

# APPLICAZIONE DELLA REGRESSIONE LOGISTICA PER LA ZONAZIONE DELLA SUSCETTIBILITA' DA FRANA IN AMBIENTE GIS

## APPLICATION OF LOGISTIC REGRESSION FOR LANDSLIDE SUSCEPTIBILITY ZONING IN GIS ENVIRONMENT

R. Marzocchi <sup>a</sup>, A. Rovegno <sup>b</sup>, B. Federici <sup>b</sup>, R. Bovolenta <sup>b</sup>, R. Berardi <sup>b</sup>

<sup>a</sup> Gter srl Innovazione in Geomatica, Gns e Gis, Via Greto di Cornigliano 6r, 16152 Genova - roberto.marzocchi@gter.it

<sup>b</sup> Dipartimento di Ingegneria Civile Chimica ed Ambientale (DICCA), Università degli Studi di Genova, Via Montallegro 1, 16145 Genova - (bianca.federici, rossella.bovolenta, riccardo.berardi)@unige.it

**PAROLE CHIAVE:** zonazione, suscettibilità da frana, analisi statistica, GIS GRASS

**KEY WORDS:** Landslide susceptibility zoning, statistical methods, GRASS GIS

### RIASSUNTO

Nel corso degli ultimi decenni sono state sviluppate diverse procedure per effettuare analisi della suscettibilità al dissesto franoso, a diverse scale e basati su diversi approcci; i più diffusi sono il metodo euristico e quello statistico. Il presente lavoro vuole approfondire l'utilizzo dell'analisi statistica bivariata e multivariata in ambiente GIS per la zonazione della suscettibilità. È stata analizzata l'influenza di diversi fattori (morfologici e geo-litologici del territorio, ma anche legati allo sviluppo antropico dell'area, alla copertura vegetale e al clima) sulla presenza e attivazione di frane superficiali (scivolamenti e colamenti). Inoltre è stata effettuata una revisione critica della scelta dell'area di calibrazione per estensione e caratteristiche morfologiche, climatiche ed antropiche. L'applicazione al territorio ligure ha permesso di evidenziare l'utilità del metodo per analisi su grandi estensioni territoriali, in tempi rapidi e con risorse relativamente limitate. Inoltre ha fatto emergere alcune criticità e l'importanza di un'analisi bivariata preliminare su ogni singolo fattore per valutarne la distribuzione statistica e la reale influenza sull'occorrenza di frane nell'area in studio, al fine di ottenere una corretta zonazione mediante la successiva analisi statistica multivariata.

### ABSTRACT

Over the past few decades different procedures have been developed to perform analysis of landslide susceptibility, at different scales, and based on different approaches. The most common methods are heuristic and statistical ones. The present research aims to investigate the use of GIS-based bivariate and multivariate statistical analysis for susceptibility zoning. The influence of different factors (morphology and geo-lithology of the territory, but also anthropic development, vegetation cover and climate) on the occurrence and triggering of slides and flows have been analyzed. In addition, a critical review of the choice of the calibration area by extension and morphological, climatic and anthropogenic characteristics has been performed. The application to the Ligurian territory has highlighted the usefulness of the method for analyzing large areas, quickly and with relatively limited resources. Some critical issues raised, as well as the importance of bivariate analysis of each individual factor prior to assess the statistical distribution and the real influence on the occurrence of landslides in the study area, in order to obtain a correct zoning through the next multivariate statistical analysis.

### 1. INTRODUZIONE

Circa il 10% del territorio nazionale (ISPRA, 2013) è sottoposto ad elevato rischio idrogeologico (inondazioni e frane). Nel corso degli ultimi decenni sono state sviluppate diverse procedure per effettuare analisi di previsione spaziale dei fenomeni franosi, a diverse scale e basati su diversi approcci. In particolare, per quel che concerne l'analisi di suscettibilità al dissesto da frana, si sono individuati principalmente due approcci: euristico e statistico (Varnes and IAEG 1984; Hutchinson 1995; Fell et al., 2008).

I metodi euristici si basano sul cosiddetto "giudizio esperto" della persona che conduce l'analisi, ad es. volta alla perimetrazione di frane esistenti o alla redazione di carte di suscettività. Sono generalmente basati su criteri di carattere geologico/geomorfologico e fanno largo uso di indicatori diretti e informazioni "puntuali".

I metodi statistici mirano a correlare informazioni legate a fenomeni franosi occorsi nel passato con un insieme di fattori che si suppongono responsabili, in misura diretta o indiretta, della propensione all'innesco di frane e delle loro conseguenze (Cascini, 2008). Essi risultano in genere più idonei per essere applicati a zone ampie e soprattutto diversificate e si avvalgono vantaggiosamente dell'ambiente GIS per la gestione e integrazione delle cartografie tematiche. Essi possono però richiedere processi di calcolo anche complessi (Guzzetti et al., 1999).

Nell'ambito della gestione del rischio da frana, gli autori ritengono che uno strumento in grado di eseguire analisi di suscettibilità su grandi estensioni territoriali, in tempi rapidi e con risorse relativamente limitate, possa rivelarsi di notevole aiuto.

All'interno di un Progetto di Ricerca di Interesse Nazionale, volto alla "Mitigazione del rischio da frana mediante interventi

sostenibili”, di cui una corretta zonazione della suscettibilità su area vasta è componente fondamentale, è stato approfondito un precedente lavoro (Natali & Federici, 2010) con lo scopo di meglio analizzare l'influenza di alcuni parametri sul potenziale innesco di frane superficiali.

I fattori di propensione al dissesto franoso inizialmente presi in considerazione (Natali & Federici, 2010; Bovolenta et al., 2011) erano pendenza, litotipo ed uso del suolo. Tali fattori sono anche quelli normalmente presi in considerazione nelle zonazioni effettuate a scopo di pianificazione territoriale, che spesso si dimostrano non del tutto accurate.

A seguito di un'approfondita analisi della letteratura, nonché di esperienze dirette maturate in campo, è stata individuata una serie di ulteriori indicatori di carattere geologico, geomorfologico, climatico ed antropico e sono state effettuate verifiche e correlazioni con l'Inventario dei Fenomeni Franosi in Italia (progetto IFFI), per perfezionare il modello di suscettibilità al dissesto franoso precedentemente implementato. L'intento è stato quello di comprendere l'influenza di tali fattori sulla genesi del fenomeno franoso, analizzando con attenzione le possibili criticità relative all'area di calibrazione e all'utilizzo dei fattori stessi.

I cinematismi su cui è stata incentrata l'analisi sono scivolamenti e colamenti (Glossario Internazionale delle Frane - UNESCO WP/WLI, 1993).

Le analisi sono state interamente condotte con il software free ed open source GRASS GIS (Neteler et al., 2012).

## 2. METODOLOGIA ADOTTATA

I metodi statistici correlano una mappa inventario dei fenomeni franosi accaduti nel passato con i fattori predisponenti. Nella fattispecie, in questo lavoro è stata svolta un'attenta analisi della letteratura pertinente, al fine di individuare i principali fattori predisponenti ai dissesti franosi, ed è stato scelto il modello di regressione logistica quale modello statistico per la valutazione della suscettibilità al dissesto.

### 2.1 Fattori predisponenti

Ogni tipologia di movimento franoso presenta caratteristiche peculiari ed è condizionato da fattori differenti. Per poter determinare la suscettibilità da frana in un'area vasta è necessario riconoscere i fattori che più degli altri creano situazioni favorevoli all'innesco dei cinematismi considerati e che sono quindi da tenere in considerazione nella costruzione del modello.

Come dimostrano molteplici articoli pubblicati sul tema dell'analisi della suscettibilità da frana, la scelta di tali parametri non è univoca e tantomeno codificata. La letteratura consultata è tuttavia concorde nella necessità di considerare le caratteristiche morfologiche e geo-litologiche del territorio, ma anche caratteri legati allo sviluppo antropico dell'area, alla copertura vegetale e al clima (Ayalew & Yamagishi, 2005; Cencetti et al., 2010; Dai & Lee, 2002; Dahal et al., 2008; Pauditš & Bednárík, 2002; Lee, 2007; Pradhan & Lee, 2010).

Nel presente lavoro sono state considerate tutte le categorie di fattori che, come osservato in letteratura (Tab. 1 e 2), vengono utilizzate in studi analoghi, ossia fattori geomorfologici, litologici, climatici e di copertura del suolo.

Le *caratteristiche litologiche* sono di comprovata importanza, in quanto è ragionevole pensare che le proprietà del terreno determinanti per la sua stabilità, come la resistenza e la permeabilità, siano ad essa legate. In particolare, nelle aree

potenzialmente interessate dall'innesco di frane superficiali e colate di detrito, giocano un ruolo importante i depositi di copertura. Purtroppo, però, la disponibilità di mappe che forniscono questo tipo di informazione è scarsa, almeno in Italia.

Letteratura	geologici					climatici		
	litologia	giacitura	distanza dagli allineamenti tettonici	tipo di coltre superficiale	spessore della coltre superficiale	aggressività climatica	precipitazioni	
<b>a</b>	x		x	x			x	
<b>b</b>	x							
<b>c</b>	(x)		x	x				
<b>d</b>	x			x	x			
<b>e</b>	x	x				x		
<b>f</b>	x							
<b>g</b>	x	x	x					

Letteratura	morfologici					copertura del suolo			
	pendenza	caratteristiche morfologiche	Esposizione	quota	distanza da rete di drenaggio	accumulazione	uso del suolo	vegetazione	distanza da rete stradale
<b>a</b>	x		X		x		x	x	
<b>b</b>	x			x					
<b>c</b>	x		X		x		x	x	
<b>d</b>	x		X	x		x	x		x
<b>e</b>	x						x		
<b>f</b>	x	(x)	X	x	(x)		x		
<b>g</b>	x		X	x					x

Tabelle 1 e 2. Utilizzo dei diversi fattori predisponenti in vari lavori di letteratura considerati: **a)** Pradhan & Lee, 2010; **b)** Pauditš & Bednárík, 2002; **c)** Lee, 2007; **d)** Dahal et al., 2008; **e)** Cencetti et al., 2010; **f)** Dai & Lee, 2002; **g)** Ayalew & Yamagishi, 2005 [(x) indica un debole suggerimento all'utilizzo].

Parametri quasi sempre considerati in letteratura sono quelli di tipo geomorfologico. Primo fra tutti la *pendenza*, ma anche la direzione in cui è orientato il versante, ossia la sua *esposizione* e il valore di *quota* che in alcuni regioni si è rivelato particolarmente significativo per determinare alcune caratteristiche (es. presenza o meno di vegetazione) che statisticamente possano determinare una maggiore o minore suscettibilità da frana. Infine la direzione e la quantità di deflusso superficiale e sotterraneo sono sicuramente fattori fortemente correlati a dissesti franosi superficiali come quelli oggetto del presente studio. Il processo con cui, dalle zone morfologicamente convesse, l'acqua di pioggia va concentrandosi nelle zone morfologicamente concave è noto come *accumulazione* e il suo valore numerico è una misura dell'area di terreno che fornisce acqua superficiale ad un'area dove l'acqua si accumula. Tale parametro consente quindi di

definire le zone di maggior accumulo dell'acqua di pioggia e quindi maggiormente suscettibili al dissesto.

Fattori antropici come l'uso del suolo e la distanza dalla rete stradale possono essere anch'essi correlabili a dissesti franosi di tipo superficiale. Nel caso di strade in zone montuose gli scavi per la loro realizzazione possono costituire un elemento critico dal punto di vista meccanico e soprattutto le strade stesse possono influire sul comportamento idrogeologico del versante modificando talvolta in maniera significativa le modalità di ruscellamento e/o infiltrazione delle acque piovane.

Infine, le precipitazioni brevi e intense e quelle eccezionali o prolungate sono fattori molto importanti per l'innescare dei fenomeni d'instabilità dei versanti. Un indice di aggressività climatica di facile applicabilità, adatto alla realtà del territorio nazionale e valido per molte tipologie di dissesto, è l'indice di Fournier modificato da Arnoldus (Arnoldus, 1977; Cencetti et al., 2010), indicato come  $F_{FAO}$  e ricavabile dalla relazione:

$$F_{FAO} = \frac{\sum_{i=1}^{12} P_i^2}{P} \quad (1)$$

dove  $P$  (mm) è la precipitazione media annua e  $p_i$  (mm) è la precipitazione media del mese  $i$ -esimo. Uno studio relativo a 292 località italiane (Scrini et al., 2006) ha dimostrato che l'indice  $F_{FAO}$  è molto ben correlato con altre variabili climatiche associabili all'insorgenza o alla riattivazione di dissesti idrogeologici di varia natura ed è quindi in grado di fornire una valutazione sintetica della probabilità di eventi piovosi notevoli, sia intensi di breve durata che diluiti nel tempo.

## 2.2 Il modello di regressione logistica

Il metodo della regressione logistica (Guzzetti et al., 1999; Dai & Lee, 2002; Chen & Wang, 2007) è un caso particolare di modello lineare generalizzato, particolarmente adatto nei casi in cui la variabile dipendente  $Y$  sia dicotomica.

Nella regressione logistica, la suscettibilità di un'area viene rappresentata dalla variabile dipendente  $Y$ , a cui si associa il valore 1 o 0 per indicare rispettivamente la presenza o l'assenza di frana. Lo scopo dell'analisi è quello di stimare la probabilità che un'area sia classificata nell'una o nell'altra condizione (assenza o presenza di frana,  $Y=0$  o  $Y=1$ ). Siccome le due condizioni sono mutuamente esclusive, se è nota la probabilità di ricadere in una categoria, è nota anche la probabilità di ricadere nell'altra. La probabilità condizionata che la frana occorra è quindi data da:

$$P = [1 / X_i] \quad (2)$$

dove il vettore  $X$  rappresenta il vettore dei diversi fattori predisponenti  $X_i$ , detti anche regressori, predittori, o variabili indipendenti. Tale probabilità, nel modello di regressione logistica, viene espressa con la seguente formulazione:

$$z = \text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = b_0 + \sum_{i=1}^n b_i X_i \quad (3)$$

con  $b_i$  i coefficienti di regressione.

Tale relazione permette dunque di correlare le variabili indipendenti  $X_i$ , con la funzione *logit* e, di conseguenza, con la probabilità  $p$  che si verifichi l'evento "occorrenza del fenomeno franoso":

$$p = \frac{e^z}{1 + e^z} \quad (4)$$

Tale analisi statistica è stata effettuata in ambiente GIS.

I GIS sono, per loro natura, ambienti particolarmente adatti per condurre una procedura di zonazione della suscettibilità, come

suggerito dalle Linee Guida del JTC-1 (Fell et al., 2008). Un GIS permette, infatti, di gestire agevolmente le informazioni georiferite relative ad estese porzioni di territorio, consente di interrogare ed intersecare tra loro diversi livelli informativi e di rendere i risultati facilmente leggibili ed immediatamente utilizzabili ai fini della pianificazione territoriale. Nel corso dell'analisi, è stato utilizzato il software free ed open source GRASS GIS (Geographic Resource Analysis Support System), nella sua versione 7.0 (Neteler et al., 2012) per la sua elevata capacità di elaborazione e per la possibilità di realizzare procedure automatiche *ad hoc* per lo specifico problema. Il software GRASS permette la gestione e l'elaborazione di dati digitali in formato sia raster che vettoriale, mediante un vasto set di comandi a disposizione che permette di eseguire in tempi rapidi diverse elaborazioni. Nella fattispecie, per il calcolo dei coefficienti di regressione  $b$ , è stato utilizzato il comando *r.regression.multi*, che utilizza il metodo dei minimi quadrati ed è adatto a modelli lineari come quello ottenuto con la funzione *logit*. Siccome esso è concepito per l'analisi di regressione multipla lineare, la calibrazione è eseguita sul modello di regressione linearizzato dalla trasformazione *logit*. Il comando richiede in input le mappe raster dei predittori  $X_i$  e la mappa *logit* della probabilità di frana osservata sulla base dell'inventario dei fenomeni franosi (cartografia IFFI) nell'area scelta per la calibrazione, associando rispettivamente probabilità 1 alle aree in frana (scivolamenti e colamenti), 0 alle restanti aree. L'area da utilizzare per la fase di calibrazione non deve necessariamente coincidere con l'area su cui si vuole stimare la suscettibilità, bensì può esserne un sottoinsieme. La dimensione e le caratteristiche dell'area scelta influenzano il risultato finale, come si illustrerà nel seguito.

I coefficienti di regressione  $b$ , stimati a valle della fase di calibrazione del modello, vengono utilizzati per stimare la funzione *logit*( $p$ ) e, di conseguenza, la probabilità  $p$  che si verifichi l'evento "occorrenza del fenomeno franoso", in formato raster, teoricamente applicando il modello a qualsiasi porzione e/o estensione territoriale. La procedura restituisce in output anche la mappa raster dei residui (differenza tra i valori di *logit* reali e stimati) ed un file con i valori numerici dei coefficienti di regressione  $b$ , il valore dell'indice *AIC* - Akaike Information Criterion (Akaike, 1973) e del coefficiente di correlazione  $R^2$  per ciascuna variabile indipendente  $X_i$ .

L'indice  $R^2$  quantifica la bontà di adattamento del modello di regressione lineare ai dati. Un valore uguale a 1 indica un adattamento perfetto del modello ai dati nel caso di un modello di regressione lineare. Quanto più  $R^2$  si avvicina a 1, tanto meglio i valori stimati approssimeranno i valori osservati.

Nel caso in cui la probabilità del fenomeno dipenda da più di una variabile indipendente, si può determinare un valore di  $R_i^2$  relativo a ciascuna variabile indipendente  $X_i$ . In tal caso, esso rappresenta l'aumento di varianza condizionata dovuta all'inserimento di quella variabile nel modello rispetto a quando non se ne tiene conto. In altre parole,  $R_i^2$  rappresenta una misura di quanto l'introduzione della variabile  $i$ -esima contribuisca a migliorare l'adattamento del modello ai dati osservati. Si osservi che il coefficiente di determinazione  $R_i^2$  ha il limite di essere una funzione non decrescente del numero di regressori e pertanto non risulta adatto ad effettuare confronti tra modelli a diverso numero di regressori.

Per il confronto fra modelli è stato utilizzato l'indice *AIC* - Akaike Information Criterion (Akaike, 1973). Tale indice è basato sul concetto di entropia dell'informazione ed offre una misura relativa delle informazioni perse quando un dato modello è usato per descrivere la realtà. Nel caso generale è definito come:

$$AIC = 2k - 2\ln(L) \quad (5)$$

dove  $k$  è il numero di parametri del modello statistico e  $L$  è il valore massimizzato della funzione di verosimiglianza del modello stimato. La regola è quella di preferire sempre i modelli con  $AIC$  più basso.

Nella regressione logistica, il modello è tanto più completo quanto maggiore è il numero di variabili indipendenti  $X_i$  considerate, ma solo a condizione che queste abbiano effettivamente un ruolo influente nel determinare il valore della variabile dipendente  $Y$ . La scelta delle variabili più significative da inserire nel modello di regressione è stata quindi effettuata utilizzando il metodo *stepwise*, basato sul confronto dei valori di extradevianza delle diverse configurazioni del modello, usando l'indice  $AIC$  per avere una misura della qualità della stima del modello statistico, tenendo conto sia della bontà di adattamento che della complessità del modello stesso.

In generale, le variabili indipendenti  $X_i$  devono essere *operative* (avere un certo grado di affinità con la variabile dipendente), *complete* (abbastanza ben rappresentate su tutta l'area di studio), *non-uniformi* (variabili spazialmente), *misurabili* (poter essere espresse in una delle scale di misurazione: nominale, ordinale o cardinale) e *non-ridondanti* (non dovrebbero incidere allo stesso modo su due differenti conseguenze).

Scopo del presente lavoro è stato, appunto, quello di ottenere utili indicazioni circa la scelta ottimale delle variabili, aspetto fortemente critico in un metodo statistico. In particolare, il modello è stato calibrato su regioni diverse per estensione e caratteristiche fisiche, antropiche e climatiche e le variabili statistiche sono state criticamente analizzate. Nella fattispecie, i valori dei coefficienti di regressione  $b_i$ , dell'indice  $AIC$  e del coefficiente  $R_i^2$  sono stati oggetto di un'attenta analisi, confrontando i risultati della calibrazione sulle diverse aree, come illustrato nei paragrafi seguenti.

### 3. AREA DI STUDIO E DATI UTILIZZATI

L'analisi di suscettibilità al dissesto franoso è stata condotta calibrando la procedura sull'intera regione Liguria e poi distintamente sulle 4 provincie di Imperia, Savona, Genova, La Spezia (Fig. 1), utilizzando la carta dei confini amministrativi reperibile sul sito dell'ISTAT (Istituto Nazionale di Statistica).

Come dato fondamentale per la calibrazione, è stata utilizzata la carta Inventario dei Fenomeni Franosi in Italia - IFFI (APAT, 2007). I dati nazionali sono attualmente gestiti dalle singole Regioni e, in particolare, quelli liguri sono reperibili dal Portale Cartografico Regionale e aggiornati al 2012. Dall'inventario sono state selezionate le "aree soggette a frane superficiali" e i "colamenti rapidi", ossia le tipologie di movimento a cui si rivolge la procedura di zonazione della suscettibilità descritta e in base alle quali si vuole calibrare il modello di previsione. La mappa vettoriale è stata rasterizzata e riclassificata, assegnando il valore 1 alle celle in frana e 0 a tutte le altre, creando così la mappa *logit* (Fig 1).

Per quanto concerne i parametri geomorfologici, essi sono stati ricavati da un modello digitale del terreno (DTM) con risoluzione 20x20 m, liberamente scaricabile dal sito dell'Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA).

Per le caratteristiche litologiche è stata utilizzata una cartografia vettoriale geo-litologica (scala 1:100.000) fornita dal Settore Ambiente della Regione Liguria, data l'indisponibilità di dati a scala di maggior dettaglio (es. carte del progetto CARG) in formato vettoriale e quindi facilmente utilizzabili per le procedure di classificazione.

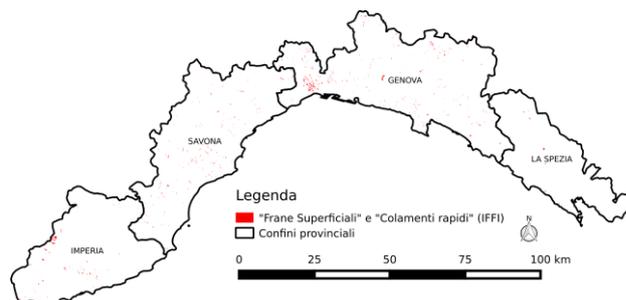


Figura 1. Mappa *logit* per la Regione Liguria, con indicazione dei confini amministrativi delle 4 Province. In rosso sono indicate le aree definite come "Frane superficiali" e "Colamenti rapidi" all'interno dell'inventario dei fenomeni franosi (cartografia IFFI), a cui è stata assegnata probabilità pari 1; le restanti aree, in bianco, hanno invece probabilità pari a 0.

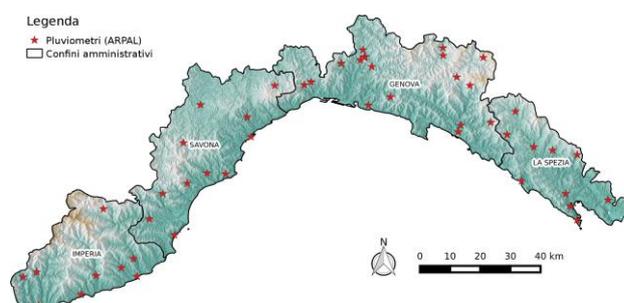


Figura 2. La distribuzione dei pluviometri utilizzati per il calcolo dell'aggressività climatica (ARPAL, 2012).

Sia la cartografia di uso del suolo, a scala 1:10.000 e aggiornata al 2012, sia la cartografia della rete infrastrutturale, anch'essa a scala 1:10.000 aggiornata al 2003, sono state reperite presso il Portale Cartografico Regionale in formato vettoriale.

Infine, per la valutazione dell'aggressività climatica, sono stati estratti dall'Atlante climatico della Liguria (ARPAL, 2012) sia la posizione geografica sia i dati osservati dai pluviometri con almeno 10 anni di misure di precipitazione disponibili (Fig. 2). I valori di aggressività climatica dei punti campionati sono stati quindi interpolati per ottenere una copertura completa dell'area di studio, con risoluzione spaziale di 1x1 km. Per individuare l'interpolatore ottimale sono stati applicati differenti metodi di interpolazione: l'inverso della distanza al quadrato (*IDW* - Inverse Distance Weighted), utilizzando i tre o quattro punti più vicini, le *spline* regolarizzate con la tensione (*RST* - Regularized Splines with Tension), il diagramma o tassellazione di Voronoi (noto anche come poligoni di Thiessen) e l'interpolatore *NN* (Natural Neighbors). Tra questi, è stato infine scelto di applicare l'algoritmo di Watson per l'interpolazione Natural Neighbor di Sibson (Sibson, 1980, Sibson, 1981; Watson, 1992), attraverso il comando di GRASS *r.surf.nmbathy*. Il metodo si basa sulla triangolazione di Delaunay e definisce il valore in ogni punto come media dei valori dei punti campionati vicini, pesati in base alla porzione di area utilizzata di ogni poligono di Thiessen. L'algoritmo è affidabile per dataset molto ampi e dati campionati in maniera irregolare, ma fornisce un risultato solamente nell'area compresa tra i punti noti. Per ottenere la copertura dell'intera area regionale è stata completata la porzione mancante calcolando i poligoni di Thiessen, ossia assegnando a ciascun punto il valore del punto campionato più vicino. La mappa dell'aggressività climatica finale è rappresentata in Figura 3.

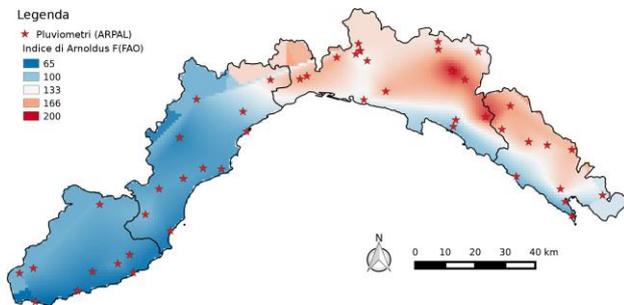


Figura 3. Mappa dell'aggressività climatica.

#### 4. L'INFLUENZA DEI DIVERSI FATTORI

Le variabili messe in gioco sono sia di tipo numerico che di tipo categoriale, ossia variabili i cui valori sono informazioni di tipo qualitativo. Sono variabili quantitative, ad esempio, la quota, la pendenza, l'accumulazione o la distanza dalle strade, mentre il litotipo, l'uso del suolo e l'esposizione si pongono tra le variabili categoriali. Ai fini dell'applicazione dell'analisi statistica, i valori possibili di ciascuna variabile sono stati suddivisi in classi identificate da un numero intero progressivo a partire da 1.

Per le variabili numeriche, le classi corrispondono a sotto intervalli di tutti i possibili valori della variabile; per le variabili categoriali, ogni classe raccoglie invece categorie simili. Per esempio, l'aggressività climatica viene suddivisa in 3 classi ( $F_{FAO} < 100$ ,  $100 < F_{FAO} < 200$  e  $F_{FAO} > 200$ ), la quota in classi di 250 m s.l.m., il logaritmo della mappa di accumulazioni in 6 classi che vanno da 1 (spartiacque) a valori superiori a 6 (reticolo idrografico o zona umida).

Nel caso delle variabili categoriali, la suddivisione in classi pone il problema di decidere l'ordine secondo il quale numerarle. L'adozione di un modello di regressione logistica comporta, infatti, la stima del valore di *logit* attraverso una funzione lineare delle variabili indipendenti  $X_i$ ; siccome il valore di *logit* così stimato varia linearmente all'aumentare della variabile, anche la suscettibilità, espressa come funzione esponenziale di *logit*, avrà un andamento monotono al crescere del valore della variabile (andamento crescente o decrescente in base al segno del coefficiente di regressione  $b$ ). Tale andamento monotono deve però essere effettivamente riscontrato sui valori osservati: al progressivo aumento del valore della variabile (ad esempio, incrementando progressivamente la classe di uso del suolo da 1 al suo valore massimo), dovrebbe effettivamente corrispondere una tendenza costante della probabilità di dissesto all'aumento o al decremento. Se questo non si osserva, l'adozione della relazione lineare tra *logit* e la variabile risulterebbe inappropriata, portando, di fatto, ad una stima del coefficiente di regressione poco rappresentativa della reale influenza che la variabile possiede nel determinare la suscettibilità. Per valutare quindi l'influenza della singola classe di un parametro sulla probabilità di dissesto, è stata prevista una fase di analisi statistica bivariata preliminare all'assegnazione delle classi.

La suscettibilità da dissesto franoso corrisponde alla probabilità condizionata che si verifichi l'evento "occorrenza del fenomeno franoso" data una combinazione di fattori. Se si considera l'occorrenza del fenomeno come dipendente da una sola variabile, la probabilità condizionata è definita dal teorema di Bayes:

$$P(\text{evento}|X) = \frac{P(\text{evento} \cap X)}{P(X)} \quad (6)$$

La probabilità condizionata da ciascuna classe di una variabile indipendente è stata calcolata sulla base dei risultati ottenuti col comando di GRASS *r.stats*. In input sono fornite la mappa di una delle variabili indipendenti, suddivisa in classi numerate secondo un ordine casuale e la mappa delle aree in frana superficiale secondo la cartografia IFFI, i cui valori sono rispettivamente 1 o 0 e rappresentano la suscettibilità osservata. In questo modo sono stati conseguiti due risultati distinti:

- riduzione del numero di classi (si veda in Tab. 3 la riclassificazione delle mappe di uso del suolo e litotipo);
- ordinamento monotono delle classi per i fattori predittivi di tipo categoriale.

Litologia		Uso del suolo	
1	Depositi	1	Altro
2	Rocce magmatiche	2	Edificato
3	Rocce metamorfiche	3	Aree agricole
4	Rocce sedimentarie	4	Bosco
5	Altro	5	Vegetazione sparsa
		6	Roccia affiorante o vegetazione assente

Tabella 3. Riclassificazione delle mappe di litologia ed uso del suolo.

In certi casi, tuttavia, si può osservare, anche per le variabili numeriche, una mancata corrispondenza tra l'ordine delle classi e l'ordine di influenza delle stesse. Si osservi, ad esempio, l'andamento della probabilità di evento franoso condizionata dalla variabile quota (Fig. 4); essendo però tali variabili numeriche dotate di ordinamento proprio, si è ritenuto di mantenere quest'ultimo, nonostante possa determinare un andamento non monotono della suscettibilità.

Per ogni predittore, l'analisi statistica bivariata ha quindi individuato la presenza o meno di una correlazione con l'accadimento del fenomeno franoso.

Nelle seguenti Figure 5-12 sono mostrate le mappe così riclassificate sulla base dell'analisi statistica bivariata.

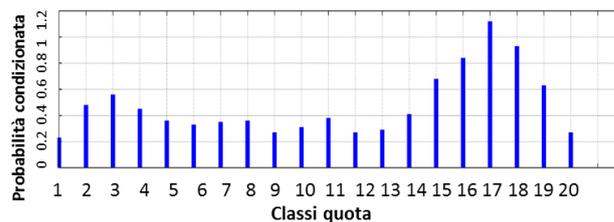


Figura 4. Probabilità condizionata dell'evento "occorrenza del fenomeno franoso" data dalla quota.

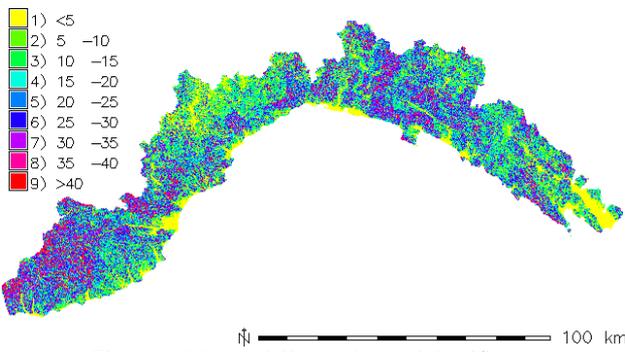


Figura 5. Mappa delle pendenze riclassificata.

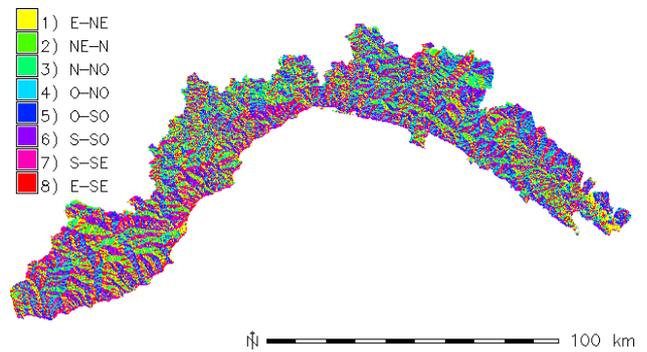


Figura 9. Mappa dell'esposizione riclassificata.

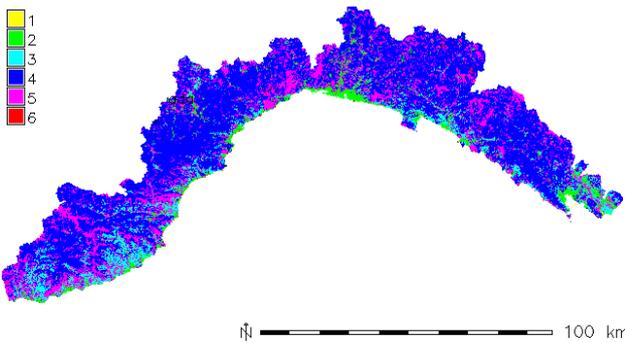


Figura 6. Mappa dell'uso del suolo riclassificata (si veda la Tab. 3 per la corrispondenza tra numero della classe e tipologia di suolo).

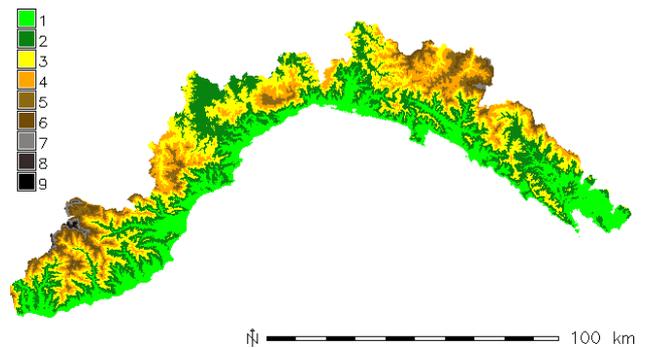


Figura 10. Mappa delle quote riclassificata (classi di 250 m, dal livello medio del mare a 2250 m s.l.m.).

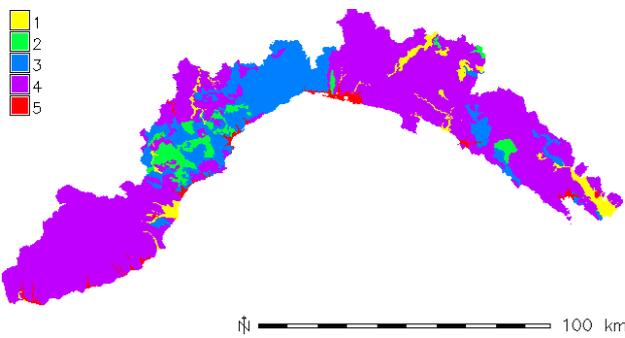


Figura 7. Mappa dei litotipi riclassificata (si veda la Tab. 3 per la corrispondenza tra numero della classe e litotipo).

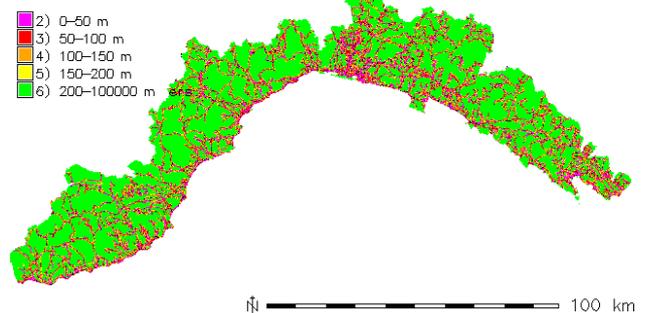


Figura 11. Mappa della distanza dalle strade riclassificata.

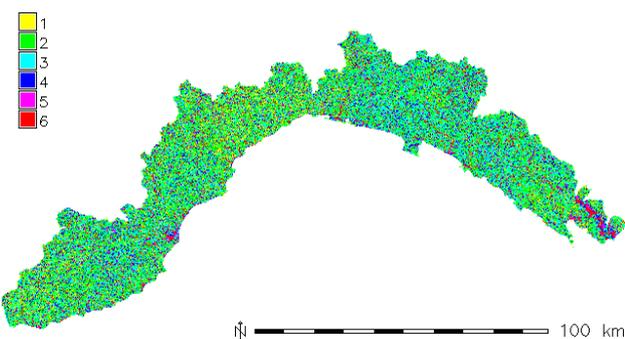


Figura 8. Mappa del Log (accumulazione) riclassificata.

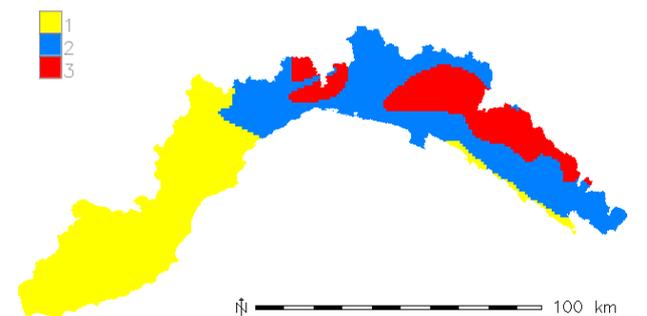


Figura 12. Mappa dell'aggressività climatica riclassificata (la classe 1 corrisponde a  $F_{FAO} < 100$ ; la classe 2 a  $100 < F_{FAO} < 200$ ; la classe 3 a  $F_{FAO} > 200$ ).

## 5. L'INFLUENZA DELL'AREA DI CALIBRAZIONE

Effettuata l'analisi statistica bivariata ed individuata la classificazione ottimale di tutti i fattori, si è proceduto all'analisi statistica multivariata con il metodo della regressione logistica, inserendo gli otto predittori, e valutando per ogni modello l'indice  $AIC$  e, per ogni predittore incluso nel modello, i coefficienti  $b_i$  e  $R_i^2$ . Il modello è stato prima applicato all'intero territorio regionale e poi alle singole province.

Relativamente all'intero territorio regionale, il migliore modello si è rivelato quello che tiene conto di tutti gli 8 suddetti parametri, come mostrato in Figura 13, dove è raffigurato l'andamento di  $AIC$  all'inserimento dei vari predittori. Al contrario, per alcune province (Imperia e La Spezia), si è rilevato migliore il modello che considera solo 7 variabili indipendenti, escludendo, nella fattispecie, l'aggressività climatica, perché statisticamente poco significativa, essendo scarsamente variabile, come mostrato in Figura 12.

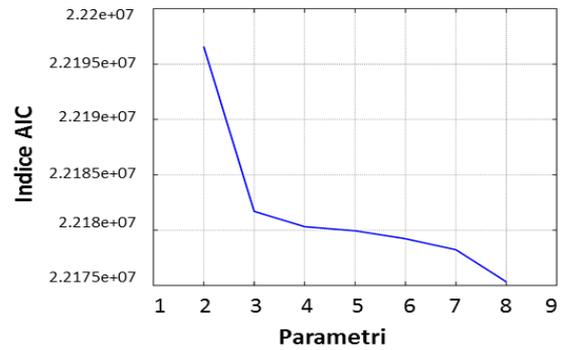


Figura 13. Andamento di  $AIC$  per la calibrazione sulla regione Liguria. Si può denotare una sostanziale decrescita all'aumentare del numero dei parametri.

I grafici delle figure 14 e 15 mostrano l'andamento dei valori dei coefficienti  $b_i$  e  $R_i^2$  per i modelli ottenuti effettuando la calibrazione prima sull'intera regione e poi sulle singole province, ossia al variare dell'area di calibrazione.

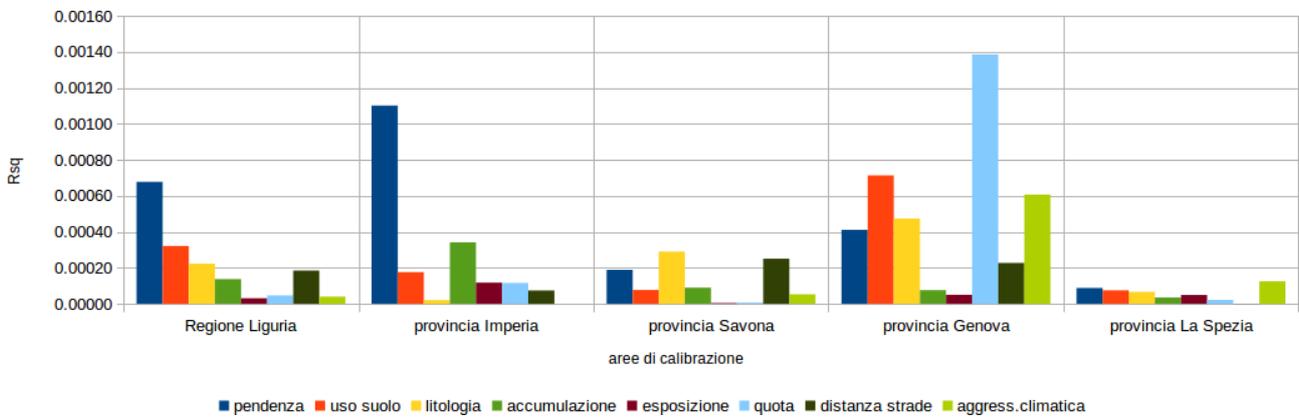


Figura 14 – Confronto dei valori di  $R^2$  per i vari predittori al variare dell'area di calibrazione

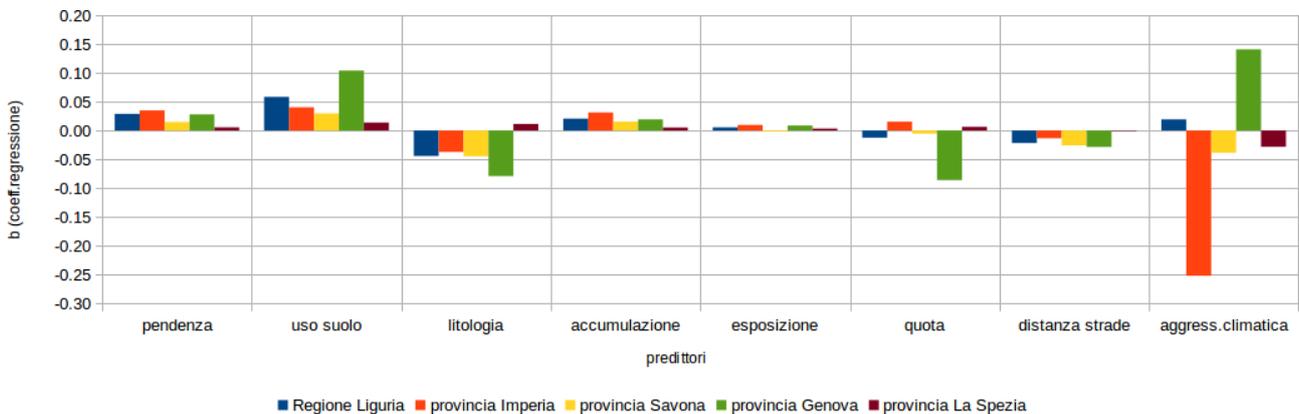


Figura 15 – Confronto dei valori dei coefficienti di regressione  $b$  per ciascun predittore al variare dell'area di calibrazione

$R_i^2$  e il valore assoluto di  $b_i$  danno un'utile indicazione circa l'influenza di ciascun predittore sulla mappa di suscettibilità. Il confronto fra le diverse aree di calibrazione ha permesso di valutare comportamenti coerenti o meno dei diversi predittori e quindi di valutarne indirettamente l'affidabilità prevista per le diverse aree di calibrazione. Per esempio, pendenza, uso del suolo e accumulazione sembrano avere un comportamento stabile, sia per entità che per influenza (positiva o negativa); un

valore stabile si ha anche per la variabile “distanza dalle strade”, sebbene i coefficienti applicati abbiano sempre valori molto bassi, mentre strani comportamenti si hanno per quota, esposizione dei versanti, litotipo e aggressività climatica.

La variabile quota presenta comportamenti fortemente variabili, in funzione delle diverse aree di calibrazione. Effettivamente, almeno per la Liguria, essa sembra essere poco influente nei confronti dell'occorrenza di frane sul territorio, in quanto

scarsamente correlata ad altri fattori climatici, morfologici o antropici, come invece può accadere in altre aree geografiche (es. area alpina).

Per quel che concerne l'esposizione, come si evince dall'analisi bivariata (Fig. 16), questo parametro, pur quantitativo, dovrebbe comunque essere trattato come i parametri categoriali. Infatti, considerare 8 diverse classi, come stabilito inizialmente, appare eccessivo e ciò è testimoniato sia dalla bassa probabilità condizionata, sia dall'andamento non monotono della probabilità condizionata al variare delle classi.

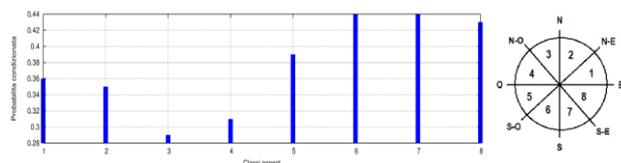


Figura 16. Suddivisione iniziale della variabile esposizione in 8 classi

È sembrato pertanto ragionevole raggruppare le classi, così da ottenere una suddivisione in settori più significativa per la regione di interesse. È stata quindi applicata la procedura sul territorio della regione Liguria, valutando il miglioramento che si ottiene riducendo il numero di classi della variabile esposizione da 8 a 4 e a 2, secondo due diverse classificazioni (Fig. 17):

- divisione nei quattro quadranti principali, ordinati in senso antiorario a partire dal quadrante di Nord-Est;
- divisione in due sole classi, comprendenti rispettivamente i quadranti a Nord e i quadranti a Sud.

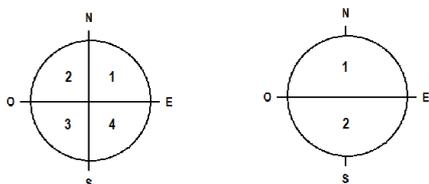


Figura 17. Modifiche alla classificazione della mappa dell'esposizione dei versanti.

La decrescita dell'indice  $AIC$  suggerisce che, almeno per la regione Liguria, il modello che considera una suddivisione in 2 sole classi è quello che meglio considera l'influenza dell'esposizione dei versanti sulla suscettibilità da frana. Appare evidente come per diverse aree di calibrazione andrebbe comunque analizzato l'andamento della probabilità condizionata e quindi valutata la migliore riclassificazione del parametro esposizione in funzione delle caratteristiche regionali.

Per quel che concerne la variabile litotipo essa sembrerebbe poco o altamente influente in funzione delle diverse aree di calibrazione, come dimostrato dall'alta variabilità del coefficiente di correlazione  $R^2$ . Questo sembra probabilmente dovuto al basso grado di dettaglio della cartografia utilizzata e quindi poco utile per valutare una reale correlazione al fenomeno franoso, in quanto assolutamente non in grado di segnalare la locale presenza di depositi o altre formazioni geologiche di carattere locale che abbiano un'influenza sull'occorrenza delle frane. È nostra opinione che l'eventuale disponibilità di mappe geologiche in formato vettoriale a scala più alta (1:25.000) e/o di informazioni stratigrafiche (database geologico dei sondaggi geognostici) potrebbero indubbiamente essere molto più indicati per l'applicazione della procedura.

Infine l'aggressività climatica  $F_{FAO}$  sembra essere un buon parametro da considerare nel caso di aree di calibrazione estese con caratteristiche climatiche eterogenee, come nel caso dell'applicazione sull'intero territorio regionale, mentre risulta assolutamente fuorviante se applicato ad aree più ristrette con caratteristiche climatiche omogenee (es. provincia di Imperia o di La Spezia).

## 6. CONCLUSIONI

A conclusione del presente lavoro, si può indubbiamente confermare che l'analisi statistica multivariata condotta in ambiente GIS rappresenta un valido e pratico strumento per eseguire la zonazione della suscettibilità da frana di vaste estensioni territoriali, in tempi rapidi e con risorse limitate.

Il metodo consente, infatti, di tener conto dei numerosi fattori che influenzano la suscettibilità e di correlarli ad essa in modo oggettivo. L'utilizzo di un software GIS permette inoltre di gestire agevolmente le diverse informazioni georiferite relative a vaste aree e di ottenere un prodotto finale facilmente leggibile, immediatamente utilizzabile ai fini della pianificazione territoriale e che si presta ad essere rapidamente aggiornato, quando necessario.

L'applicazione al territorio ligure e soprattutto l'analisi effettuata confrontando diverse aree di calibrazione per estensione e caratteristiche morfologiche, climatiche ed antropiche, ha altresì permesso di esaminare alcuni aspetti di maggiore criticità e soprattutto l'importanza di un'analisi bivariata preliminare su ogni singolo fattore per valutare la distribuzione statistica, ossia la reale influenza sull'occorrenza di frane ed ottenere una corretta classificazione ai fini della successiva analisi statistica multivariata.

I prossimi passi della ricerca saranno un'estensione dell'applicazione della procedura su aree geografiche limitrofe, ma con caratteristiche differenti, per confrontare i risultati e valutare il diverso comportamento di alcuni parametri morfologici (quota ed esposizione dei versanti) e dell'aggressività climatica. Attualmente le analisi si stanno estendendo alla regione Piemonte che presenta una forte variabilità morfologica e climatica, essendo caratterizzata da estese aree pianeggianti (piana del Po), collinari (Langhe e Monferrato), appenniniche ed infine alpine.

Scopo finale della ricerca sarà quello di redigere linee guida volte a supportare gli operatori del settore (Enti Pubblici, professionisti) che intendano avvalersi di metodi statistici in ambiente GIS per la zonazione della suscettibilità da frana.

## RINGRAZIAMENTI

Il presente lavoro, sviluppato dall'Unità di Ricerca di Genova, è stato finanziato con il contributo di finanziamenti del Progetto di Ricerca di Interesse Nazionale - PRIN 2010-2011 su la "Mitigazione del rischio da frana mediante interventi sostenibili", coordinato dall'Università degli Studi di Salerno.

## BIBLIOGRAFIA

Akaike H., 1973. Information theory and an extension of the maximum likelihood principle. In: Petrov B. N. and Csáki F. (eds.), *2nd International Symposium on Information Theory*, Akadémia Kiadó, Budapest, 267-281.

APAT, 2007. Rapporto sulle frane in Italia. Il Progetto IFFI - Metodologia, risultati e rapporti regionali APAT, Rapporti 78/2007 ISBN 978-88-448-0310-0.

- Arnoldus H.M.J., 1977. Methodology used to determine the maximum potential average annual soil loss due to sheet and rill erosion in Morocco. *FAO Soil bulletin* 34:39-48.
- ARPAL 2012. Atlante climatico della Liguria <http://www.arpal.gov.it/homepage/meteo/analisi-climatologiche/atlante-climatico-della-liguria.html> (7 Apr. 2015).
- Ayalew L. & Yamagishi H., 2005. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. *Geomorphology* 65:15-31.
- Bovolenta R., Federici B. & Natali D., 2011. Realizzazione automatica di carte di pericolosità da frana superficiale utilizzando il GIS open source GRASS. - X Giornata Mondiale dell'Acqua. Frane e Dissesto Idrogeologico: consuntivo. Atti dei Convegni Lincei, Accademia Nazionale dei Lincei, Roma, ISSN 0391805; ISBN 9788821810367, 169-177.
- Cascini L., 2008. Applicability of landslide susceptibility and hazard zoning at different scales. *Engineering Geology*. 102:164-177.
- Cencetti C., De Rosa P., Fredduzzi A. & Minelli A., 2010. La valutazione della suscettibilità da frana tramite i metodi dell'Analisi Condizionale e degli Indici di Franosità: un'applicazione tramite GRASS al bacino del T. Ventia (Umbria settentrionale). *Geomatic Workbooks*, 9: 41-54.
- Chen H. & Wang J., 2007, "Landslide hazard mapping using logistic regression model in Mackenzie Valley, Canada", *Natural Hazards*, 42(75-89).
- Dahal R.K., Hasegawa S., Nonomura A., Yamanaka M., Masuda T. & Nishino K., 2008. GIS-based weights-of-evidence modelling of rainfall-induced landslides in small catchments for landslide susceptibility mapping, *Environmental Geology* 54 (2): 314-324, DOI:10.1007/s00254-007-0818-3.
- Dai F.C. & Lee C.F., 2002. Landslide characteristics and slope instability modeling using GIS, Lantau Island, Hong Kong *Geomorphology* 42:213-228.
- Fell R., Corominas J., Bonnard C.H., Cascini L., Leroi E. & Savage B., 2008. Guidelines for landslide susceptibility, hazard and risk zoning for land use planning. *Engineering Geology*, 102: 85-98, ISSN 0013-7952.
- Guzzetti F., Carrara A., Cardinali M. & Reichenbach P., 1999. Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. *Geomorphology* 31:181-216.
- Hutchinson J.N., 1995. Landslide hazard assessment. In: *Proc VI Int Symp on Landslides*, Christchurch, 1: 1805-1842.
- ISPRA, 2013. Linee guida per la valutazione del dissesto idrogeologico e la sua mitigazione attraverso misure e interventi in campo agricolo e forestale. ISPRA, Manuali e Linee Guida 85/2013 ISBN 978-88-448-0586-9.
- Lee S., 2007. Application and verification of fuzzy algebraic operators to landslide susceptibility mapping *Environ Geol* 52:615-623.
- Natali D. & Federici B., 2010. Realizzazione di carte della pericolosità da frana su aree estese: una procedura automatica in GRASS, *Geomatic Workbooks* 9:121-126.
- Neteler M., Bowman M.H., Landa M. & Metz M., 2012. GRASS GIS: a multi-purpose Open Source GIS. *Environmental Modelling & Software*, 31: 124-130.
- Pauditš P. & Bednárík M., 2002. Using GRASS in evaluation of landslide susceptibility in Handlovská kotlina Basin. *Proceedings of the Open source GIS - GRASS Users Conference, 2002 - Trento, Italy, 11-13 September*.
- Pradhan B. & Lee S., 2010. Regional landslide susceptibility analysis using back-propagation neural network model at Cameron Highland, Malaysia *Landslides* 7:13-30.
- Scrinzi G, Gregori E, Giannetti F, Galvagni D, Zorn G, Colle G & Andrenelli M., 2006. Un modello di valutazione della funzionalità protettiva del bosco per la pianificazione forestale: la componente stabilità dei versanti rispetto ai fenomeni franosi superficiali. *Foresta* 3 (1): 98-155.
- Sibson R., 1980. A vector identity for the Dirichlet tessellation. In *Mathematical Proceedings of the Cambridge Philosophical Society*, 87: 151-155.
- Sibson R. 1981. A brief description of natural neighbour interpolation. In Barnett V. (ed.), *Interpreting Multivariate Data*, Wiley, New York, USA, 21-36.
- Varnes, D.J. and the International Association of Engineering Geology Commission on Landslides and Other Mass Movements, 1984. *Landslide hazard zonation: A review of principles and practice*. Natural Hazards, vol.3, Paris, France. UNESCO, 63 pp.
- Watson D.F., 1992. *Contouring: A Guide to the Analysis and Display of Spatial Data*. Pergamon Press, Oxford, UK, 321 pp.
- WP/WLI-International Geotechnical Societies UNESCO Working Party on World Landslide Inventory (1993). *Multilingual Landslide Glossary*. Bitech Publisher, Richmond, British Columbia, Canada, 59 pp.