione 2.1. Questo valore va a incrementare slope_i.

Viene infine restituito slope_i.

Riassumendo l'algoritmo individua le porzioni di linee di livello che si trovano tutte in una stessa direzione rispetto a b_i e per ognuna di esse calcola la pendenza, ossia la ripidità tra la linea e la linea precedente, dove la linea precedente è più vicina a b_i . slope_i è la somma di queste pendenze.

La complessità computazionale è data dal ciclo in riga 4 ed è quindi pari a $\mathcal{O}(\text{card}(E))$.

Algoritmo 4

L'algoritmo denominato `computeExpposure` è l'algoritmo che effettivamente calcola l'expposure associato a b_i . All'interno dell'algoritmo vengono richiamati tutti e tre gli algoritmi precedentemente definiti.

Algorithm 4: `computeExpposure`

Input : $B, r, \alpha_{slope}, E, Z$
Output: B
Result: Calcola l'expposure di tutte le b_i

```
1 for each  $b_i \in B$  do
2      $buffer_i = st\_buffer(b_i, r);$ 
3      $V = averageElevationNearZones(Z, E, buffer_i);$ 
4      $b_i = elevationBuild(V, E, b_i);$ 
5      $slope_i = slopeFactor(E, b_i, buffer_i);$ 
6     for each  $v_\gamma \in V$  do
7         Calcola Size;
8         Calcola  $Exp_{b_i, \gamma}$ ;
9     end
10    Calcola  $Exp_{b_i};$ 
11    Aggiorna il valore di  $Exp_{b_i}$  in  $b_i;$ 
12 end
13 return  $B;$ 
```

Commenti:

Nelle righe 1-11 è presente un ciclo che scandisce tutte le b_i in B . Per ogni b_i viene calcolato un buffer circolare di raggio r . Viene successivamente costruito un insieme V che contiene tutte le zone che intersecano il buffer, dove a ogni zona viene associata una quota media (richiamando l'algoritmo 1).

Nella riga 4 la tupla b_i viene aggiornata inserendo un valore che rappresenta una stima della sua quota (richiamando l'algoritmo 2).

Nella riga 5 viene quindi calcolato $slope_i$ (richiamando l'algoritmo 3).

Dalla riga 6 inizia un ciclo che scandisce gli elementi in V e per ognuno calcola $Size$ e $Exp_{b_i, \gamma}$.

In riga 10 si calcola Exp_{b_i} (tenendo conto anche del fattore α_{slope} come mostrato nell'Eq.7) che va ad aggiornare il valore nella tupla b_i .

L'algoritmo ritorna come risultato l'insieme B dove ogni tupla b_i viene arricchita con un stima della quota e un valore di $Explosure$.

La complessità computazionale è data dai due cicli presenti alle righe 1 e 10 ed è quindi pari a $\mathcal{O}(\text{card}(V) \times \text{card}(B))$.

2.3 UN METODO PER CLASSIFICARE LE LINEE FERROVIARIE

Il metodo proposto per classificare le stazioni (vedi Sez.2.2.2) può essere utilizzato anche per classificare le linee ferroviarie. Occorre tuttavia estendere il NMC a causa della diversa rappresentazione geometrica dei due elementi.

2.3.1 *Metodo di calcolo*

Poiché il NMC è stato progettato per classificare stazioni, rappresentate attraverso dei punti geometrici, non può essere utilizzato così com'è per classificare le linee ferroviarie rappresentate invece attraverso linee geometriche. Di seguito in Fig.25 un esempio di tratta ferroviaria.

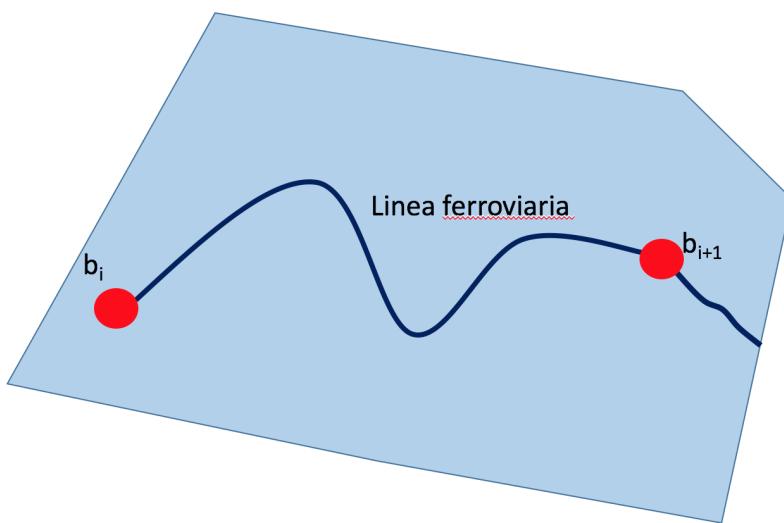


Figura 25: Esempio di rappresentazione geometrica di una linea ferroviaria

L'idea di fondo è campionare con un certo passo, le linee ferroviarie. Fatto ciò si avrà una sequenza di punti sui quali poter applicare il NMC.

Nonostante l'idea sembri essere molto semplice, ci sono tre questioni sulle quali bisogna fare delle riflessioni e proporre delle soluzioni:

1. che passo di campionamento? la scelta di un passo piccolo giova alla correttezza del calcolo finale di *exposure* della tratta ma aumenta significativamente i tempi di esecuzione del calcolo stesso.
2. le linee sono semplici? le linee ferroviarie potrebbero avere delle traiettorie complesse, il metodo deve essere in grado di gestirle.
3. come aggregare i punti campionati? dopo aver campionato una certa tratta e aver determinato i valori di *exposure* di ogni punto della sequenza che la compongono, bisognerà aggregare in qualche modo tali risultati per poter aver un

risultato utile e consultabile. Il campionamento potrebbe generare un numero considerevole di punti che presi singolarmente non forniscono un contributo significato al problema in questione (vedi 1)

L'estensione del NMC (d'ora in avanti indicato con NMCE) per le linee ferroviarie propone una soluzione per ognuna di queste problematiche. Nell'ordine in cui sono state precedentemente presentate:

1. Il passo di campionamento scelto è di 300 mt, ma può essere facilmente modificato essendo un parametro del metodo. Di seguito in Fig.26 un esempio di campionamento.

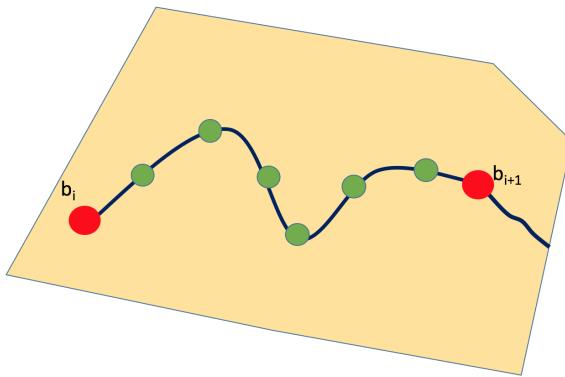
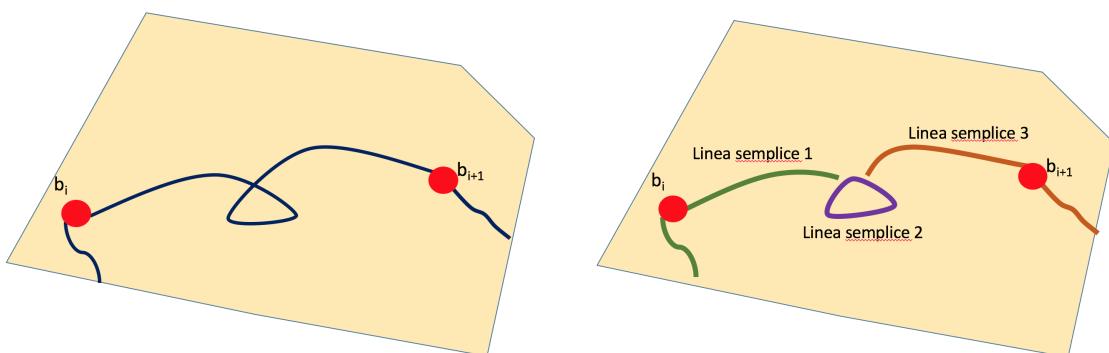


Figura 26: Esempio di campionamento di una linea ferroviaria

2. nel caso in cui una tratta abbia una traiettoria complessa, il NMCE rileva tale situazione e provvede a dividere la linea in due o più linee semplici. In Fig.27 un esempio esplicativo.



(a) Esempio di una linea ferroviaria complessa (b) Esempio di linee semplici ottenute dividendo la linea in Fig. 27a

Figura 27: Esempio di una linea ferroviaria complessa e della divisione in più linee semplici

Come si evince dalla Fig.27b, la linea ferroviaria complessa in Fig.27a viene spezzata in tre linee semplici.

3. una volta campionata la tratta semplice, come in Fig.26, si utilizza il NMC per determinare il valore di *exposure* di ogni punto che la compone.

La linea ferroviaria viene quindi divisa in segmenti da 1 km ciascuno (fatta eccezione per l'ultimo segmento che potrebbe essere più piccolo) come in Fig.37a. Quindi il valore di *exposure* del segmento viene calcolato facendo la media aritmetica dell'*exposure* dei punti campionati al suo interno. Di seguito in Fig.37b si propone un'esempio esplicativo, il colore del segmento tende sempre più al rosso all'aumentare del valore medio di *exposure* dei punti contenuti nel segmento.

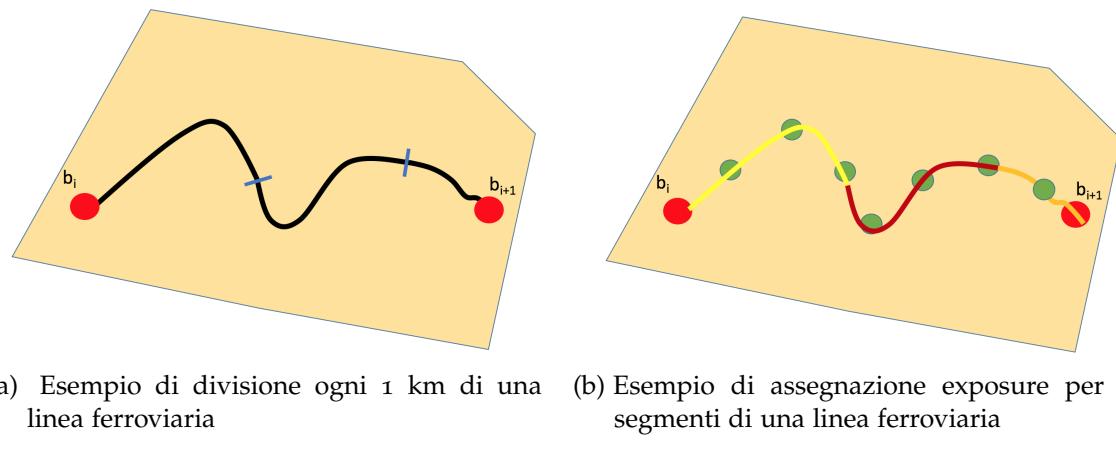


Figura 28: Esempio di divisione ogni 1 km di una linea ferroviaria e assegnazione dell'*exposure* ad ogni suo segmento

Per ogni segmento si tiene traccia anche del valore massimo e della somma dell'*exposure* dei punti al suo interno. Queste informazioni potrebbero in alcuni casi essere utili. Ad esempio se ci fosse un punto il cui valore di *exposure* risulti essere molto alto rispetto gli altri punti, la media aritmetica non sarebbe sufficiente a rilevare tale situazione.

2.3.2 L'algoritmo

UN CASO DI STUDIO: L'ABRUZZO

Il focus del presente studio è posto, come è stato già preannunciato in precedenza, unicamente sulla regione Abruzzo, la cui rete ferroviaria, nonostante non abbia una estensione considerevole, è immersa in un territorio eterogeneo ed esposto a diversi livelli di rischio di natura idrogeologica. La ridotta estensione dei dati in esame, ha consentito di focalizzarsi su una analisi puntuale e non dispersiva dei metodi proposti, al fine di valutarne accuratamente la validità ed i limiti. Il presente *report* è stato prodotto a fronte di un lavoro di laboratorio volto alla costruzione dei seguenti dati:

- Una classifica delle stazioni, in base al potenziale rischio idrogeologico a cui esse sono esposte;
- Una classifica dei tratti di ferrovia maggiormente esposti a rischio di smottamenti del terreno su cui sono posti i binari.

La produzione di tali classifiche ha l'obiettivo di consentire una fruibilità facile ed immediata dei dati per quegli enti preposti al monitoraggio e alla manutenzione delle tratte e stazioni ferroviarie.

Di seguito si mostrerà una breve istantanea della attuale situazione osservata nel territorio abruzzese, con particolare attenzione ai dati selezionati, tra quelli disponibili, e alla loro provenienza.

3.1 TERRITORIO

L'Abruzzo (rappresentata in fig.29) è una regione a statuto ordinario dell'Italia peninsulare, compresa tra il mare Adriatico e l'Appennino centrale, con capoluogo L'Aquila. La regione è geograficamente ed economicamente parte dell'Italia centrale, mentre dal punto di vista storico e linguistico è inserita all'interno dell'Italia meridionale.

Occupava una superficie di 10.831 km² e ha una popolazione di 1.322.349 abitanti. È diviso in quattro province: L'Aquila, Chieti, Pescara e Teramo, e in 305 comuni. Confina a nord con le Marche, ad est con il mare Adriatico, ad ovest con il Lazio e a sud con il Molise. Si divide principalmente in una parte costiera nel versante orientale con le spiagge dell'Adriatico, e in una parte montuosa dal lato occidentale con il Gran Sasso d'Italia (2.914 m s.l.m.), la Majella (2.793 m s.l.m.) e il Sirente-Velino (2.487 m s.l.m.) che costituiscono i tre massicci montuosi più alti dell'intera catena appenninica.



Figura 29: La regione Abruzzo

3.2 RETE FERROVIARIA

Per quanto riguarda il trasporto ferroviario, vi è in Abruzzo una forte disparità tra quello moderno sulla costa (anche se con numero e qualità delle corse e del servizio imparagonabili rispetto all'asse ferroviario "tirrenico" e non è prevista alcuna linea di TAV) e quello delle zone interne, molto carente in termini di modernità e qualità del servizio e in attesa da decenni di interventi di potenziamento e ammodernamento (vedi in particolare la linea Pescara-Avezzano-Roma). Complessivamente comunque la rete ferroviaria abruzzese in attivo è abbastanza sviluppata e si estende per 524 km, mentre l'estensione totale complessiva della rete (attiva o dismessa) è pari a 648 km.

3.2.1 *Stazioni Ferroviarie*

I dati iniziali a nostra disposizione circa i nodi ferroviari ci forniscono le informazioni relative alle 114 stazioni distribuite sul territorio abruzzese mostrate in fig.30.



Figura 30: Stazioni ferroviarie presenti in Abruzzo

3.2.2 *Linee Ferroviarie*

Sul territorio abruzzese sono presenti nove tratte ferroviarie:

- Bologna - Bari
- Ortona - Crocetta
- Marina di San Vito - Castel di Sangro
- Roma - Pescara
- Avezzano - Roccasecca
- Archi stazione - Atessa
- Sulmona - Carpinone
- Rieti - L'Aquila - Sulmona
- Teramo - Giulianova

Di queste tratte solo alcune sono completamente interne alla regione, altre invece la attraversano (avendo origine e/o fine al di fuori dei confini); riguardo queste ultime si è scelto di considerare unicamente le porzioni che cadono nella regione Abruzzo mostrate in fig.[31](#).



Figura 31: Linee ferroviarie presenti in Abruzzo

3.3 ZONE

L'intero territorio abruzzese è stato completamente suddiviso in 22.000 zone più piccole la cui area media è pari a $0,5 \text{ km}^2$. Il dataset relativo a queste zone ,che ci è stato fornito, è stato costruito mediante un algoritmo che segue un approccio basato sul *diagramma di Voronoi*, tale da costruire una partizione completa del territorio determinata dalle distanze rispetto ad un determinato insieme discreto di elementi dello spazio (nel nostro caso un insieme finito di punti). Il risultato di tale partizionamento dello spazio è visibile nella Fig.32.



Figura 32: Partizionamento completo della regione Abruzzo

Ad ogni z_k è associato un valore numerico compreso tra 0 e 1, il quale quantifica la pericolosità determinata dal rischio idrogeologico della zona alla quale si riferisce.

3.4 BASI DI DATI SPAZIALE

Si descrive in seguito la progettazione e conseguente realizzazione della basi di dati geografica realizzata per l'effettiva implementazione degli algoritmi precedentemente descritti.

I dati raccolti nella basi dati sono relativi:

- alle stazioni e linee ferroviarie abruzzesi con informazioni relative al nome e alla geometria;
- alla geometria dei confini dell'area geografica;
- alla geometria delle zone che vanno a partizionare in modo completo l'area con relative informazioni di rischio idrogeologico;
- alla geometria delle linee di livello presenti nel territorio di riferimento con la quota associata;
- ai risultati ottenuti dalle elaborazioni.

3.4.1 Progettazione concettuale

Si procede ora con l'illustrare la progettazione concettuale della base di dati spaziale. Per ogni entità verranno mostrate:

- gli attributi
- l'attributo identificante
- i vincoli di partecipazione nelle relazioni, con la notazione (min, max)

In Figura 33 si riporta il modello E-R della base di dati spaziale, mentre in Tabella 1 sono elencate le entità con i relativi attributi, descrizione e attributi identificanti, ed infine in Tabella 2 l'elenco delle associazioni.

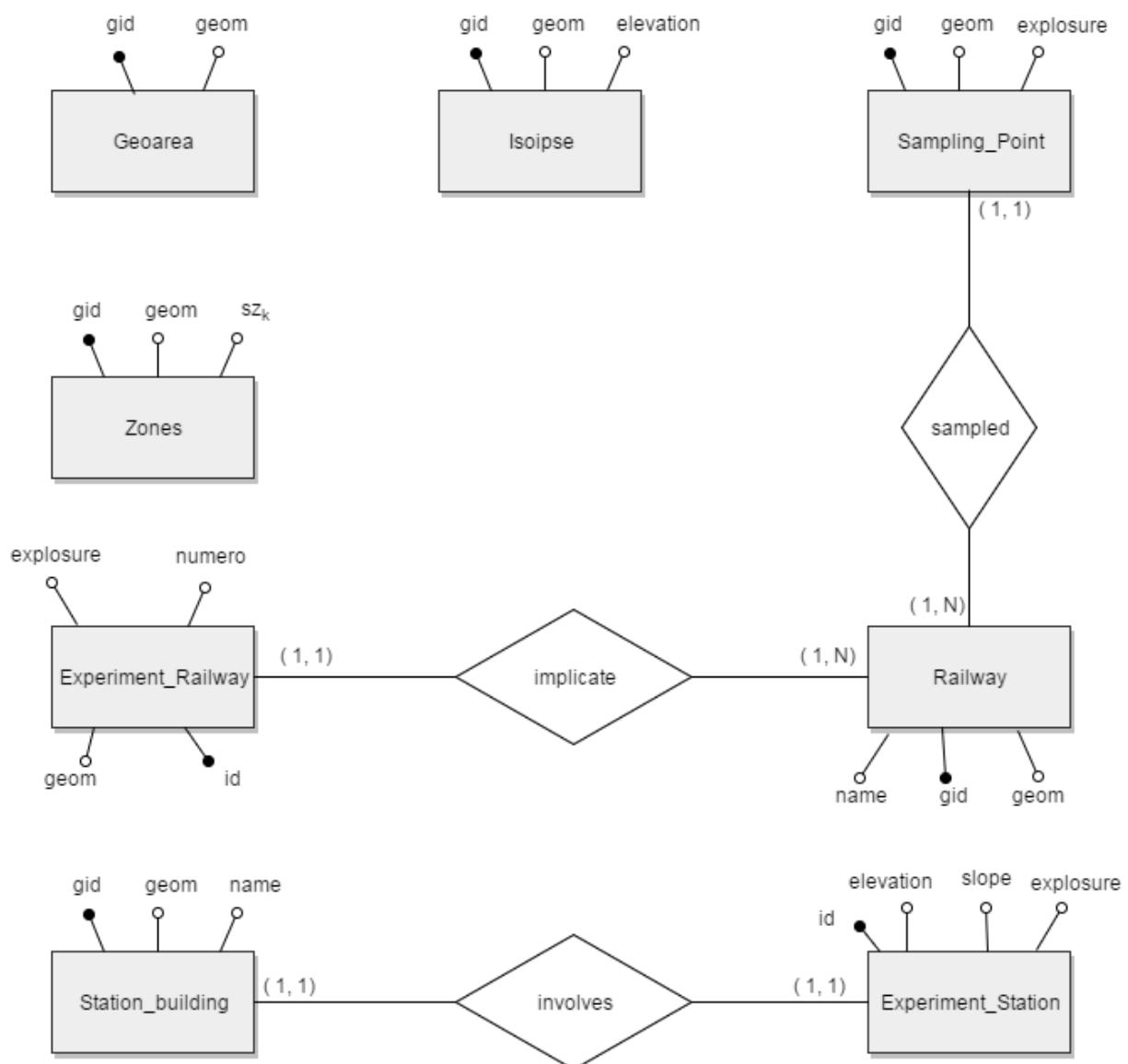


Figura 33: Diagramma E-R della base di dati

Note:

l'attributo *numero* dell'entità *Experiment_railway* rappresenta il numero di tratte all'interno della linea. Nel caso di divisione delle linee in tratte da un km, allora corrisponde al km lungo la tratta.

Entità	Descrizione	Attributi	ID
Geoarea	Entità relativa al territorio di interesse. Nel nostro caso è la regione Abruzzo.	<i>gid</i> <i>geom</i>	<i>gid</i>
Zones	Entità relativa al territorio di interesse. Nel nostro caso è un partizionamento completo della regione Abruzzo.	<i>gid</i> <i>sz_k</i> <i>geom</i>	<i>gid</i>
Station_building	Entità relativa alle stazioni presenti nel territorio di riferimento.	<i>gid</i> <i>name</i> <i>geom</i>	<i>gid</i>
Isoipse	Entità relativa alle isoipse presenti nel territorio di riferimento.	<i>gid</i> <i>elevation</i> <i>geom</i>	<i>gid</i>
Railway	Entità relativa alle linee ferroviarie presenti nel territorio di riferimento.	<i>gid</i> <i>name</i> <i>geom</i>	<i>gid</i>
Sampling_point	Entità relativa ai punti individuati come campionamento delle linee ferroviarie presenti nel territorio di riferimento.	<i>id</i> <i>geom</i> <i>exposure</i>	<i>id</i>
Experiment_station	Entità relativa agli esperimenti effettuati sulle stazioni presenti nel territorio di riferimento.	<i>id</i> <i>elevation</i> <i>slope</i> <i>exposure</i>	<i>id</i>
Experiment_railway	Entità relativa agli esperimenti effettuati sulle linee ferroviarie presenti nel territorio di riferimento.	<i>id</i> <i>geom</i> <i>numero</i> <i>exposure</i>	<i>id</i>

Tabella 1: Tabella delle entità

Associazione	Descrizione	Entità coinvolte	Attributi
involves	Relazione che evidenzia il coinvolgimento di una sola stazione in ogni esperimento.	<i>station_building</i> <i>experiment_station</i>	<i>nessuna</i>
implicated	Relazione che evidenzia il coinvolgimento di almeno una linea ferroviaria in uno o più esperimenti.	<i>railway</i> <i>experiment_railway</i>	<i>nessuna</i>
sampled	Relazione che lega ogni linea ferroviaria con uno o più punti di campionamento.	<i>railway</i> <i>sampling_point</i>	<i>nessuna</i>

Tabella 2: Tabella delle associazioni

3.4.2 Progettazione logica

Il diagramma E-R in figura 33 viene tradotto nello schema logico rappresentato graficamente in figura 34

Visual Paradigm Standard (Università degli Studi dell'Aquila)

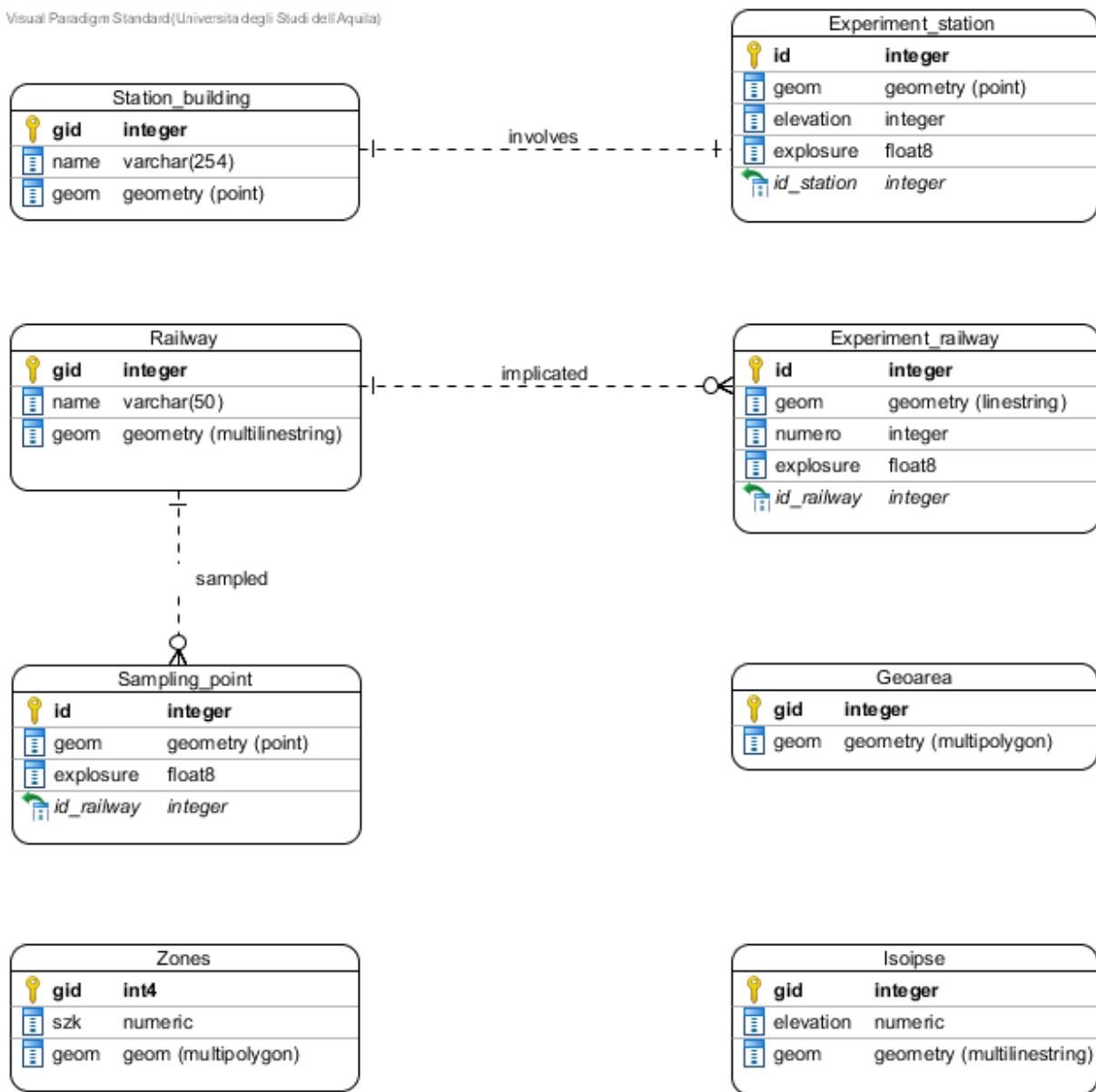


Figura 34: Schema logico

RISULTATI E LORO DISCUSSIONE

Nella sezione n° 4 verranno illustrati e discussi alcuni dei risultati più significativi ottenuti dall'applicazione dei metodi esposti (vedi Sez. 2) per il calcolo del ranking, rispettivamente delle stazioni e delle tratte ferroviarie. Per i risultati completi si rimanda alla lettura dell'appendice B.

4.1 RANKING E VALUTAZIONE DELLE STAZIONI FERROVIARIE

Applicando il metodo discusso nella Sez. 2.2.2 abbiamo ottenuto i risultati che andremo qui di seguito ad illustrare. Tali risultati sono stati utilizzati per realizzare una classifica delle stazioni in base al loro grado di esposizione al pericolo idrogeologico.

Il valore di *exposure* ottenuto per le stazioni in analisi varia da un valore minimo di 0 ad un massimo di 50. A fronte di tali risultati si è deciso di definire tre diverse fasce di esposizione al pericolo in base alle quali aggregare le stazioni all'interno della classifica:

- *High*
- *Moderate*
- *Low*

La Tabella n°3 mostra la suddetta suddivisione in fasce di rischio, gli intervalli in cui le stazioni sono state divise secondo il valore di *exposure*, il numero e la percentuale sul totale delle stazioni appartenenti ad una determinata fascia. Inoltre ad ogni fascia è stato associato un colore identificativo.

Fascia	Intervallo valori Exp_b_i	# b_i	%
High	$20 < \text{Exp_b}_i < 50$	6	5,26%
Moderate	$1 < \text{Exp_b}_i < 20$	62	54,39%
Low	$0 < \text{Exp_b}_i < 1$	46	40,35%

Tabella 3: Aggregazione delle stazioni b_i in base al valore di *exposure*

Come si evince dalla tabella 3 tale suddivisione in fasce non è lineare ma bensì frutto di un'analisi critica (vedi Sez. 4.1.1) del metodo discusso (vedi Sez. 2.2.2) e dei risultati ottenuti. Inoltre è possibile notare come la maggioranza delle stazioni presenta un livello di esposizione al rischio *Moderate* mentre solo il 5,26% presenta valori di *exposure* tali da far emergere il bisogno di controlli più frequenti (fascia *High*).

Nella tabella 4 si riportano i primi dieci record, appartenenti alla classifica delle stazioni ordinati in modo decrescente rispetto il valore di *exposure*. In riferimento alla singola stazione, per ogni record vengono quindi mostrati la posizione all'interno della classifica, l'identificativo, il nome e il suo valore di *exposure*.

n.	id Stazione	Nome Stazione Ferroviaria	Valore di <i>exposure</i>
1	114	Acciano	49,85
2	54	Fontecchio	42,68
3	44	Sant'Ilario	31,73
4	38	Pettorano Sul Gizio	31,14
5	94	Isca d'Archi	22,71
6	66	Prezza	20,54
7	33	Bussi	19,98
8	81	Capistrello	18,57
9	91	Civiluparella	16,32
10	63	Sella di Corno	14,24

Tabella 4: Classifica delle prime dieci stazioni ordinata in base al valore di *exposure*

4.1.1 *Analisi critica del Nuovo Metodo Proposto*

Osservando i risultati ottenuti dall'applicazione del metodo di cui abbiamo discusso nella Sez. 2.2.2 abbiamo potuto osservare come i valori di *exposure* delle stazioni ferroviarie non presentano una distribuzione

"omogenea/lineare" lungo l'asse dei numeri reali bensì risultano "aggregati" in gruppi di valori i quali ci hanno guidato nella definizione dei range di valori per ogni fascia che abbiamo presentato nella Tab. 3. Il NMC restituisce valori in , quindi utilizzare un'equazione lineare per aggregare le stazioni in 3 fasce in base al loro valore di *exposure* risulterebbe errato. Al fine di poter valutare le performance del metodo proposto nella Sez. 2.2.2 abbiamo confrontato la classificazione in fasce delle stazioni ottenuta in base ai valori di *exposure* con la classificazione in fasce delle stazioni ferroviarie presenti sul territorio realizzata dai membri dei gruppi partecipanti al Laboratorio di Basi di Dati II nell'A.A. 2016/2017 tramite un'indagine visiva del territorio ove tali edifici sono situati. L'ipotesi di lavoro implicita è che tale classificazione "umana", indicata d'ora in avanti come classificazione ufficiale, sia corretta per definizione. Pertanto i risultati ottenuti dal metodo discusso nella Sez. 2.2.2 saranno tanto più soddisfacenti quanto più questi si avvicineranno ai dati ufficiali.

La tabella che riporta la classificazione in fasce ufficiale delle stazioni ferroviarie nella sua completezza è presente in appendice B.3, mentre qui di seguito, in tabella

[5](#), sono stati riportati i record delle stazioni che sono stati classificati ufficialmente in fascia *High*.

id Stazione	Nome Stazione Ferroviaria
44	Sant'Ilario
65	Aversa
54	Fontecchio
38	Pettorano sul Gizio
114	Acciano
62	Vigliano d'Abruzzo

Tabella 5: Elenco delle stazioni ferroviarie ufficialmente in fascia *High*

Lo studio delle performance del metodo è stato realizzato mediante la *tabella di contingenza* (detta anche *matrice di errore* o *di confusione*). Ogni colonna di tale tabella rappresenta le istanze in una fascia "prevista" (*Predicted Classes*) mentre le righe rappresentano le istanze in una fascia reale (*Actual Classes*). La corrispondente *tabella di contingenza* (tabella 6) riassume i risultati ottenuti con il metodo.

		Predicted Classes			Totale
		High	Moderate	Low	
Actual Classes	High	4	2	0	6
	Moderate	2	53	6	61
	Low	0	7	40	47
Totale		6	62	46	114

Tabella 6: *Tabella di Contingenza* del metodo NMC discusso nella Sez. 2.2.2

Osservando i dati contenuti nella tabella 6 è possibile ricavare le seguenti deduzioni:

- Delle 6 stazioni classificate ufficialmente in fascia *High* il metodo ha previsto che 2 di queste sono in fascia *Moderate*;
- Delle 61 stazioni classificate ufficialmente in fascia *Moderate* il metodo ha previsto che 2 di queste sono in fascia *High* mentre 6 sono in fascia *Low*;
- Delle 47 stazioni classificate ufficialmente in fascia *Low* il metodo ha previsto che 7 sono in fascia *Moderate*;
- Le stazioni correttamente classificate dal nostro metodo sono localizzate lungo la diagonale della tabella.

Dall'esame della *letteratura* del settore del *Machine Learning* emerge che la maniera più semplice di effettuare misurazioni circa le performance della predizione restituita dal metodo adottato consiste nel ricondurre il problema della classificazione dei risultati ottenuti al caso di *tabella di contingenza binaria*, ossia al caso nel quale sia coinvolta

una sola classe alla volta. Il problema può essere quindi formulato come segue:

Dati n valori (v_1, v_2, \dots, v_n) e una classe (C), costruire la tabella di contingenza binaria che riassume come detti valori sono classificati sia nel mondo reale che secondo quanto stimato dall'algoritmo del quale si vuole misurare la efficacia. Evidentemente il valore v_1 potrà cadere in C oppure no, idem per v_2, \dots, v_n .

Dalla tabella 6 è possibile ricavare le seguenti *tabelle di contingenza* riferite relativamente alla fascia *High* (Tabella 7), alla fascia *Moderate* (Tabella 8) e alla fascia *Low* (Tabella 9).

		Predicted Classes		Totale
		High	Not High	
Actual Classes	High	4	2	6
	Not High	2	106	108
Totale		6	108	114

Tabella 7: Tabella di contingenza binaria riferita alla fascia *High*

Segue un commento alla tabella 7

Prima Riga:

- Il numero 4 denota le stazioni che il metodo ha classificato correttamente in fascia *High* (da questo momento chiameremo questi casi *Veri Positivi*);
- Il numero 2 denota le stazioni che il metodo **non** ha classificato **correttamente** in fascia *High* ma in un'altra fascia (da questo momento chiameremo questi casi *Falsi Negativi*).

Seconda Riga:

- Il numero 2 denota i casi nei quali l'algoritmo ha classificato **erroneamente** delle stazioni in classe *High* (da questo momento chiameremo questi casi *Falsi Positivi*);
- Il numero 106 denota le stazioni che l'algoritmo ha classificato correttamente come **non** di classe *High* (da questo momento chiameremo questi casi *Veri Negativi*).

		Predicted Classes		Totale
		Moderate	Not Moderate	
Actual Classes	Moderate	53	8	61
	Not Moderate	9	44	53
Totale		62	52	114

Tabella 8: *Tabella di contingenza binaria riferita alla fascia Moderate*

Segue un commento alla tabella 8

Prima Riga:

- Il numero 53 denota le stazioni che il metodo ha classificato correttamente in fascia *Moderate*;
- Il numero 8 denota le stazioni che il metodo **non** ha classificato **correttamente** in fascia *Moderate* ma in un'altra fascia.

Seconda Riga:

- Il numero 9 denota i casi nei quali l'algoritmo ha classificato **erroneamente** delle stazioni in classe *Moderate*;
- Il numero 44 denota le stazioni che l'algoritmo ha classificato correttamente come **non** di classe *Moderate*.

		Predicted Classes		Totale
		Low	Not Low	
Actual Classes	Low	40	7	47
	Not Low	6	61	67
Totale		46	68	114

Tabella 9: *Tabella di contingenza binaria riferita alla fascia Low*

Segue un commento alla tabella 9

Prima Riga:

- Il numero 40 denota le stazioni che il metodo ha classificato correttamente in fascia *Low*;
- Il numero 7 denota le stazioni che il metodo **non** ha classificato **correttamente** in fascia *Low* ma in un'altra fascia;

Seconda Riga:

- Il numero 6 denota i casi nei quali l'algoritmo ha classificato **erroneamente** delle stazioni in classe *Low*;

- Il numero 61 denota le stazioni che l'algoritmo ha classificato correttamente come **non** di classe *Low*.

Prima di elencare le metriche che abbiamo adottato per giudicare le performance del metodo è utile introdurre la tabella 10, nella quale andremo ad evidenziare i sei parametri che descrivono la *tabella di contingenza binaria*, ovvero: P, N, VP, FP, FN, VN, a ciascuno dei quali corrisponde un valore intero.

In letteratura sono state proposte molte metriche per giudicare le performance di un algoritmo che restituisca la realtà osservata, le più diffuse verranno elencate di seguito. Si segnala inoltre che il valore di ciascuna di esse esprime una probabilità, quindi oscilla tra 0 e 1.

		Predicted Classes		Totale
		Predizione Affermativa	Predizione Falsa	
Actual Classes	Veri Positivi (VP)	Falsi Negativi (FN)	Valori \in alla classe $P=VP + FN$	
	Falsi Positivi (FP)	Veri Negativi (VN)	Totali altri valori $N=FP+VN$	
Totale	VP + FP	FN + VN	$P + N =$ $VP + FP + FN + VN$	

Tabella 10: *Tabella di contingenza binaria "neutra"*

La **Percentuale dei Veri Positivi** (True Positive Rate) è definita come la percentuale dei casi positivi riconosciuti (dal metodo di classificazione adottato) correttamente come tali. E' auspicabile che tale valore sia prossimo a 1. In formule:

$$PVP = \frac{VP}{P} = \frac{VP}{(VP + FN)} \quad (11)$$

Dalle *tabelle di contingenza binaria* relative relativamente alla fascia *High*, alla fascia *Moderate* e alla fascia *Low* otteniamo i seguenti valori di PVP:

- PVP fascia *High* = 0,67
- PVP fascia *Moderate* = 0,87
- PVP fascia *Low* = 0,85

La **Percentuale dei Falsi Negativi** (False Negative Rate) o Percentuale dei Mancati Allarmi è definita in formule come segue:

$$PFN = \frac{FN}{P} = \frac{FN}{(VP + FN)} = 1 - PVP \quad (12)$$

Dalla formula si evince che esso è il complementare a PVP, ovvero che è auspicabile che il suo valore tenda a 0.

Dalle *tabelle di contingenza binaria* riferite relativamente alla fascia *High*, alla fascia *Moderate* e alla fascia *Low* otteniamo i seguenti valori di PFN:

- PFN fascia *High* = 0,33
- PFN fascia *Moderate* = 0,13
- PFN fascia *Low* = 0,15

La **Percentuale dei Veri Negativi** (True Negative Rate) indica la percentuale dei casi che non destano allarme. In formule:

$$PVN = \frac{VN}{N} = \frac{VN}{(FP + VN)} \quad (13)$$

Dalle *tabelle di contingenza binaria* riferite relativamente alla fascia *High*, alla fascia *Moderate* e alla fascia *Low* otteniamo i seguenti valori di PVN:

- PVN fascia *High* = 0,98
- PVN fascia *Moderate* = 0,83
- PVN fascia *Low* = 0,91

La **Percentuale dei Falsi Positivi** (False Positive Rate) indica la percentuale di falsi allarmi. In formule:

$$PFP = FP/N = FP/(FP + VN) = 1 - PVN \quad (14)$$

Dalle *tabelle di contingenza binaria* riferite relativamente alla fascia *High*, alla fascia *Moderate* e alla fascia *Low* otteniamo i seguenti valori di PFP:

- PFP fascia *High* = 0,02
- PFP fascia *Moderate* = 0,17
- PFP fascia *Low* = 0,09

La **Precisione** è definita dalla seguente formula:

$$P = \frac{VP}{(VP + FP)} \quad (15)$$

Dalle *tabelle di contingenza binaria* riferite relativamente alla fascia *High*, alla fascia *Moderate* e alla fascia *Low* otteniamo i seguenti valori di Precisione:

- Precisione fascia *High* = 0,67
- Precisione fascia *Moderate* = 0,85
- Precisione fascia *Low* = 0,87

L' **Accuratezza** è definita dalla seguente formula:

$$ACC = \frac{(VP + VN)}{(P + N)} = \frac{(VP + VN)}{((VP + FN) + (VN + FP))} \quad (16)$$

Dalle *tabelle di contingenza binaria* riferite relativamente alla fascia *High*, alla fascia *Moderate* e alla fascia *Low* otteniamo i seguenti valori di Accuratezza:

- Accuratezza fascia *High* = 0,89
- Accuratezza fascia *Moderate* = 0,85
- Accuratezza fascia *Low* = 0,89

4.1.2 *Discussione dei risultati ottenuti*

Alla luce dei risultati ottenuti dal confronto tra la classificazione in fasce ottenuta dall'elaborazione dell'output del metodo illustrato nella Sez. 2.2.2 e la classificazione ufficiale e dai valori delle metriche utilizzate per valutare le performance del metodo proposto è stato possibile concludere che l'algoritmo presenta sì alcune limitazioni che andremo a discutere qui di seguito ma che ha buone performance e riesce generalmente (nell'87% dei casi) a classificare correttamente le stazioni ferroviarie presenti sul territorio abruzzese.

Le limitazioni del metodo proposto derivano dal fatto che l'algoritmo, nel calcolare il *vettoreDirezionale*, prende in considerazione la curva di livello più vicina alla stazione ferroviaria a cui si riferisce e non sempre coincide con la direzione lungo la quale si propaga la pendenza più rilevante presente nelle vicinanze della stazione. Per maggiore chiarezza, qui di seguito, portiamo in esame due casi critici nei quali questa limitazione è ben visibile:

- **Aversa**

La stazione ferroviaria di Aversa viene classificata dal nostro metodo in fascia *Moderate* mentre *ufficialmente* viene classificata in fascia *High*. In figura 35 è visibile come il *vettoreDirezionale* viene costruito considerando la curva di livello più vicina (a 68m) al punto che indica la stazione di Aversa mentre il pendio con una pendenza maggiore si riferisce alla curva di livello che dista 93m dal punto. Ragion per cui, il valore di *exposure* ottenuto applicando il metodo proposto nella Sez. 2.2.2 per questa stazione ferroviaria è tale da classificare erroneamente tale edificio in fascia *Moderate*.

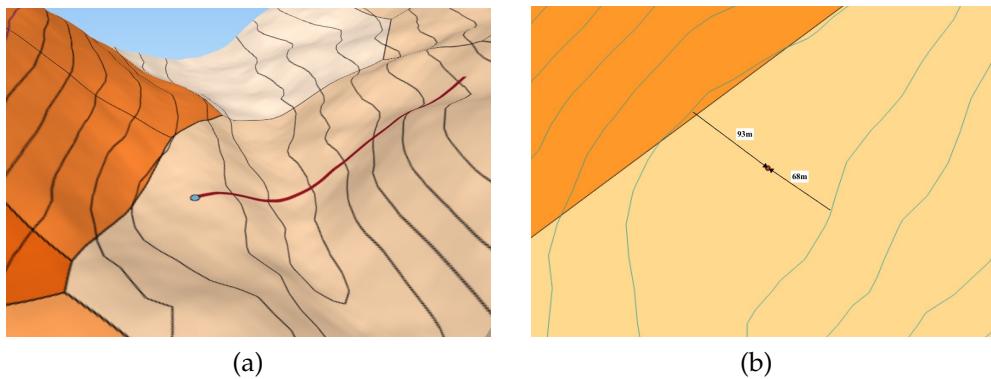


Figura 35: Stazione Ferroviaria di Aversa

- **Vigliano d'Abruzzo**

La stazione ferroviaria di Vigliano d'Abruzzo viene classificata dal nostro metodo in fascia *Moderate* mentre *ufficialmente* viene classificata in fascia *High*. In figura 36 è visibile come il *vettoreDirezionale* viene costruito considerando la curva di livello più vicina (a 38m) al punto che indica la stazione di Aversa mentre il pendio con una pendenza maggiore si riferisce alla curva di livello che dista 56m dal punto. Ragion per cui, il valore di *exposure* ottenuto applicando il metodo proposto nella Sez. 2.2.2 per questa stazione ferroviaria è tale da classificare erroneamente tale edificio in fascia *Moderate*.

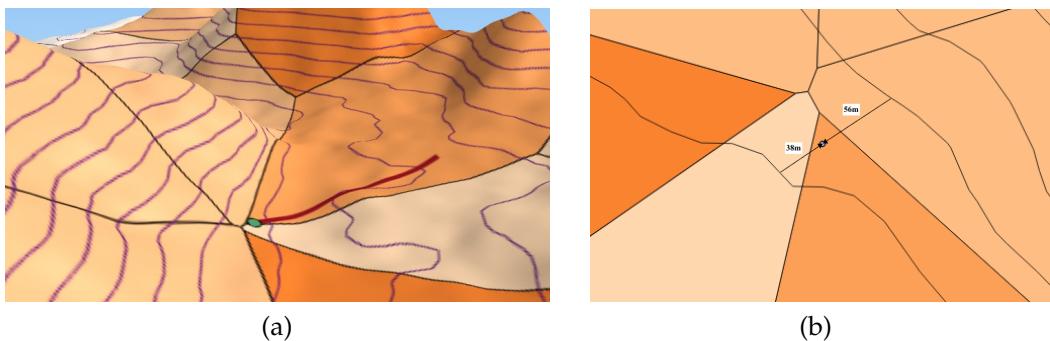


Figura 36: Stazione Ferroviaria di Vigliano d'Abruzzo

4.2 RANKING E VALUTAZIONE DELLE LINEE FERROVIARIE

4.2.1 *Discussione dei risultati ottenuti*

I risultati ottenuti applicando Il NMCE (vedi Sez. 2.3.1) per la classificazione delle linee ferroviarie sono molto positivi, infatti si è osservato che la maggior parte dei segmenti di che compongono le diverse linee ferroviarie hanno valori di *exposure* corretti. Per la validazione di questi risultati è stato utilizzato il software QGIS e un suo plugin (vedi

appendice ??) per la visualizzazione 3D di dati raster, in modo da poter verificare visivamente la correttezza dei valori di *exposure* dei segmenti calcolati dal NMCE.

In alcuni casi il valore di *exposure* dei segmenti in prossimità di una stazione sono in contrasto con la fascia assegnata a quest'ultima dal NMC. Questi casi corrispondono ai casi critici di cui abbiamo ampiamente parlato nelle sezioni 4.1.1 e 4.1.2.

Ad esempio la stazione di Aversa sappiamo essere stata classificata erroneamente nella fascia *Moderate* dal NMC. Tuttavia il NMCE classifica i segmenti in prossimità della stazione nella fascia *High* (indicata con il colore rosso) come si evince dalla Fig.37.

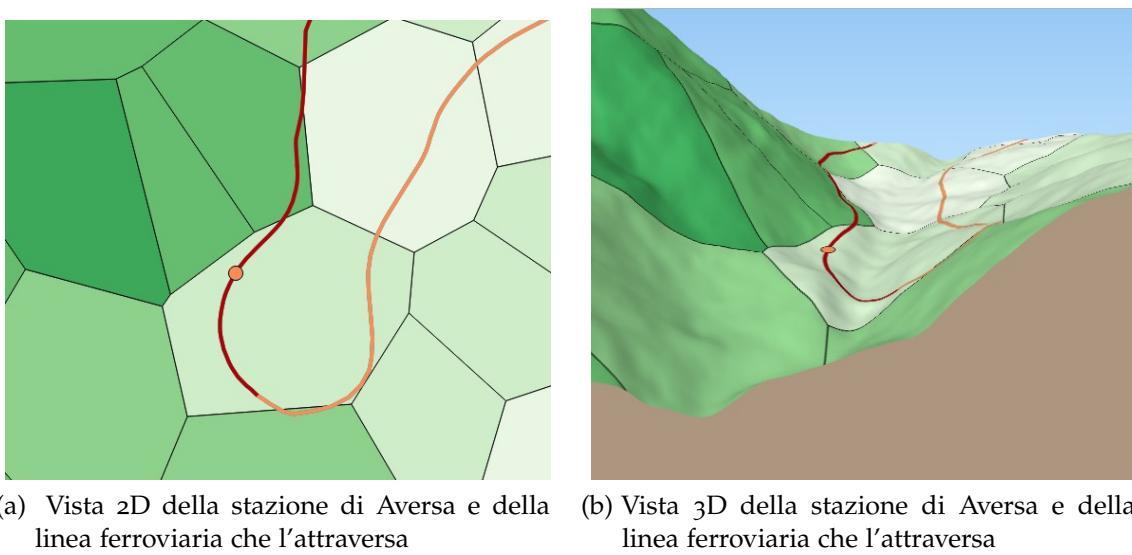


Figura 37: Vista 2D e 3D della stazione di Aversa e della linea ferroviaria che l'attraversa

CONSIDERAZIONI FINALI E PROPOSTE PER LAVORI FUTURI

In questo capitolo andremo a riportare alcune considerazioni circa i metodi discussi per il ranking delle tratte ferroviarie e delle stazioni che su di esse insistono al fine di proporre spunti per i lavori futuri. Inoltre verranno evidenziate alcune inesattezze riscontrate nei dataset utilizzati.

In conclusione è presente un riepilogo del lavoro svolto.

5.1 PROPOSTE PER LAVORI FUTURI SUL METODO DI RANKING DELLE STAZIONI FERROVIARIE

5.2 PROPOSTE PER LAVORI FUTURI SUL METODO DI RANKING DELLE TRATTE FERROVIARIE

5.3 CONSIDERAZIONI SUI DATASET

Nella valutazione delle performance dei metodi proposti non è possibile prescindere dalla consapevolezza che i dataset utilizzati potrebbero essere incompleti o potrebbero presentare imprecisioni. Uno spunto per dei lavori futuri potrebbe essere quello di utilizzare un dataset della rete ferroviaria e delle stazioni che insistono su di essa aggiornato, tale da limitare le inesattezze dei risultati. Infatti, in seguito ad accurate ricerche effettuate in rete, risulta evidente che alcune tratte, nonchè alcune stazioni situate lungo di esse, sono inattive o in fase di smantellamento. I portali che abbiamo utilizzato maggiormente sono quello proposto da Trenitalia (<http://www.rfi.it/rfi/LINEE-STAZIONI-TERRITORIO/Nelle-regioni/Abruzzo/La-rete-oggi-in:-Abruzzo>) e <http://www.ferrovieabbandonate.it/>.

Qui di seguito alcune foto di stazioni ferroviarie ormai dismesse o in disuso: Archi (Fig. 38) , Perano (Fig. 39), Pettorano sul Gizio (Fig. 40) e Civitaluparella (Fig. 41).



Figura 38: Stazione di Archi



Figura 39: Stazione di Perano



Figura 40: Stazione di Pettorano



Figura 41: Stazione di Civitaluparella

E' necessaria inoltre un'altra riflessione circa i dataset utilizzati, quella circa il dataset delle zone in cui è stato deframmentato il territorio abruzzese. Il dataset in nostro possesso è l'output di un metodo che sfrutta un algoritmo che si basa sul *Diagramma di Voronoi* per decomporre il territorio, per cui le zone ottenute potrebbero non rispecchiare in maniera coerente le caratteristiche morfologiche del terreno. Bisogna inoltre considerare che il valore numerico che quantifica la pericolosità di frana di ogni zona è strettamente legato alle curve di livello che sono presenti in tale zona e non tiene conto di quelle che sono le caratteristiche litologiche del terreno. Uno spunto per i lavori futuri potrebbe essere quello di utilizzare un dataset delle zone che presenta maggiore affidabilità circa la morfologia e le caratteristiche del territorio.

5.4 CONSIDERAZIONI FINALI

L'obiettivo di questo studio è stato fornire una classifica circa le stazioni e le linee ferroviarie in base al loro livello di esposizione al rischio frane. Sono stati proposti rispettivamente due metodi al fine di raggiungere tale scopo. Successivamente è stata progettata una base di dati spaziale atta a contenere i dataset presi in input. Nel nostro caso di studio questi ultimi hanno interessato la regione Abruzzo. La base di dati spaziale è stata implementata tramite il DBMS PostgreSQL servendoci dell'estensione spaziale PostGIS. Sono state implementate delle UDF, e definite apposite query, tramite le quali poter raggiungere l'obiettivo preposto. Dopo un'analisi dei risultati ottenuti, si è proceduto alla discussione circa le criticità dei metodi e dei dataset in input. E' opportuno sottolineare come i metodi proposti non abbiano la pretesa di restituire valori assoluti circa l'esposizione al pericolo di frane, bensì quello di fornire una metodologia valida applicabile e ovviamente migliorabile per determinare

delle classifiche grazie alle quali è possibile individuare le stazioni e le tratte ferrovia-
rie che sono maggiormente esposte al pericolo di smottamenti. I risultati finali sono
migliorabili sulla base degli spunti offerti durante la trattazione.

Parte I
APPENDICI

APPENDICE A

A.1 PROPOSTA ALTERNATIVA DI DATASET DELLE ZONE

Prima di approdare al dataset delle zone di cui abbiamo discusso nella Sez. 3.3, abbiamo ipotizzato e implementato un metodo in grado di restituirci un dataset delle zone in cui suddividere l’Abruzzo prendendo come punto di partenza il dataset delle zone prodotto dall’esecuzione della seguente *query* su Postgres:

```
CREATE TABLE GeoArea_split8 (
    id serial PRIMARY KEY,
    geom geometry (Polygon, 3004))
```

```
INSERT INTO GeoArea_split8 (geom)
SELECT ST_Subdivide(geom,8)
FROM Geo_Area
```

L’invocazione della funzione **ST_Subdivide()** sulla geometria che descrive *GeoArea* genera 24142 zone in cui è suddiviso il territorio abruzzese, di cui è possibile avere una visione d’insieme in Fig. 42. L’esame della Fig. 43 fornisce una visione più nitida

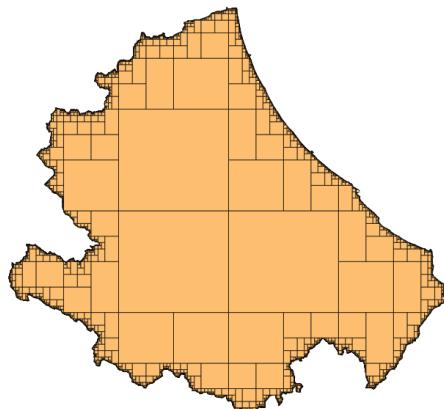


Figura 42: Visione d’insieme delle 24142 zone generate

del tipo di suddivisione effettuata da **ST_Subdivide()**. Esso in prossimità del *boundary* della *GeoArea* costruisce poligoni di forma irregolare aventi al massimo otto vertici, mentre tutte le altre aree sono dei quadrati di area crescente via via che ci si sposta verso il centro della *GeoArea*. Al fine di realizzare un’analisi quantitativa dell’esito del partizionamento della *GeoArea* si riportano in tabella 12 e in tabella 11 rispettivamente le dieci aree con area maggiore e le dieci con aree minore.

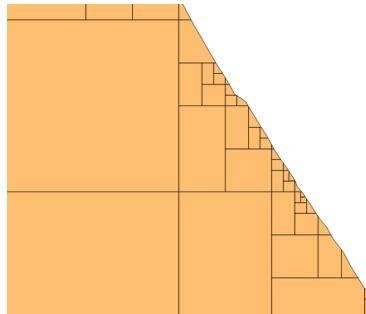


Figura 43: Particolare delle zone in prossimità del *boundary*

Area (kmq)	Perimetro (km)
$3,46 \times 10^{-11}$	$3,55 \times 10^{-5}$
$1,14 \times 10^{-10}$	$5,74 \times 10^{-5}$
$3,11 \times 10^{-10}$	$8,64 \times 10^{-5}$
$3,21 \times 10^{-10}$	$8,84 \times 10^{-5}$
$6,29 \times 10^{-10}$	$19,22 \times 10^{-5}$
$6,46 \times 10^{-10}$	$14,01 \times 10^{-5}$
$6,86 \times 10^{-10}$	$14,38 \times 10^{-5}$
$8,90 \times 10^{-10}$	$16,65 \times 10^{-5}$
$11,81 \times 10^{-10}$	$19,91 \times 10^{-5}$
$13,05 \times 10^{-10}$	$17,51 \times 10^{-5}$

Tabella 11: Top 10 aree più piccole

Come si può vedere chiaramente i valori del perimetro e dell'area delle varie zone presentano davvero molti ordini di grandezza di differenza, ragion per cui è necessario un duplice approccio di elaborazione allo scopo di ottenere un dataset delle zone più omogeneo e tale per cui le varie zone abbiano un'estensione quantomeno paragonabile tra loro:

- Frazionamenti successivi dei poligoni interni "enormi" ;
- Accorpamenti successivi dei poligoni lungo il bordo.

A.1.1 *Processamento dei poligoni interni*

Al fine di realizzare un algoritmo in grado di realizzare dei frazionamenti successivi dei poligoni interni "enormi" presenti nel dataset delle zone che abbiamo preso come punto di partenza abbiamo realizzato uno studio della funzione **ST_Subdivide()**, con lo scopo di comprendere quali sono i suoi limiti applicativi. Da questo studio è emerso che tale funzione risulta essere inapplicabile per 500 delle 24142 zone in cui abbiamo diviso il territorio, ciò ha determinato la creazione di questi poligoni interni

Area (kmq)	Perimetro (km)
1231,2	140,5
1231,2	140,5
1231,2	140,5
307,8	70,2
307,8	70,2
307,8	70,2
307,8	70,2
307,8	70,2
307,8	70,2
307,8	70,2

Tabella 12: Top 10 aree più grandi

"enormi". Questi 500 elementi presentano un'area maggiore di 0,1 kmq (valore che ci siamo prefissati come valore medio delle aree delle zone in cui dividere la *GeoArea*) e hanno numero di vertici inferiore ad 8, motivo per cui non sono processabili ulteriormente tramite la funzione **ST_Subdivide()**.

Abbiamo ricercato dunque un'altra funzione che potesse venirci in aiuto per raggiungere il nostro obiettivo e consultando il manuale PostGIS 2.3.2dev abbiamo ritenuto opportuno utilizzare la funzione **ST_Segmentize** per processare queste 500 zone "problematiche". Applicando tale funzione abbiamo ottenuto 37875 zone la cui area massima è pari a 0,0005 mq. E' stato necessario, di conseguenza, implementare un meccanismo di aggregazione di tali zone con lo scopo di ottenere zone con area media pari a 0,1 kmq. Il metodo di aggregazione che abbiamo proposto e implementato sfrutta approccio *random*: per ogni zona abbiamo utilizzato la funzione **ST_buffer()** per individuare le zone confinanti con essa, candidate ad essere aggregate con essa; Tramite la funzione **random** determino quali tra esse dovranno essere effettivamente aggregate con la zona in esame al fine di raggiungere l'obiettivo che ci siamo prefissati.

A.1.2 Processamento dei poligoni lungo il bordo

Come abbiamo già discusso nella Sez. A.1 le zone presenti sul bordo hanno un'area distante molti ordini di grandezza rispetto l'obiettivo che ci siamo prefissati. Di conseguenza è stato necessario implementare un meccanismo di aggregazione di tali zone con lo scopo di ottenere delle aree con estensione più accettabile e vicina al valore medio che ci siamo posti come obiettivo:

- Per ogni zona problematica del bordo, che andremo a chiamare z_{bordo} , andiamo a realizzare un buffer con un certo raggio (che a livello implementativo abbiamo

fissato a 150m) con lo scopo di costruire un cerchio.

- Tale cerchio verrà intersecato, tramite la funzione **ST_Intersect()**, con tutte le zone problematiche del bordo al fine di individuare quali sono le zone che intersecano la sua geometria, che andremo a chiamare $z_{cerchio}$.
- Per ogni $z_{cerchio}$, sfruttando la funzione **ST_Relate** e la definizione di *DE-9IM matrix pattern*, riusciamo a stabilire se risulta essere adiacente con la zona z_{bordo} (la funzione **ST_Relate** restituirà *true* o *false* se l'intersezione tra le due zone, rappresentate come poligoni, è una linea/geometria di dimensione 1).
- Una volta individuate tutte le zone adiacenti alla zona presa in esame, che indicheremo con $z_{adiacente}$, opero un'operazione di aggregazione ovvero unisco la z_{bordo} con la $z_{adiacente}$ più piccola. Se dall'unione risulta che l'area è ancora troppo piccola ripeto l'operazione di aggregazione.

A.1.3 Conclusioni

Le operazioni realizzate sia sul bordo che all'interno per realizzare le aggregazioni, così come le abbiamo pensate ed implementate, risultano essere molto pesanti, anzi quasi infattibili per un numero così elevato di elementi.

RISULTATI DI LABORATORIO

B.1 CLASSIFICAZIONE COMPLETA DELLE STAZIONI TRAMITE IL NMC

B.2 CLASSIFICAZIONE COMPLETA DELLE LINEE FERROVIARIE TRAMITE IL NMC

B.3 CLASSIFICAZIONE UFFICIALE DELLE STAZIONI UTILIZZATA PER LA VALIDAZIONE

Tabella 13: Range ufficiale stazioni

ID	Nome	Fascia
1	Alba Adriatica - Nereo - Controguerra	1
2	Tortoreto	1
3	Giulianova	1
5	Pineto - Atri	1
7	Montesilvano	1
8	Pescara Centrale	1
16	Casalbordino - Pollutri	1
17	Porto di Vasto	1
19	San Salvo	1
20	Teramo	1
21	San Nicolò a Tordino	1
22	Bellante - Ripattoni	1
23	Notaresco	1
24	Mosciano Sant'Angelo	1
25	Chieti	1
26	Brecciarola	1
28	Rosciano	1
29	Alanno	1
30	Scafa - San Valentino - Caramanico Terme	1
34	Popoli - Vittorito	1
35	Pratola Peligna	1
36	Sulmona	1
37	Sulmona - Introdacqua	1
42	Rivisondoli - Pescocostanzo	1

Tabella 14: Range ufficiale stazioni

ID	Nome	Fascia
46	Montenero - Valcoccchiara	1
47	Castel Di Sangro	1
48	Castel Di Sangro	1
49	Pratola Peligna Superiore	1
56	Villa Sant'Angelo	1
57	San Demetrio Vestini	1
58	Fossa	1
59	Paganica	1
61	Sassa - Torninparte	1
64	Bugnara	1
70	Collarmele	1
71	Cerchio	1
75	Avezzano	1
77	Scurcola Marsicana	1
86	Balsorano	1
87	San Pietro Avellana	1
95	Archi	1
96	Perano	1
103	Lanciano	1
106	Villa Caldari	1
107	Selceroli	1
108	Arielli	1
113	Vasto	1
4	Roseto Degli Abruzzi	2
6	Silvi	2
9	Francavilla al mare	2
10	Foro	2
11	Ortona	2
12	Ortona - Sangritana	2
13	San Vito - Lanciano	2
14	Fossacesia	2

Tabella 15: Range ufficiale stazioni

ID	Nome	Fascia
15	Torino di Sangro - Paglieta	2
18	Vasto San Salvo	2
27	Manoppello	2
31	Torre de passeri	2
32	Tocco - Castiglione	2
33	Bussi	2
39	Cansano	2
40	Campo di Giove	2
41	Palena	2
43	Roccaraso	2
45	Alfredena - Sicontrone	2
50	Raiano	2
51	Molina Aterno	2
52	Beffi	2
53	Tione degli Abruzzi	2
55	Fagnano - Campana	2
60	L'Aquila	2
63	Sella di Corno	2
66	Prezza	2
67	Goriano Sicoli	2
68	Carrito - Ortona	2
69	Pescina	2
72	Aielli	2
73	Celano - Ovindoli	2
74	Paterno - San Pelino	2
76	Cappelle Magliano	2
78	Tagliacozzo	2
79	Lanciano	2
80	Carsoli	2
81	Capistrello	2

Tabella 16: Range ufficiale stazioni

ID	Nome	Fascia
82	Canistro	2
83	Civitella Roveto	2
84	Civita D'Antino - Morino	2
85	Morrea	2
88	Ateleta	2
89	Gamberale	2
90	Quadri	2
91	Civitaluparella	2
92	Villa Santa Maria	2
93	Bomba	2
94	Isca d'Archi	2
97	Atessa	2
98	Altino	2
99	Casoli	2
100	Sant'Eusonio del Sangro	2
101	Crocetta	2
102	Castel Frentano	2
104	Treglio	2
105	San Vito Chietino	2
109	Orsogna	2
110	Filetto	2
111	Guardigrele	2
112	San Vincenzo	2
38	Pettorano Sul Gizio	3
44	Sant'Ilario	3
54	Fontecchio	3
62	Vigliano d'Abruzzo	3
65	Aversa	3
114	Acciano	3