

# *Classificazione del comportamento torrentizio dei bacini della Provincia di Trento*

## **Obiettivo del lavoro**

L'obiettivo di questo lavoro di ricerca è quello di sviluppare un metodo di classificazione dei piccoli bacini alpini a carattere torrentizio ubicati in provincia di Trento. La classificazione si basa sulla probabilità che vengano generati fenomeni torrentizi intensi, quali colate detritiche o correnti iperconcentrate, capaci di raggiungere l'area di conoide.

## **Introduzione**

La storia della gestione dei bacini idrografici in Tirolo risale al XIII secolo. Già nel 1277 si parla di danni al muro di protezione del torrente Talvera presso Bolzano. Mentre nel 1400 iniziò la costruzione dei muri di sponda lungo il torrente Fersina, finanziati con un terzo dei dazi sui prodotti forestali.

Una delle pietre miliari del controllo dei torrenti nelle Alpi è rappresentata dalla costruzione della serra di Pont'Alto nel 1537, voluta dal principe vescovo Bernado Clesio per proteggere la città di Trento. Si tratta di un passo importante nella gestione dei processi torrentizi: fu la prima volta che una struttura di mitigazione venne progettata per operare nella parte alta di un bacino idrografico. Fino a quel momento, tutti gli sforzi compiuti erano rivolti a proteggere passivamente case, infrastrutture e aree agricole. Dalla sua prima costruzione,

la serra di Pont'Alto venne distrutta varie volte dal torrente Fersina. L'ultima ricostruzione fu eseguita nel 1850, con un'altezza di circa 40 m. La stabilità del manufatto fu assicurata con la costruzione della Controserra Madruzzo, una controbriglia ai piedi dell'opera principale, conclusa nel 1882.

Nel territorio della Monarchia asburgica, dopo i distruttivi eventi alluvionali del 16-20 settembre e del 27 ottobre 1882, iniziò l'istituzione di un servizio tecnico-forestale (k.k. forsttechnische Abteilung für Wildbachverbauungen), completato nel 1884. Come già avvenuto in Francia, questo servizio adottò una gestione integrale dei bacini idrografici attraverso la combinazione di misure forestali e tecniche. In questo modo il concetto di protezione fu esteso all'intero bacino idrografico, applicando una gestione integrale (BM für Land- und Forstwirtschaft, 1984). Soluzione innovativa e quanto mai attuale.

Un altro evento da menzionare è l'alluvione del 1966. Si verificò il 4 novembre 1966 e produsse danni significativi soprattutto a Trento e alla parte orientale della provincia. Le pesanti conseguenze sull'economia provinciale e l'instabilità del territorio portarono a una forte campagna di costruzione di misure di mitigazione, che si protrasse fino agli anni '80.

Oggi, l'obiettivo principale dei servizi provinciali è quello di applicare le nuove conoscenze scientifiche per ottimizzare l'efficienza della gestione dei bacini idrografici. Proprio in quest'ottica si posizionano studi di questo tipo.

## Torrenti alpini e processi di trasporto

Per definire un torrente alpino è necessario descrivere molte caratteristiche. Una definizione generale ma esaustiva è riportata in BERGMEISTER *et al.*, 2009: un torrente alpino è un corso d'acqua naturale, con flusso idrico durevole o temporaneo. Il suo corso è almeno localmente ripido e la portata è fortemente variabile. L'erosione dei sedimenti avviene per mezzo di piene brevi e intense con un'elevata capacità di trasporto. L'erosione avviene sull'intero bacino idrografico e lungo il corso d'acqua. La deposizione dei sedimenti può avvenire sia all'esterno che all'interno degli argini. Un bacino torrentizio generale è formato da un bacino di drenaggio, l'area che raccoglie le precipitazioni, da una zona di trasporto e da un conoide di deiezione, dove si depositano la massa d'acqua e i sedimenti trasportati. Tutto ciò che accade in un torrente è dovuto alla rapida risposta idrologica del bacino e quindi all'improvviso aumento dell'onda di piena. A ciò sono collegati i processi di mobilitazione dei sedimenti, di trasporto e di conseguente deposizione.

Processi torrentizi molto intensi, come le colate detritiche e le correnti iperconcentrate, possono portare fino al fondovalle enormi quantità di materiale in pochi minuti. Il trasporto di fondo è meno intenso, ma gli eventi di lunga durata sono responsabili della mobilitazione di grandi quantità di sedimenti. Soprattutto per la gestione del rischio e la progettazione di strutture di mitigazione, la conoscenza approfondita dei processi di trasporto è di fondamentale importanza.

La granulometria del materiale movimentato differisce fortemente tra i vari processi considerati. Molti autori hanno proposto la classificazione dei diversi fenomeni in uno dei seguenti tipi: trasporto in sospensione, trasporto di fondo, corrente iperconcentrata e colata detritica. Oltre a questi, un torrente può trasportare materiale come carico disciolto, cioè costituito da rocce solubili, o per fluitazione di particelle di legno, rifiuti, serbatoi, ecc..

Il carico in sospensione e il trasporto di fondo sono tipi di trasporto selettivo, poiché solo una determinata granulometria di sedimenti viene trasportata dal flusso. Al contrario, le colate detritiche sono processi di trasporto non selettivi, poiché qualsiasi classe diametrica

viene spostata indiscriminatamente dal flusso. Una corrente iperconcentrata rappresenta un fenomeno intermedio tra le precedenti. Le colate detritiche sono caratterizzate da un regime di flusso non newtoniano nel quale i detriti e l'acqua scorrono con la stessa velocità come una miscela. L'elevata mobilità, il potenziale erosivo e la forza d'impatto di una colata detritica portano a considerare questo fenomeno torrentizio come uno dei più critici per la sicurezza di persone e infrastrutture.

## I bacini di debris flow

Un processo di debris flow si verifica generalmente in bacini caratterizzati da una piccola estensione e da un grande rilievo. Considerando un immaginario bacino da colata detritica, è possibile identificare tre aree principali: una zona di innesco, una zona di trasporto ed erosione e una zona di deposizione. VANDINE (1996) afferma che per l'innesco di una colata detritica granulare è necessaria una pendenza superiore a 25°. Il trasporto e l'erosione richiedono una pendenza superiore a 15°. Una prima deposizione parziale, sotto forma di argini detritici, avviene quando la pendenza del canale è inferiore a 15° o il grado di confinamento del flusso diminuisce fortemente. La deposizione su di un conoide avviene di solito quando la pendenza del letto del canale scende sotto la soglia dei 10°. I valori di pendenza possono variare a seconda del confinamento del flusso, delle caratteristiche dei sedimenti (matrice, litologia, geometria, granulometria, ecc.), del contenuto d'acqua e del tipo di flusso.

## Predisposizione dei bacini alle colate detritiche

Tre fattori principali influenzano la suscettibilità o la disposizione al verificarsi di colate detritiche in un bacino torrentizio (Iverson 1997):

- ripidità dei pendii e del letto del torrente (zone di potenziale innesco);
- disponibilità di sedimenti sciolti o non consolidati (rocce alterate, morene quaternarie, terreni alluvionali, ecc.);

- caratteristiche idrologiche delle potenziali zone di innesco (sufficiente disponibilità di acqua).

ZIMMERMANN *et al.*, 1997, hanno fornito una chiara spiegazione del meccanismo di formazione delle colate detritiche in un determinato bacino torrentizio. Per ogni bacino esiste una disposizione di base e una disposizione variabile, la somma dei due termini rappresenta l'effettiva disposizione all'innesco di una colata detritica. Gli eventi di pioggia o i periodi di scioglimento della neve rappresentano lo stress del sistema. Un evento di debris flow può essere innescato solo quando lo stress e la disposizione del bacino superano determinati valori, che non sono costanti ma variano nel tempo. La disposizione di base è invariabile o varia solo a lungo termine (da decenni a secoli), ed esprime la suscettibilità generale del sistema torrentizio a eventi di debris flow. La localizzazione e la magnitudo di un possibile evento sono fortemente correlate alla disposizione di base del bacino. Variabilità di medio termine (da giorni a decenni) influenzano invece la disposizione variabile del bacino, come le precipitazioni, la presenza di ghiaccio e neve, le condizioni idrologiche e i periodi di vegetazione, definendo la frequenza delle colate detritiche oltre che avendo un certo controllo anche sulla magnitudo.

### **I dati disponibili per la provincia di Trento**

La disponibilità di banche dati di notevole interesse per il territorio della provincia di Trento ha permesso di impostare un lavoro di analisi sulle stesse. L'utilizzo di tecniche di data mining e machine learning consente di ricavare da questi database complessi informazioni altrimenti poco o per nulla visibili. Tutti i dati sono stati forniti dal Servizio bacini montani (di seguito SBM) della Provincia autonoma di Trento (di seguito PAT), che li compila e li aggiorna al fine di disporre di materiali e fonti informative utilizzabili.

Il primo database analizzato riguarda i bacini idrografici e contiene informazioni, sotto forma di misure e indici, sulle caratteristiche geomorfologiche, geologiche e climatiche dei bacini. L'identificazione e la descrizione dei bacini idrografici sono state condotte dal SBM

attraverso un processo automatizzato e si basa su un DEM (Digital Elevation Model) della PAT con risoluzione di 10 metri. Un codice, scritto in R, estrae il reticolo idrografico provinciale e definisce un codice identificativo univoco del torrente. Attraverso una procedura GIS, per ogni torrente vengono identificati il bacino idrografico e il conoide ad esso collegato. In questo modo è possibile creare un sistema duale bacino-conoide caratterizzato da un codice torrente comune. Si ha così la possibilità di valutare, separatamente o in maniera combinata, le informazioni sulle caratteristiche geomorfologiche del conoide, del bacino e del relativo torrente. È importante segnalare che questo scripting considera solo i bacini aventi un conoide ben riconoscibile. I corsi d'acqua con conoidi erosi, a causa della grande capacità di trasporto del fiume di fondovalle o di un fondovalle molto stretto, non vengono quindi rilevati dalla procedura automatizzata. Considerando che questi database sono costruiti per descrivere la probabilità di generare processi torrentizi intensi, quali le colate detritiche, sono rappresentati solo piccoli bacini alpini. Per questo motivo, sono stati esclusi anche i bacini con un bacino idrografico superiore a 10 km<sup>2</sup>, considerati inclini a generare solo portate liquide o eventi di trasporto di fondo. Queste esclusioni portano a una rappresentazione parziale dell'intera area provinciale. Tale fattispecie non è da considerarsi problematica in quanto l'analisi si concentra solo sui processi che potenzialmente raggiungono l'area del conoide, dove sono generalmente presenti infrastrutture e persone. Allo stesso tempo, questa errata rappresentazione del territorio reale può portare a difficoltà e incertezze nella creazione dei modelli. In totale il sistema elaborato dal SBM porta a rilevare 1697 bacini idrografici.

Il database è costituito dai 20 descrittori delle caratteristiche del bacino idrografico elencati nella Tabella 1. La maggior parte di essi riportano informazioni e indici geomorfologici, ma sono presenti anche quantili di precipitazione e informazioni geologiche. Per la descrizione statistica dei dati e per una precisa definizione dei parametri disponibili si rimanda all'elaborato originale.

Il secondo database a disposizione è relativo

Tabella 1 – Caratteri descrittivi dei bacini idrografici.

| Indice     | Unità              | Descrizione  |
|------------|--------------------|--|
| classid    | -                  | Codice identificativo del bacino (lo stesso del conoide collegato) |
| barea      | m <sup>2</sup>     | Superficie del bacino  |
| bperim     | m                  | Perimetro del bacino   |
| bhmin      | m s.l.m.           | Altitudine minima  |
| bhmax      | m s.l.m.           | Altitudine massima   |
| bhmed      | m s.l.m.           | Altitudine media   |
| bslo_min   | °                  | Pendenza minima  |
| bslo_max   | °                  | Pendenza massima   |
| bslo_med   | °                  | Pendenza media   |
| bgravelius | -                  | Indice di Gravelius del bacino                                     |
| bicompatt  | -                  | Indice compattezza del bacino                                      |
| bmelton    | -                  | Numero di Melton del bacino  |
| bintipsom  | -                  | Integrale ipsometrico del bacino                                   |
| bstadio    | -                  | Indice di stadio del bacino  |
| big        | -                  | Indice geologico del bacino (da D'Agostino 1996)                   |
| big1045mea | -                  | Indice geologico del bacino delle zone di innesco                  |
| bquantmm   | mm/h               | Quantile di precipitazione centenaria di durata 1 h                |
| bdensdren  | km/km <sup>2</sup> | Densità di drenaggio del bacino                                    |
| blungasta  | m                  | Lunghezza asta torrentizia   |
| bslo_asta  | °                  | Pendenza media dell'asta torrentizia                               |
| bffactor   | -                  | Fattore di forma del bacino  |

alle caratteristiche geomorfologiche delle aree di conoide cartografate dal SBM. Il metodo precedentemente descritto per creare il database ha identificato un numero totale di 2584 conoidi sull'intero territorio provinciale. Solo i conoidi direttamente collegati a uno dei 1697 bacini idrografici identificati in precedenza vengono utilizzati per l'analisi. L'elenco degli 8 indici per la descrizione delle caratteristiche morfologiche dei conoidi è presente in Tabel-

la 2. Le modalità di calcolo dei parametri e la descrizione statistica dei dati sono disponibili nell'elaborato originale al quale si rimanda.

Il terzo database utilizzato riporta informazioni sugli eventi verificatisi all'interno del territorio provinciale, raccolte e descritte secondo diverse fonti e con varia attendibilità. Questo catasto comprende ben 3522 eventi relativi all'attività di torrenti, fiumi e laghi, avvenuti su bacini e/o conoidi di deiezione. Sono riportati fe-

Tabella 2 – Caratteri descrittivi dei bacini idrografici.

| Indice     | Unità          | Descrizione  |
|------------|----------------|--|
| classid    | -              | Codice identificativo del conoide                        |
| ambito_cnr | -              | Codice del bacino di 1° ordine                           |
| id_opera   | -              | Codice identificativo di eventuali opere di mitigazione  |
| danno_pote | -              | Indice delle infrastrutture potenzialmente danneggiabili |
| rasta      | -              | Codice identificativo del torrente                       |
| cslo_med   | °              | Pendenza media del conoide                               |
| carea      | m <sup>2</sup> | Area del conoide   |
| evol       | m <sup>3</sup> | Volume stimato del conoide                               |

nomeni a partire dall'anno 1570, principalmente dal 1750, fino al 2016. È facile immaginare che la risoluzione spaziale, così come la completezza e l'affidabilità delle informazioni dipendono fortemente dalla finestra temporale considerata e dalla fonte dei dati. In generale, l'incertezza sulla veridicità delle informazioni cresce con la distanza temporale. Le informazioni disponibili provengono inoltre da diverse fonti, caratterizzate da diversa precisione e affidabilità. Le fonti informative possono essere riassunte nei seguenti gruppi:

- scheda profilo eventi, informazioni sugli eventi registrate dal SBM per ogni evento noto verificatosi negli ultimi 10-15 anni;
- banca dati ARCA, eventi descritti dal Servizio Geologico della PAT, esclusi i movimenti di massa come frane, caduta massi, frane profonde, valanghe di roccia e altro (molti eventi sono classificati come frane ma probabilmente legati all'attività torrentizia);
- cronistoria degli eventi, riepilogo descrittivo degli eventi passati, incluso nel progetto delle misure di mitigazione per un determinato bacino idrografico;
- articoli di giornali e bollettini locali;
- altre fonti di dati esterne, come ad esempio i registri e gli archivi comunitari e parrocchiali.

Un catasto eventi di questo tipo non può assicurare la completezza delle serie di eventi

di un bacino, soprattutto considerando bacini non occupati da importanti attività umane (uso residenziale, industriale, turistico o agricolo) o infrastrutture (ferrovie, strade e altro), come la maggior parte dei bacini alpini o i bacini completamente coperti da foreste. A questo si aggiunge che gli eventi di minore magnitudo e intensità spesso non vengono registrati, così come gran parte degli eventi che si fermano nella zona di trasporto e non raggiungono l'area del conoide. Oltre agli eventi non registrati, bisogna considerare anche il conteggio multiplo di alcuni di questi eventi, che generano danni in diversi punti dell'area di deposizione. Un limite cruciale è anche la definizione di ogni processo, che in molti casi non è precisa e univoca. Ciò è dovuto principalmente alla natura instabile dei fenomeni (ad esempio, la continua transizione da colata detritica a trasporto di fondo, passando per una corrente iperconcentrata e viceversa lungo un singolo canale torrentizio) e alle diverse conoscenze e obiettivi dei ricercatori. Anche l'indicazione di una magnitudo, quando presente, non è omogeneamente affidabile. Tutte queste considerazioni sulle incertezze implicano la necessità di una revisione e di un filtraggio incisivo degli eventi utilizzati per la costruzione, il fitting e la validazione dei modelli. Per ogni singolo elemento del catasto eventi del SBM, sono disponibili le informazioni riportate in Tabella 3.

Tabella 3 – Informazioni presenti nel catasto eventi del SBM.

| <b>Indice</b> | <b>Descrizione</b>   |
|---------------|--|
| id_evsto      | Codice identificativo evento                                   |
| tipoevento    | Tipo di evento (per es. piena liquida, colata detritica, ecc.) |
| att_tipoev    | Attendibilità dell'indice <i>tipoevento</i>                    |
| dataevento    | Data di accadimento dell'evento                                |
| att_dataev    | Attendibilità dell'indice <i>dataevento</i>                    |
| localita      | Località di accadimento dell'evento                            |
| att_locali    | Attendibilità dell'indice <i>localita</i>                      |
| descrizion    | Descrizione dell'evento  |
| danni         | Descrizione dei danni  |
| opere_post    | Opere di mitigazione costruite dopo l'accadimento dell'evento  |
| fonte         | Fonte informativa utilizzata per il record                     |
| corsoacqua    | Nome del corso d'acqua   |
| note          | Note varie   |

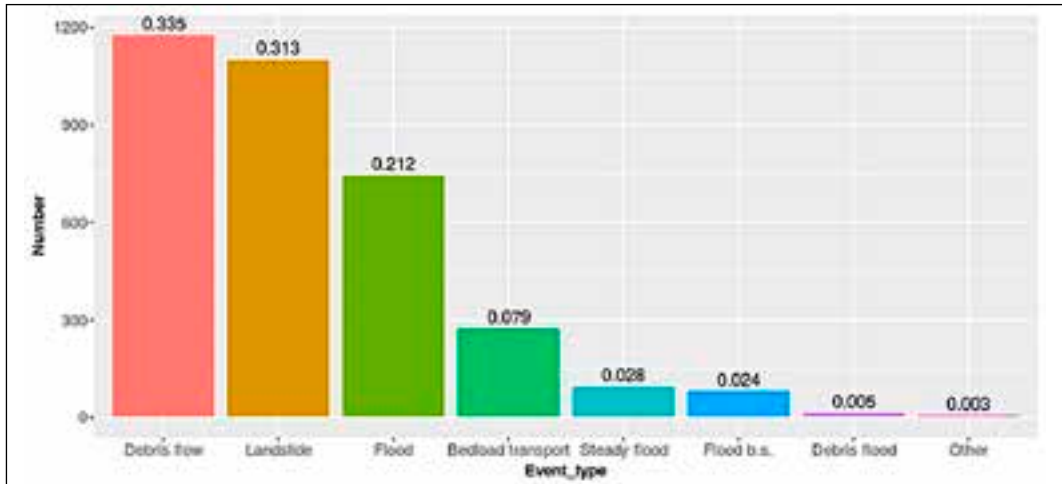


Figura 1 – Classificazione dei record presenti nel catasto eventi SBM (i numeri sopra le colonne indicano la frequenza relativa).

I 3522 record contenuti nel catasto eventi sono suddivisi in 8 diversi tipi di processo. Come rappresentato graficamente in Figura 1. Un terzo del totale è costituito da fenomeni di debris flow e un ulteriore 31% è rappresentato da frane. Questi ultimi eventi sono descritti in modo grossolano e la loro affidabilità è generalmente bassa. Con ogni probabilità non si tratta di frane vere e proprie, ma più probabilmente di processi torrentizi non ben definiti o di eventi misti. In questo senso un'importante fonte di incertezza è dovuta alla mancanza dell'uso del nome tecnico di colata detritica, o del suo sinonimo inglese debris flow, nella parte italiana delle Alpi. Marchi e Cavalli (2007) ricordano che all'epoca dell'alluvione del 1966 e negli anni immediatamente successivi la maggior parte delle colate detritiche non erano riconosciute e analizzate come tali in Italia. L'errata o completa omissione della classificazione dei processi come colate detritiche nella parte italiana delle Alpi è stata sottolineata anche da Castiglioni (1971), mentre il termine specifico Mure è sempre stato ben consolidato ad esempio nella letteratura austriaca.

### L'analisi dei dati e il preprocessing

Tutte le analisi di seguito descritte sono state eseguite principalmente in ambiente R, uti-

lizzando il software RStudio. Il software QGIS è stato utilizzato nella fase iniziale di selezione degli eventi e per la rappresentazione dei risultati dell'analisi.

Sui database disponibili viene utilizzata la tecnica della Principal component analysis seguita poi da un modello di regressione logistica. Il modello non deve essere solo creato, ma anche testato. Per realizzare questi due compiti, è necessario filtrare i dati inaffidabili e dividere il database risultante in sottoinsiemi separati. Il modello di regressione logistica viene quindi addestrato e testato per poi venire applicato all'intero database. Alcune delle fasi comuni della costruzione del modello sono:

- pre-elaborazione dei dati predittivi;
- stima dei parametri del modello;
- selezione dei predittori;
- valutazione delle prestazioni;
- messa a punto delle regole di classificazione.

La prima fase dell'analisi consiste nel filtrare i dati degli eventi ed è stata eseguita in ambiente GIS. Sono stati filtrati gli eventi torrentizi aventi buona affidabilità per quanto riguarda tipo di processo e localizzazione, ottenendo così un database di 2338 eventi (sui 3522 dell'intero database). Visto l'interesse ai processi legati ai torrenti non vengono considerati i record classificati come frane, valanghe o piene fluviali.

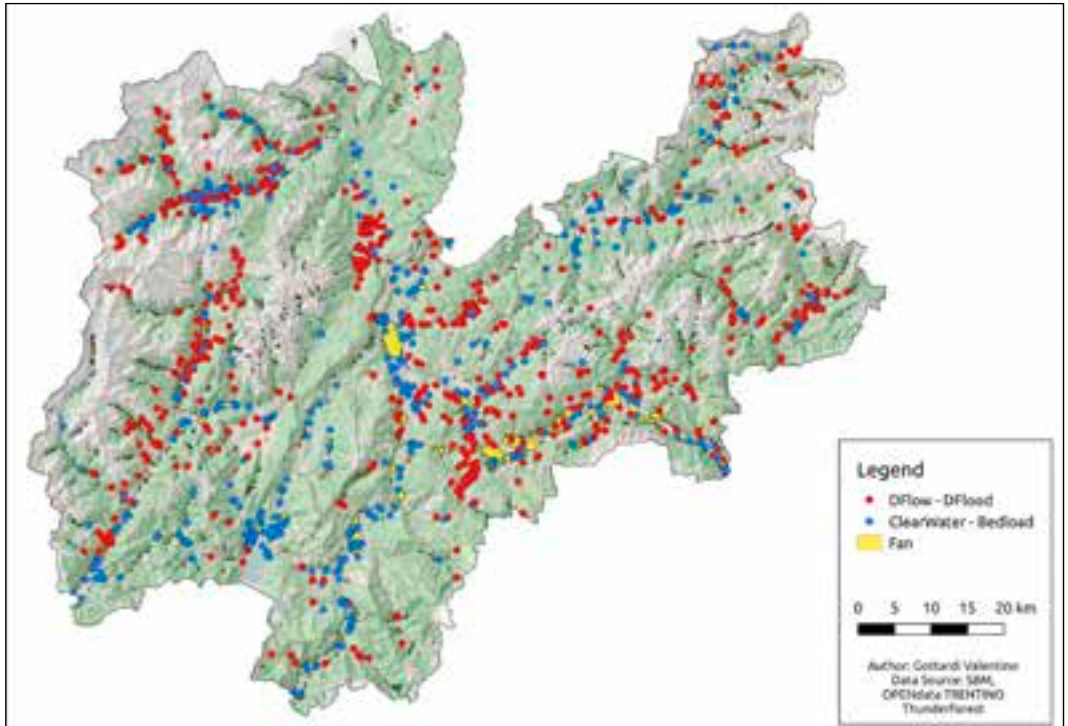


Figura 2 – Distribuzione spaziale dei processi di trasporto selettivo fluviale (blu) e di trasporto torrentizio di massa (rosso) - eventi sull'intero territorio provinciale.

Dall'operazione di filtraggio, gli eventi legati al comportamento fluviale selettivo o a processi di bassa intensità sono 715 di cui 516 eventi di solo deflusso liquido e 199 eventi di trasporto solido di fondo. I processi più intensi, legati ad un comportamento torrentizio sono 824 di cui 813 eventi di colata detritica e 11 eventi di corrente iperconcentrata. Questi 1539 processi di trasporto selettivo fluviale e di trasporto torrentizio di massa, con identificazione e localizzazione certa, sono distribuiti lungo le principali valli del Trentino, come mostrato in Figura 2. I punti blu rappresentano i processi fluviali (piena liquida o trasporto al fondo), mentre i punti rossi mostrano i processi torrentizi (colata detritica o corrente iperconcentrata).

Poiché vogliamo prevedere la probabilità che le colate detritiche raggiungano l'area del conoide, è necessario selezionare ulteriormente gli eventi, in base alla loro localizzazione. Gli eventi segnalati nel bacino, ma descritti come dannosi per le infrastrutture sul conoide, ven-

gono ricollocati manualmente. Considerando solo le osservazioni all'interno del perimetro dei conoidi. Per l'intero database dei conoidi (2586 in totale) si arriva a 406 eventi. Sono stati rilevati 204 eventi di processi fluviali (132 eventi di piena liquida e 72 eventi di trasporto al fondo). Il numero di processi torrentizi è invece pari a 202 (198 colate detritiche e 4 correnti iperconcentrate).

Per una prima valutazione i sistemi duali bacino-conoide sono stati classificati in base al tipo di eventi verificatisi durante il periodo di osservazione. Per ben 1538 bacini, che rappresentano il 90,6% del totale, non sono stati segnalati eventi nell'area di conoide. Il 6,1% dei bacini ha sperimentato solo processi ad alta intensità mentre il 2,1% dei bacini ha avuto solo processi fluviali. Il restante 1,2% ha mostrato un comportamento misto con segnalazioni di processi sia di alta che di bassa intensità. I conoidi con processi storici sono in totale 159, pari al 9,4% del totale. Il metodo proposto cer-

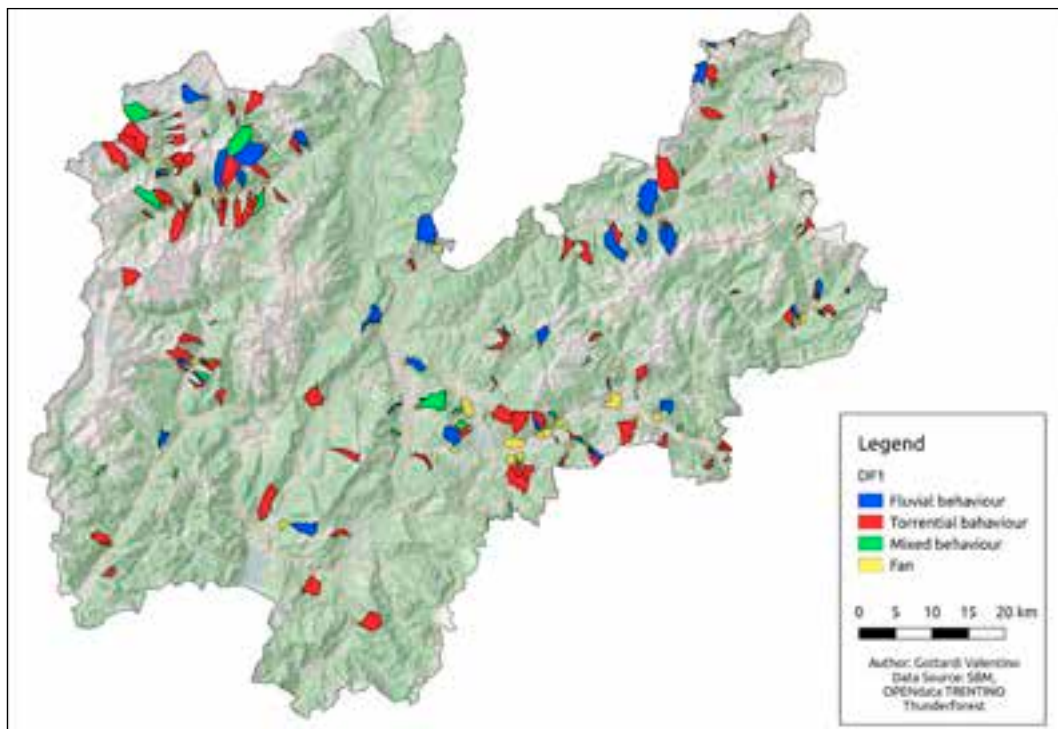


Figura 3 – Distribuzione spaziale dei sistemi bacino-conoide attivi. In rosso i bacini con comportamento torrentizio; in blu i bacini con comportamento fluviale; in verde i bacini con comportamento misto; in giallo i conoidi.

ca quindi di estrapolare da questi bacini attivi informazioni sul più probabile comportamento dei bacini per i quali non si hanno informazioni dirette sull'attività torrentizia. I 159 sistemi bacino-conoide attivi utilizzati per la creazione del modello, suddivisi in base al tipo di processo dominante, vengono rappresentati in Figura 3.

### La creazione del modello

Per addestrare e testare il modello vengono utilizzati solo i bacini attivi in quanto solo i bacini con un comportamento ben definito possono fornire informazioni utili. In questo caso, dei 159 bacini attivi, vengono esclusi i 20 che mostrano un comportamento misto in modo da utilizzare solo informazioni derivanti dai 103 bacini con un chiaro comportamento torrentizio e dai 36 bacini con un chiaro comportamento fluviale.

Una volta selezionati i bacini può iniziare la costruzione del modello. In questo caso il modello predittivo che verrà applicato ai dati vede due approcci diversi e consecutivi:

- Principal component analysis come input per superare il problema dell'elevata collinearità tra i predittori;
- regressione logistica per separare i bacini dominati da processi torrentizi da quelli dominati da processi fluviali.

Come appena accennato, la Principal component analysis consente di trattare variabili altamente correlate, caratteristica presente nei database a disposizione. Considerando inoltre la forte perdita di informazioni derivante dall'eventuale selezione manuale delle variabili caratterizzate da elevata collinearità, l'applicazione di uno strumento capace di operare anche in queste condizioni risulta particolarmente indicata. Questo metodo fornisce una visione semplificata dell'intero set di dati utilizzando



un insieme meno numeroso di caratteristiche, le componenti principali (in seguito PC), che sono linearmente non correlate tra loro. Tutta l'analisi viene eseguita garantendo la minor perdita possibile di varianza dei dati. La Principal component analysis nota anche con l'abbreviazione PCA, fa parte dell'ampio campo delle procedure di Machine learning e Data mining ed è una tecnica di semplificazione dei grandi insiemi di dati generalmente utilizzata nella statistica multivariata. Lo scopo di questa tecnica è di ridurre il numero di variabili che descrivono un insieme di dati in un insieme di variabili latenti meno numerose. Essa esamina le correlazioni tra gli oggetti originali e condensa le informazioni contenute in questi oggetti in un gruppo più piccolo di componenti garantendo al contempo la minor perdita possibile di informazioni, mantenendo così la varianza dei dati (TUFFERY, 2011).

La procedura statistica della PCA converte un insieme di osservazioni di variabili eventualmente correlate in un insieme di valori di variabili linearmente non correlate, utilizzando una trasformazione ortogonale. I dati originali sono espressi in termini di queste nuove variabili, chiamate Principal component. Considerate nel loro insieme, le nuove variabili rappresentano la stessa quantità di informazioni delle variabili originali e, allo stesso modo, anche la stessa quantità di varianza dei dati (KUHN e JOHNSON, 2013). Si consideri che la prima PC non solo rappresenta la maggior varianza tra le nuove variabili, ma anche la maggior varianza che una singola PC può rappresentare (KUHN e JOHNSON, 2013). È anche possibile calcolare la frazione di varianza rappresentata da sottoinsiemi di PCs semplicemente sommando le loro varianze e dividendo i risultati per la varianza totale del set di dati. Questa peculiarità della PCA permette di massimizzare la varianza rappresentata riducendo allo stesso tempo il numero di variabili considerate mantenendo solo le prime PCs in modo da semplificare il dataset senza compromettere l'informazione generale.

Il modello di regressione logistica viene applicato ai bacini attivi caratterizzati da un comportamento ben definito. Idealmente, un modello dovrebbe essere valutato su campioni che non sono stati utilizzati per la costruzione o per il fitting del modello stesso, al fine di for-

nire un senso non amplificato dell'efficacia del modello (KUHN e JOHNSON, 2013). Per valutarne il funzionamento è quindi importante disporre di due diversi insiemi di campioni: il primo viene utilizzato per creare e addestrare il modello e il secondo per valutare la sua capacità previsionale. Sono chiamati rispettivamente sottoinsieme di addestramento (train subset) e sottoinsieme di test (test subset).

Dato il basso numero di bacini a disposizione si deve ricorrere ad una tecnica di ricampionamento. Le differenze tra le tecniche risiedono solitamente nei vari metodi utilizzati per scegliere i sottocampioni (KUHN e JOHNSON, 2013). In questo caso si utilizza il metodo del bootstrapping. Esso si basa su un campionamento casuale con sostituzione. Il termine sostituzione significa che se un elemento viene selezionato, è ancora disponibile per un'ulteriore selezione. Il sottoinsieme ottenuto, chiamato campione bootstrap, è caratterizzato dalla stessa dimensione dell'insieme di dati originale. Ciò comporta la rappresentazione multipla di alcuni punti di dati in questo sottoinsieme. Altri punti di dati non vengono selezionati affatto e sono denominati campioni out-of-bag.

Il bootstrapping viene eseguito più volte, generalmente 1000 o 10000 volte. Il risultato di questa iterazione è un campione bootstrap utilizzato per costruire ed allenare il modello. Il campione out-of-bag viene predetto e utilizzato per stimare le prestazioni del modello. Metriche come la sensibilità, la specificità e l'Indice J Youden vengono utilizzate per valutare le performance del modello proposto.

Una volta costruiti i campioni di training e test, viene applicata una regressione logistica utilizzando solamente le prime 6 PCs calcolate utilizzando la PCA in modo da semplificare i parametri di ingresso senza perdere una eccessiva variabilità del dato. Il risultato dell'analisi è una variabile binaria del tipo 0-1 o improbabile-probabile. Le due classi di bacini rappresentano la variabile dipendente binaria che sarà predetta sulla base di molti predittori o variabili indipendenti (le PCs).

Per una variabile dipendente binomiale generale, la probabilità corrispondente della classe prevista "1" può variare tra 0 (certamente il valore "0", indicato come un "insuccesso" o un "non caso") e 1 (certamente il

valore “1”, indicato come un “successo” o un “caso”). La regressione logistica in sé modella semplicemente la probabilità di un evento in base ai predittori e non esegue una classificazione statistica. Il modello può essere utilizzato come classificatore binario quando si accoppiano le probabilità previste e un valore di soglia. In base a questa soglia, la probabilità stimata ottenuta viene classificata in classi. Gli elementi con probabilità superiore al valore di cutoff vengono classificati nella classe “1”, mentre la classe “0” viene assegnata a probabilità inferiori. Per perfezionare la capacità predittiva del modello è stata scelta la soglia di predizione che massimizza il parametro denominato Indice J di Youden e che risulta pari a 0,5222838, di poco superiore alla soglia standard dello 0,5. L'indice J di Youden va da 0 a 1. Il valore è 0 quando il modello è inutile. Il valore di 1 indica che non ci sono falsi positivi o falsi negativi, quindi il test è perfetto. Questo indice assegna lo stesso peso ai valori falsi positivi e falsi negativi.

Una volta che il modello è stato costruito e testato sui bacini attivi, viene applicato all'intero database per effettuare la vera e propria predizione del comportamento dei bacini per i quali non esistono eventi registrati. Per motivi di spazio, per la descrizione matematica della procedura di PCA e della regressione logistica si rimanda all'elaborato originale.

## I risultati del modello

Le prestazioni del modello addestrato possono essere confrontate con altri metodi presenti in letteratura. Per avere una semplice panoramica dei metodi, vengono utilizzati due approcci diversi. Il primo utilizza solo parametri relativi alla morfologia del bacino idrografico. Il secondo si basa sulla morfometria del bacino e del conoide. La selezione dei metodi utilizzati per il confronto viene fatta cercando di utilizzare solo le caratteristiche già disponibili. In questo modo i risultati del modello proposto e dei modelli disponibili sono ben confrontabili.

Tra i sistemi di classificazione basati solo sui parametri morfometrici del bacino idrografico sono stati selezionati i lavori di D'AGOSTINO (1996) e di DE SCALLY e OWENS (2004). Questi

autori hanno proposto un approccio simile basato sul numero di Melton, che differisce solo in base alla soglia fissata. D'AGOSTINO (1996), analizzando 62 eventi storici avvenuti in bacini di piccole e medie dimensioni nella parte orientale del Trentino, ha definito una soglia pari a 0,50 per distinguere tra bacini con processi esclusivamente fluviali e bacini con processi ad alta intensità. Lo stesso approccio è utilizzato anche nel lavoro di DE SCALLY e OWENS (2004) per l'analisi sui bacini idrografici della catena alpina della Nuova Zelanda. In questo caso la soglia proposta è fissata a 0,75.

Gli stessi autori hanno proposto anche dei sistemi di classificazione basati su parametri morfometrici sia del bacino che del conoide. Oltre al numero di Melton viene considerata anche la pendenza media del conoide. D'AGOSTINO (1996) classifica un bacino come propenso a generare fenomeni intensi se la pendenza media del conoide, espressa in gradi, supera il valore del numero di Melton moltiplicato per un fattore più o meno conservativo variabile da 7 a 14. Diversamente da quanto appena descritto, DE SCALLY e OWENS (2004) hanno affermato che un bacino è incline a generare flussi detritici se, contemporaneamente, sia il numero di Melton sia la pendenza media del conoide superano determinate soglie. Colate detritiche e correnti iperconcentrate sono attese in bacini con un numero di Melton superiore a 0,75 e una pendenza media del conoide superiore a 7,5°. Se questi due requisiti non vengono rispettati, si prevedono solo processi di trasporto di fondo.

Per confrontare i risultati del modello proposto con quelli dei modelli presenti in letteratura vengono utilizzate alcune metriche statistiche tra le quali la statistica Kappa che considera l'accuratezza che sarebbe generata semplicemente dal caso (KUHN e JOHNSON, 2013), la sensibilità che è la capacità di individuare la classe positiva e la specificità che è la capacità di individuare la classe negativa. Quando il modello non utilizza pesi o penalità differenziate per errori di falsi positivi o falsi negativi, è possibile avere un unico parametro di misura. L'indice J di Youden sintetizza l'entità di entrambi i tipi di errore misurando le proporzioni di campioni correttamente previsti per entrambi i gruppi di eventi e di non eventi. Tra i metodi testati, il più alto valore di Youden's J

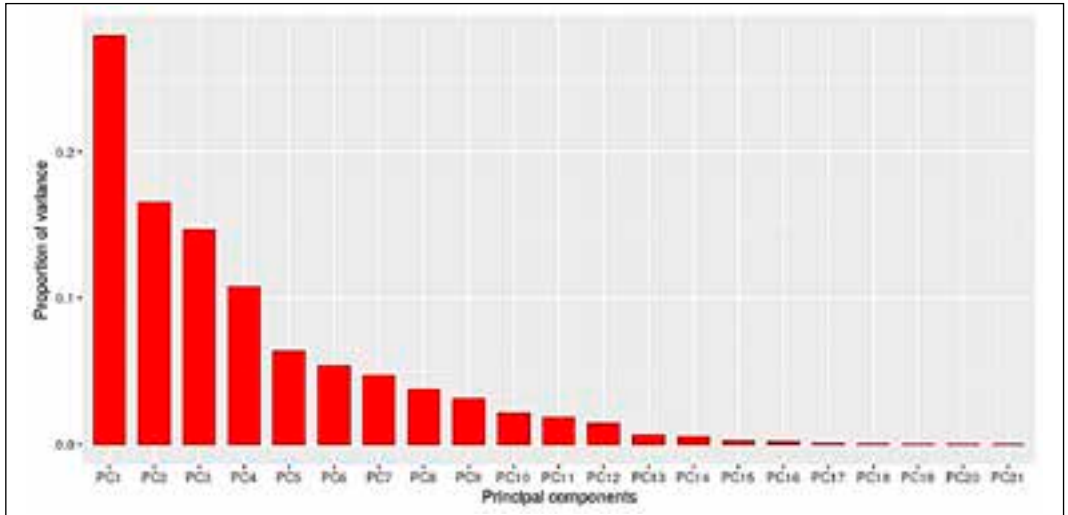


Figura 4 – Scree plot della Principal component analysis condotta sui bacini attivi.

è raggiunto dal modello proposto (0,4182), seguito dai modelli di DE SCALLY e OWENS (2004) e dall'approccio di D'AGOSTINO (1996).

L'applicazione della Principal component analysis al database dei bacini attivi restituisce le PCs derivanti dalle variabili originali. Come già spiegato, la prima PC rappresenta la maggior parte della varianza totale e le successive hanno un'importanza decrescente. La Figura 4 rappresenta graficamente l'importanza delle componenti calcolate. Il grafico, definito scree plot, mostra quanta varianza ogni PC rappresenta rispetto alla varianza totale dei dati. Scegliendo le prime 6 PCs, si riesce a rappresentare l'81,5% della varianza totale, raggiungendo così un buon compromesso tra semplificazione del modello e rappresentatività dell'informazione disponibile.

Le prestazioni della regressione logistica, condotta utilizzando le prime 6 PCs, possono essere espresse graficamente tracciando la probabilità prevista e le osservazioni. I bacini vengono ordinati secondo una probabilità crescente in modo da formare una graduatoria. Ogni punto è tracciato rispettando il ranking di probabilità e colorato in riferimento alle osservazioni effettive. In Figura 5 viene illustrata la probabilità che ciascuno dei 139 bacini utilizzati per la creazione del modello generi processi torrentizi intensi fin sul conoide. Il colore del punto indica il comportamento effettivamente

osservato. Per la maggior parte dei bacini con comportamento torrentizio (punti colorati in turchese) è stata prevista un'alta probabilità di generare colate detritiche o correnti iperconcentrate (probabilità superiore alla soglia di 0,50, indicata dalla linea rossa). Una buona capacità di separazione viene evidenziata anche per i bacini idrografici a dominante comportamento fluviale (colorati in arancio).

Dopo la costruzione e l'addestramento del modello, questo viene applicato all'intero database dei sistemi bacino-conoide. La distribuzione della probabilità prevista per tutti i 1697 bacini è rappresentata graficamente nella Figura 6 come spiegato per il grafico della precedente Figura 5 e utilizzando la soglia di cutoff ottimizzata pari a 0,5222838 per separare i bacini con predominante comportamento fluviale da quelli con comportamento torrentizio.

I processi di trasporto al fondo e di portata liquida sono considerati dominanti per 210 bacini. Questi bacini con bassa probabilità di debris flow rappresentano il 12,38% dell'intero database. Un comportamento torrentizio, con colate detritiche e correnti iperconcentrate, è previsto per i restanti 1482 bacini del territorio provinciale pari all'87,33% dei bacini alpini considerati. È interessante notare che solo 159 bacini, il 9,37% del totale, hanno riportato eventi sul conoide durante la finestra temporale considerata.

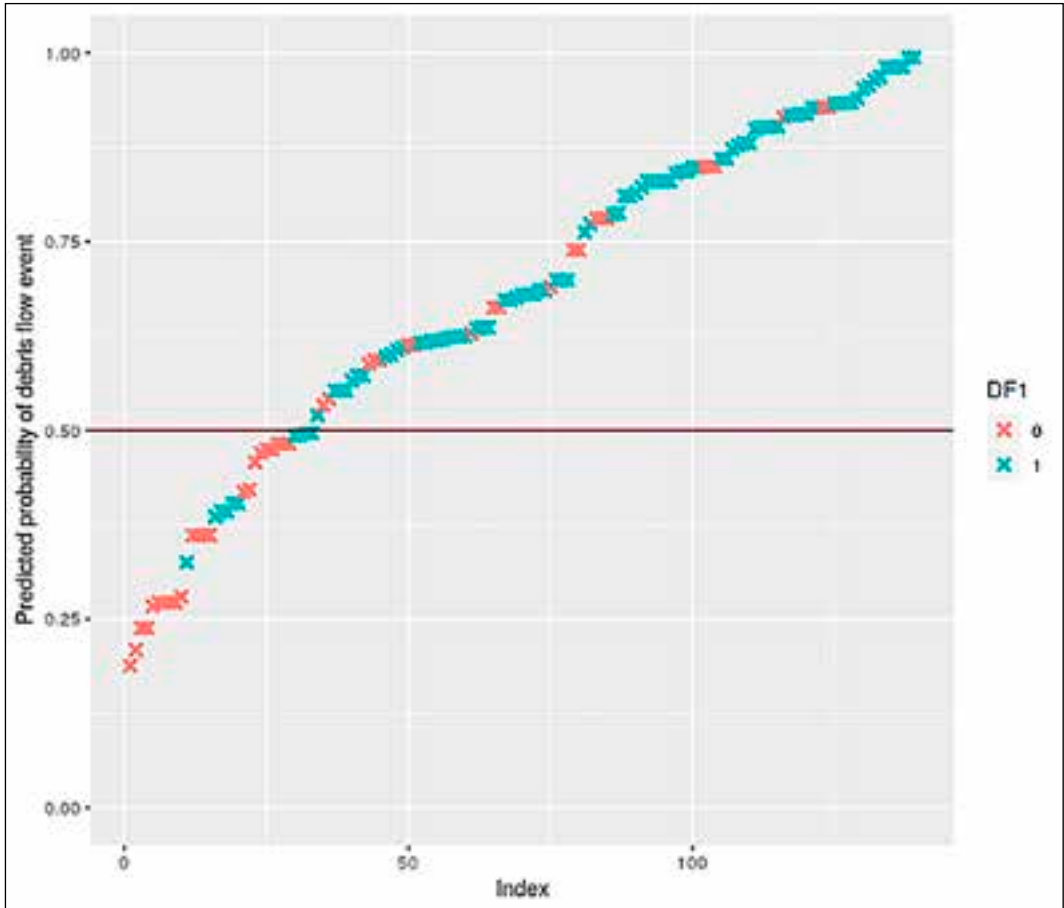


Figura 5 – Regressione logistica: ordinamento delle probabilità calcolate. In turchese bacini con comportamento torrentizio; in arancione bacini con comportamento fluviale.

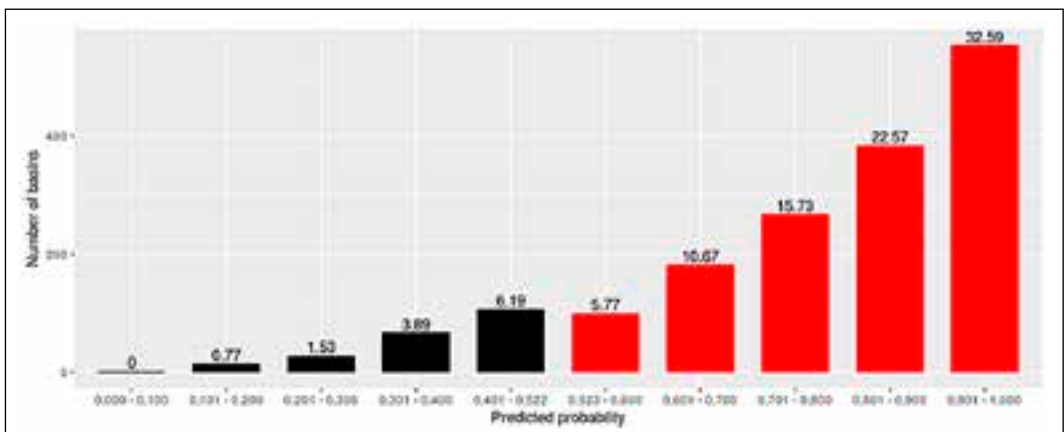


Figura 6 – Probabilità prevista e distribuzione nelle dei bacini in classi di probabilità.

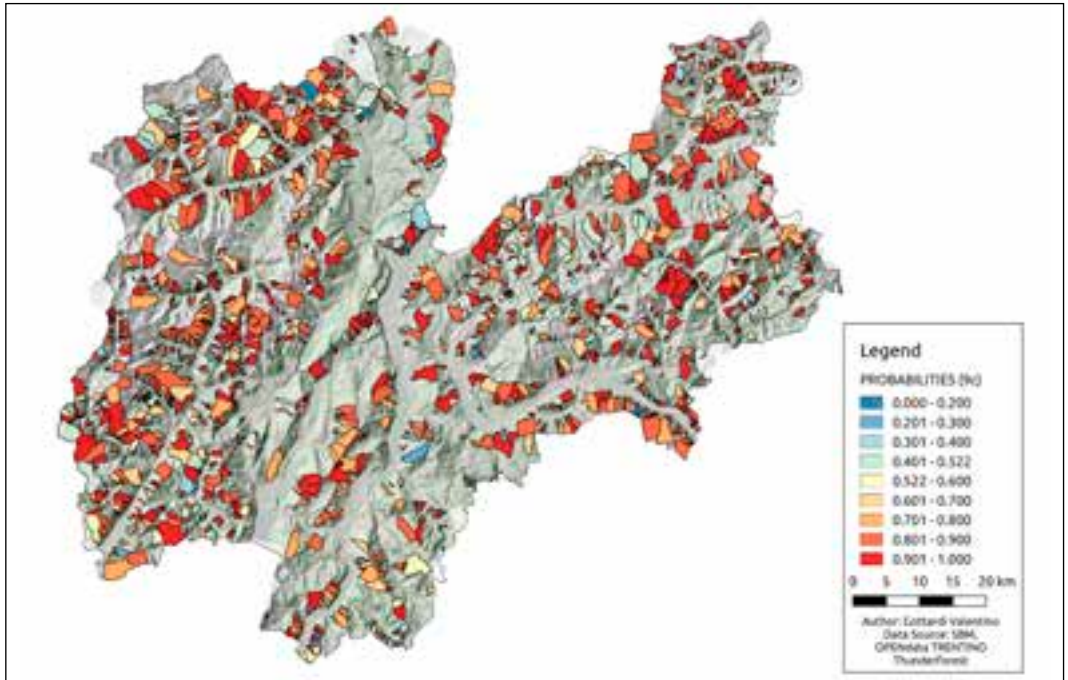


Figura 7 – Risultati modellazione: probabilità prevista suddivisa in 9 classi.

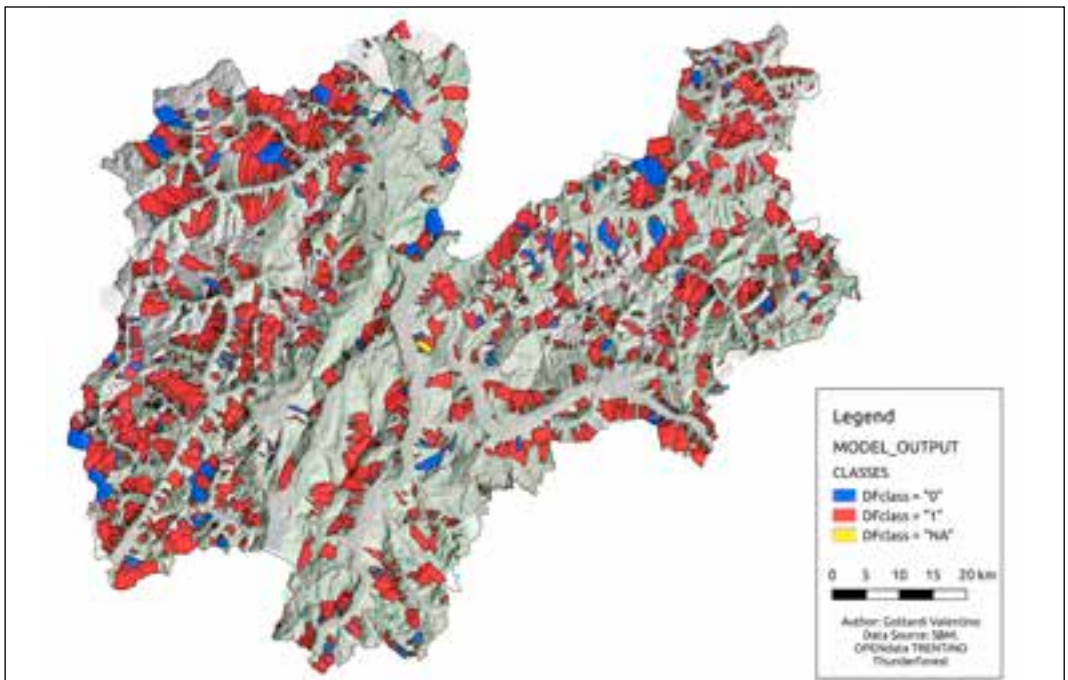


Figura 8 – Risultati modellazione: classi previste.

I risultati della modellazione condotti sull'intero database sono graficamente rappresentati nelle seguenti figure. La Figura 7 rappresenta i bacini ordinati in 9 classi sulla base della probabilità calcolata di generare e trasportare processi torrentizi fino al conoide. La Figura 8 rappresenta invece la classificazione del comportamento atteso dei bacini mettendo a confronto la probabilità calcolata e la soglia ottimizzata di classificazione.

## Discussione dei risultati e conclusioni

Sebbene con qualche elemento di incertezza, l'utilizzo dei database disponibili ha consentito di creare un modello di classificazione del comportamento atteso di piccoli bacini idrografici del territorio della provincia di Trento. Il risultato può essere considerato soddisfacente anche in considerazione degli scarsi risultati ottenuti da altri approcci presenti in letteratura che utilizzano solo pochi parametri previsionali.

Le caratteristiche morfologiche dei 1697 sistemi bacino-conoide sono state create con un approccio automatizzato e devono essere considerate generalmente robuste. Più importante risulta il filtraggio e la selezione degli eventi segnalati. Questo è dovuto principalmente alle diverse fonti informative utilizzate per la costruzione del catasto degli eventi. I 3522 eventi totali sono caratterizzati da forti differenze per quanto riguarda epoca di accadimento e attendibilità. Come è ovvio pensare, i vecchi registri parrocchiali o gli articoli di giornale riportano descrizioni imprecise del processo torrentizio occorso, concentrandosi spesso solo sui danni e sulle vittime. A seguito della selezione, solo 269 eventi verificatisi nelle aree di conoide sono stati reputati affidabili e quindi utilizzati per la creazione del modello.

La selezione dei soli bacini attivi e caratterizzati da un comportamento ben preciso, ovvero quelli che hanno evidenziato solo fenomeni intensi o solo fenomeni di lieve entità, ha portato ad una forte riduzione del numero di sistemi bacino-conoide disponibili per le operazioni di training e fitting del modello, che si attesta a 139 bacini.

L'esigua numerosità dei record a dispo-

zione e l'elevata collinearità di molte delle variabili ha imposto l'utilizzo di tecniche di ricampionamento e di analisi avanzata per l'estrazione di informazioni da database complessi. L'applicazione della regressione logistica dà luogo a un modello di output intelligente e di facile lettura e interpretazione.

Sorgono alcune considerazioni riguardanti il comportamento osservato dei bacini.

- È altamente possibile che, durante il periodo post-glaciale, grandi colate detritiche si siano verificate anche in bacini oggi considerati dominati solo da processi fluviali. Questo antico cambiamento comportamentale non ha influito sull'utilità di una classificazione in quanto siamo interessati al comportamento attuale dei torrenti. Tuttavia, le caratteristiche geomorfologiche create nel passato possono essere responsabili di errori di previsione quando vengono utilizzate come predittori dei modelli.
- Nei bacini con processi molto intensi, le portate liquide e il trasporto di fondo possono essere trascurati e quindi non venir registrati. Questo vale anche per tutti i fenomeni torrentizi verificatisi in bacini situati in aree remote e non antropizzate.
- La selezione dei soli bacini che presentano eventi in area di conoide durante la finestra temporale analizzata porta a perdere bacini potenzialmente molto informativi ma che hanno generato fenomeni solo in area di innesco e di trasporto.
- I futuri cambiamenti climatici, e quindi le possibili alterazioni del comportamento dei torrenti, non possono essere presi in considerazione (distribuzione annuale delle precipitazioni, scioglimento del permafrost nelle regioni alpine, ecc.).
- La geometria dei conoidi è spesso soggetta a forzanti esterne al bacino contribuente, come l'erosione dovuta al fiume di fondovalle o l'alimentazione incrociata con bacini limitrofi, portando a incerta classificazione degli eventi osservati.

In conclusione, il modello proposto, pur presentando numerosi aspetti migliorabili, indica efficacemente punti di forza e di debolezza dei database provinciali finora disponibili. Il miglioramento dell'omogeneità della registrazione

degli eventi e la pulizia dei segnali meno informativi porteranno probabilmente a una migliore valutazione del catasto degli eventi e a una previsione più solida e affidabile della procedura automatizzata.

## Ringraziamenti

Il presente articolo è tratto dal lavoro di tesi dell'autore, dal titolo originale "Classification of torrential behaviour of torrents in the Province of Trento", per il conseguimento della laurea magistrale nel corso di studi sui rischi naturali alpini "Alpine Naturgefahren/Wildbach- und Lawinenverbauung" presso il Dipartimento di ingegneria civile e rischi naturali dell'Università BOKU di Vienna. Si ringrazia il Prof. Johannes Hübl della BOKU, il dott. Ruggero Valentinotti e il dott. Gabriele Bertoldi del Servizio Bacini montani della Provincia autonoma di Trento.

## BIBLIOGRAFIA

BERGMEISTER, K., 2009 – *Schutzbauwerke gegen Wildbachgefahren: Grundlagen, Entwurf und Bemessung*, Beispiele. John Wiley & Sons.

Bundesministerium für Land- und Forstwirtschaft, 1984 – 100 Jahre Wildbachverbauung in Österreich. In self publishing, Wien

CASTIGLIONI, G.B., 1971 – *Le calamità naturali nelle Alpi*. Pages 7-37 in Atti del XXI Congresso Geografico Italiano, Verbania, Italy (in Italian).

D'AGOSTINO, V., 1996 – *Analisi quantitativa e qualitativa del trasporto solido torrentizio nei bacini montani del Trentino Orientale*. Associazione Italiana di Ingegneria Agraria, 1.

DE SCALLY, F. A., & OWENS, I. F., 2004 – *Morphometric controls and geomorphic responses on fans in the Southern Alps, New Zealand*. Earth Surface Processes and Landforms, 29(3), 311-322.

IVERSON, R. M., 1997 – *The physics of debris flows*. Reviews of geophysics, 35(3), 245-296.

KUHN, M., & JOHNSON, K., 2013 – *Applied predictive modeling* (Vol. 26). New York: Springer.

MARCHI, L., & CAVALLI, M., 2007 – *Procedures for the documentation of historical debris flows: application to the Chieppena Torrent (Italian Alps)*. Environmental management, 40(3), 493-503.

TUFFÉRY, S., 2011 – *Data Mining and Statistics for Decision Making*. John Wiley & Sons, Ltd.

VANDINE, D. F., 1996 – *Debris Flow Control Structures for Forest Engineering*. Working paper 22/1996, Ministry of Forest Research Program, British Columbia

ZIMMERMANN, M., MANI, P., GAMMA, P., GSTEIGER, P. HEINIGER, O. & HUNZIKER, G., 1997 – *Murengefahr und Klimänderung – ein GIS-basierter Ansatz*. Zuerich.

## Valentino Gottardi

E-mail: valentino.gottardi@provincia.tn.it  
Forestate Provincia Autonoma di Trento  
Servizio Foreste Corpo Forestale Trentino  
Via G. B. Trener, 3 – 38121 Trento

**PAROLE CHIAVE:** *colata detritica, bacino idrografico, Principal Component Analysis, regressione logistica, comportamento torrentizio*

**KEYWORDS:** *debris flow, watershed, Principal Component Analysis, Logistic regression, torrential behaviour*

## RIASSUNTO

La disponibilità di banche dati relative a caratteri geomorfologici e climatici dei bacini alpini e di un catasto eventi per la provincia di Trento ha permesso la creazione di un modello di classificazione della probabilità di avvenimento di fenomeni di colata detritica in area di conoide. Il modello proposto utilizza la Principal Component Analysis e la regressione logistica. I risultati vengono confrontati con alcuni modelli previsionali presenti in letteratura.

## SUMMARY

The availability of databases on geomorphological and climatic characteristics of the Alpine basins and an event cadaster for the province of Trento allowed the creation of a model for classifying the probability of occurrence of debris flow phenomena in fan areas. The proposed model uses Principal Component Analysis and logistic regression. The results are compared with some prediction models in the literature.