

PIETRO PATRIZIO CIRO AUCELLI (\*), ALDO CINQUE (\*\*) & ANTONIO ESPOSITO (\*\*)

## L'USO DI TECNICHE DI CLASSIFICAZIONE NON SUPERVISIONATA PER LA MAPPATURA AUTOMATIZZATA DI FORME A PARTIRE DA MODELLI DIGITALI DEL TERRENO

**ABSTRACT:** AUCELLI P.P.C., CINQUE A. & ESPOSITO A., *Unsupervised classification techniques for automated mapping of landform elements from digital elevation models*. (IT ISSN 1724-4757, 2005).

We report the preliminary results of a methodology for the automated mapping of some landform features. Our investigation has still to be refined and extended to a wider area to give reliable results for the study of the dissectional and tectonic history of the study region, but it is already possible to discuss some advances we have achieved in automatic landform classification for the study area.

The Trigno River catchment was selected to test the method. In the study area, due to the almost complete lack of river terraces, the reconstruction of old stages of valleys development must be attempted using other characters of the valley-side slopes. Among others, the widespread, gentle erosional glacia that overhang the present thalwegs at different elevations can be used as indicators of discontinuous downcutting and as proxy of ancient base levels. Most the methods that are currently used to characterize terrain features by means of DEMs extract the information contained in each cell of the raster and are able to quantify the local surface geometry very accurately (within the resolution limit of the DEM used). However, to make discriminations among landforms which differ in terms of spatial and contextual proprieties, it is essential to take into account also the properties of neighbour cells. In this sense, we propose a procedure that, instead of using a classic neighbourhood analysis, partitions the DEM by submitting slope angle and elevation value to a k-means algorithm for the identification of homogeneous landform elements. In this way, we expect to obtain patches of contiguous cells whose boundaries reflect characteristic landform elements of the study area. Mean values of (i) slope angle, (ii) tangential curvature and (iii) profile curvature of the cells belonging to each patch, are then used as «feature vectors» for an unsupervised landform classification of the area under examination. For the Trigno River test area a 10-cluster solution has been chosen by a trial and error procedure. The resulting automated mapping is judged very good as it recognizes ten different units, each of them matches very well with ones of the geomorphological units that have been surveyed in the test area. A comparison with a cell-by-cell unsupervised classification shows that the proposed method pro-

duces more accurate results. In particular a good discrimination between summit erosional surfaces and valley bottom were achieved. Further developments, aimed to the refinement of the classification results, are also briefly discussed.

**KEY WORDS:** Unsupervised classification, Glacis, DEM, GIS, Molise Region (Italy).

**RIASSUNTO:** AUCELLI P.P.C., CINQUE A. & ESPOSITO A., *L'uso di tecniche di classificazione non supervisionata per la mappatura automatizzata di forme a partire da modelli digitali del terreno*. (IT ISSN 1724-4757, 2005).

Il seguente lavoro espone l'uso di una procedura per la produzione di cartografia automatica di forme di interesse che si avvale dell'uso della classificazione non supervisionata implementata utilizzando l'algoritmo *k-means* ed i modelli digitali del terreno. Analizzare l'informazione contenuta nelle singole celle del raster risulta insufficiente per individuare automaticamente le forme del rilievo tradizionalmente cartografate. Il metodo proposto, benché ancora da affinare sotto vari aspetti, tiene conto di tale problematica attraverso una procedura volta ad individuare i bordi degli elementi comunemente cartografabili nell'area di studio fornendo risultati di maggiore utilità rispetto a metodi che tengono conto di un solo parametro morfometrico o di valori estratti da singole celle del raster.

**TERMINI CHIAVE:** Classificazione non supervisionata, Glacis, DEM, GIS, Regione Molise.

### INTRODUZIONE

L'interpretazione di carte topografiche per la realizzazione di una carta geomorfologica prevede necessariamente una fase in cui si provvede a riconoscere e delimitare «forme particolari» della superficie terrestre che sono il prodotto di determinati processi considerati rilevanti per la descrizione degli eventi che hanno determinato l'evoluzione di una particolare area di studio. Forme semplici, quali ad esempio certi edifici vulcanici monogenici, sono facilmente individuabili grazie alla loro geometria conica e quindi addirittura rapportabili alle regole della geometria euclidea. Ma la maggioranza delle forme naturali sono geo-

(\*) Università degli Studi del Molise, Dipartimento S.T.A.T., via Mazzini 8 - 67100 Isernia.

(\*\*) Università degli Studi di Napoli Federico II, Dipartimento di Scienze della Terra, largo S. Marcellino 10 - 80100 Napoli.

Lavoro stampato con i contributi del fondo di ricerca dipartimentale 40/138-106/6 anno 2003, Resp. A. Cinque.

metricamente complesse specie se derivano dall' azione di più processi. Le loro proprietà geometriche non sono quindi esprimibili secondo le regole della geometria euclidea. Inoltre la maggior parte degli elementi del paesaggio non sono caratterizzati da confini, presentando spesso delle zone marginali di transizione con altri elementi. In tale frequente caso mappare dei confini tra le forme è un processo che implica un elevato grado di soggettività e richiede un notevole livello di esperienza. Ciò non solo perché la capacità di riconoscere forme di genesi diversa dipende da quanti modelli interpretativi possiede il geomorfologo, ma anche perché la stessa fase di analisi della superficie topografica (sia essa fatta sul campo, che su carte o su foto aeree, etc.) viene svolta con intensità ed accuratezza variabili da persona a persona. Per non dire che anche al singolo geomorfologo, quando lavora a lungo su aree vaste, può capitare di usare criteri di analisi e livelli di attenzione un po' diversi da luogo a luogo. L'esigenza di un tipo di analisi meno soggettiva e, possibilmente, più rapida (per operare su vaste aree), ha reso necessaria l'introduzione, in geomorfologia, di parametri che permettessero di caratterizzare le proprietà geometriche del rilievo e di cogliere, entro una data area, quali variazioni di tali parametri fossero significative di passaggi da un elemento ad un altro. Ad esempio, le derivate di primo e secondo ordine (che sono relative alla pendenza e la curvatura della superficie) possono essere proficuamente utilizzate per classificare la superficie topografica in vari elementi di interesse geomorfologico e quantificare le loro proprietà geometriche. L'approccio quantitativo è divenuto di ampia utilizzazione grazie alla diffusione di computer a basso costo e alla crescente disponibilità di modelli digitali del terreno (DEM). Ciò ha reso possibile effettuare automaticamente vari tipi di analisi della superficie topografica. I DEM di uso più comune sono costituiti da una griglia di valori altimetrici la cui spaziatura costituisce la risoluzione del DEM. In questo caso la superficie terrestre non viene rappresentata come una funzione continua  $z = f(x, y)$  ma da valori discreti campionati secondo lo schema della griglia. Esistono attualmente vari software, commerciali e non, in grado di gestire modelli digitali del terreno. Attraverso tali pacchetti è possibile realizzare DEM a partire da curve di livello tramite algoritmi di interpolazione ed effettuare vari altri tipi di elaborazioni (Mitasova & Hofierka, 1993). Sicuramente una delle applicazioni dei DEM che ha avuto maggior successo nell'ambito della cartografazione automatica è stata l'estrazione dei reticoli e dei bacini idrografici (Nogami, 1995). Diverse varianti di algoritmi atti allo scopo sono infatti oggi implementati in pacchetti GIS di ampia diffusione, mentre la mappatura di forme del rilievo più complesse risulta essere molto meno formalizzata e quasi del tutto assente nel panorama software attuale. Le tecniche di classificazione automatica sono state sviluppate nell'ambito della «Pattern Recognition» e sono principalmente suddivisibili in due tipologie: supervisionata o non supervisionata (Webb, 2002).

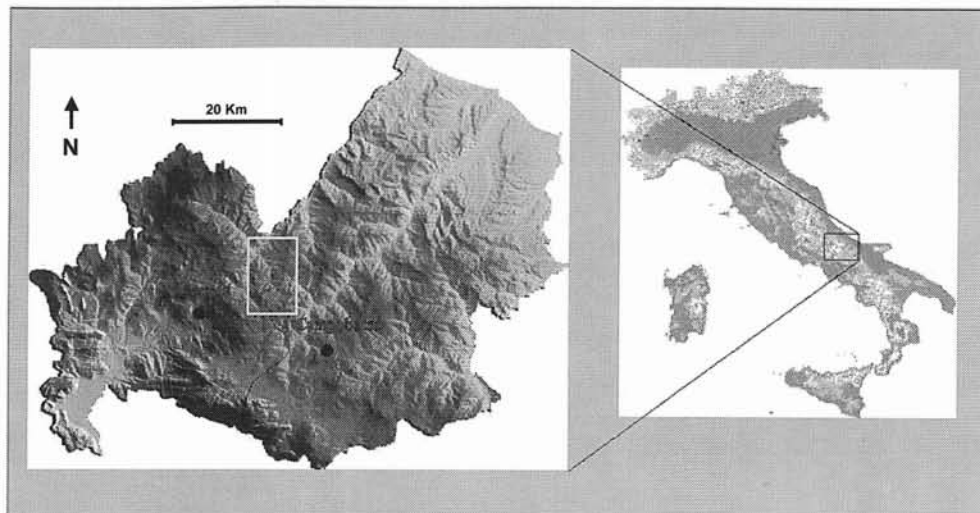
La classificazione supervisionata implica innanzitutto la selezione di campioni rappresentativi (tipici) del morfotipo che si intende far riconoscere automaticamente dal software. Da questi campioni (insieme di celle), vengono estratti

parametri che l'algoritmo è in grado di utilizzare per apprendere le caratteristiche del morfotipo e quindi individuare in altre aree. Le reti neurali artificiali (ANN) sono un tipico esempio di tale classe di algoritmi. La classificazione non supervisionata invece utilizza algoritmi in grado di individuare gruppi (o *cluster*) (Jain & alii, 1999; Webb, 2002) all'interno di un set di dati multivariato. Tale tecnica non classifica i gruppi, che devono essere invece classificati a posteriori dall'operatore, ma aiuta ad estrarre, dall'interno dei dati, delle strutture che sarebbero difficilmente individuabili manualmente.

Data anche l'analogia tra la struttura *raster* dei dati multispettrali dei satelliti e quella dei modelli digitali del terreno, molte tecniche utilizzate nell'ambito della cartografazione automatica, derivano direttamente da quelle utilizzate nel campo del *remote sensing*, dove tali tecniche di analisi multivariata sono da tempo di uso comune. Nei lavori di Irvin & alii (1997) e Burrough & alii (2000), si espone l'uso di algoritmi per il *clustering* iterativo al fine di partizionare un DEM in poche classi associabili ad alcune specifiche forme presenti nel territorio in esame utilizzando vari parametri morfometrici. Le reti neurali artificiali sono state usate da Brown & alii (1998), per il riconoscimento automatico di particolari forme del rilievo analizzando i parametri morfometrici estratti dalle celle del *raster*. Le proprietà associate alle singole celle del *raster* non forniscono le informazioni necessarie a caratterizzare l'andamento della topografia di forme complesse la cui estensione spaziale ovviamente richiede tecniche di analisi che tengano conto delle proprietà spaziali di tutte le celle che ricoprono la superficie della forma di interesse. Comunemente un'analisi di tale tipo si avvale di finestre di dimensioni fisse in termini di celle, la cosiddetta *neighborhood analysis*. Un approccio simile è stato utilizzato da Dikau & alii (1995) per realizzare una mappa delle forme del rilievo a piccola scala del Nuovo Messico muovendo una finestra all'interno del *raster* e calcolando alcuni valori statistici all'interno della stessa per classificare la cella centrale come appartenente ad una particolare forma del rilievo. La forma regolare della finestra imposta nelle tecniche di *neighborhood analysis* determina spesso risultati che non rispecchiano le reali caratteristiche delle forme del rilievo terrestre a causa della forma irregolare dei bordi che esse presentano. Ciò rende necessario sperimentare approcci alternativi. Il presente lavoro propone un metodo per la classificazione delle forme del rilievo che si avvale del *clustering* iterativo applicato ai modelli digitali del terreno. Il metodo, attraverso opportune trasformazioni del set di dati iniziale, definisce nuove proprietà morfometriche che meglio si prestano a simulare quelle fasi di generalizzazione che automaticamente la mente umana esegue quando il geomorfologo passa dalla «lettura» di una carta topografica al riconoscimento ed alla cartografia delle forme di interesse.

Sulla base di queste considerazioni, da qualche tempo è stata avviata una ricerca volta ad individuare, tramite l'uso di Sistemi Informativi Geografici, metodi automatici utili alla cartografazione di particolari superfici ed al loro ordinamento finalizzata alla ricostruzione tridimensionale di antiche morfologie vallive (Aucelli & alii, 2002).

FIG. 1 - Inquadramento geografico dell'area di studio.



## L'AREA DI STUDIO

Pur avendo finalità metodologiche, la ricerca in corso, parte da casi reali che divengono anche le aree test sulle quali verificare l'efficacia dei metodi di analisi che via via vengono sviluppati. Tra le tante aree test adottate, in questo articolo viene considerata quella del Torrente Vella (fig. 1) un affluente in destra idrografica del Fiume Trigno che è uno dei maggiori corsi d'acqua che drenano il versante Adriatico dell'Appennino Molisano. Il bacino ha una superficie di 51 km<sup>2</sup> con quota e pendenza media rispettivamente di 622 m s.l.m. e 10 gradi. La massima differenza altimetrica tra la foce, situata a 315 m s.l.m., e lo spartiacque, raggiunge localmente i 600 m. Le principali litologie affioranti consistono in depositi di mare profondo quali argille e argille marnose. A causa della bassa permeabilità ed alta erodibilità delle suddette litologie, l'idrografia superficiale è caratterizzata da un'alta densità di drenaggio e da un pattern sub-dendritico e valli a V. Il fondovalle risulta essere molto angusto e l'alveo del T. Vella risulta confinato tra stretti fianchi vallivi. Ne consegue che i terrazzi fluviali sono scarsi o difficilmente riconoscibili. Risultano invece abbastanza estesi vari lembi di paleosuperfici sommitali (Aucelli & alii, 2001) e numerosi *glacis* di erosione sospesi lungo i fianchi vallivi (Glacis Vallivi Sospesi - GVS di Aucelli, 1999). Questi ultimi sono testimoni di più fasi erosionali e sono interpretati come resti di antichi fianchi vallivi la cui genesi è fortemente collegata alle cicliche fasi di svassamento del fondovalle e susseguente reincisione per ripresa dei fenomeni di approfondimento del reticolo idrografico. Si tratta quindi di relitti la cui attenta analisi può essere preziosa per la ricostruzione degli antichi livelli di base locali e di perturbazioni tettoniche e/o climatiche.

## MATERIALI E METODI

Il modello digitale del terreno utilizzato (fig. 2) ha una risoluzione di 40 m per cella ed è stato generato tramite

interpolazione delle linee di livello digitalizzate da una tavoletta topografica IGM in scala 1:25.000. Le elaborazioni sono state effettuate in ambiente Linux/UNIX utilizzando il software GIS GRASS versione 5 e l'ambiente statistico «R» integrato allo stesso GIS tramite apposita libreria «R-GRASS» (Bivand, 1999). I valori statistici di seguito descritti sono stati estratti dal DEM direttamente in GRASS tramite un modulo scritto appositamente in linguaggio C (r.fpstat), mentre per l'algoritmo k-means si è utilizzata la versione fornita con il pacchetto *mva* che fa parte della dotazione standard della versione base di «R».

Dal DEM sono stati derivati i parametri di pendenza, curvatura del profilo e curvatura tangenziale. La curvatura

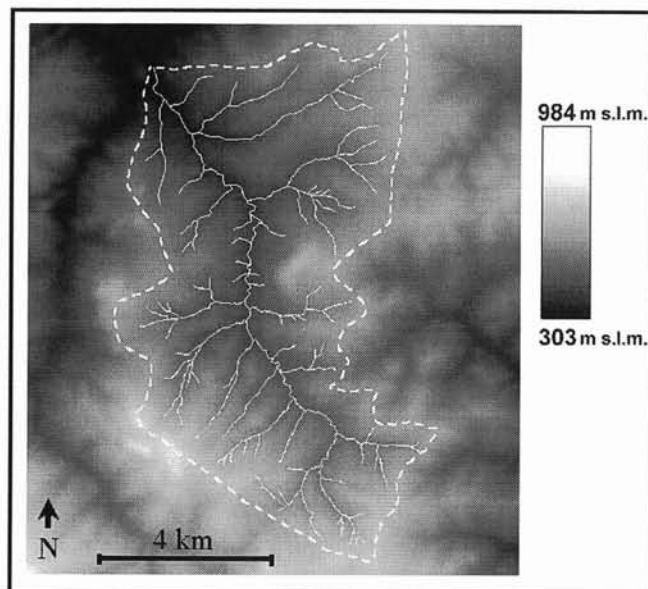


FIG. 2 - Modello digitale del terreno del bacino del T. Vella.

del profilo rappresenta la variazione dell'angolo di pendenza lungo la direzione di massima pendenza, mentre la curvatura tangenziale rappresenta la curvatura della superficie nella direzione ortogonale alla massima pendenza (*profile curvature e tangential curvature sensu Mitasova & Hofierka, 1993*).

Successivamente sono stati individuati dei domini spaziali attraverso il partizionamento del raster dell'area di studio. Questo è stato ottenuto sottoponendo i *layer* di quota e pendenza precedentemente derivati, ad un algoritmo che effettua un *clustering* iterativo, nel corso del quale i dati vengono in tal modo suddivisi in gruppi o *cluster* corrispondenti a gruppi di celle del DEM. Per effettuare il *clustering* si è utilizzato l'algoritmo *k-means* il cui obiettivo è quello di minimizzare la funzione:

$$e^2(S,L) = \sum_{J=1}^K \sum_{i=1}^{n_i} \|x_i^{(j)} - c_j\|^2$$

dove S è l'insieme dei *cluster*, L è il set di dati da partizionare, K è il numero di *cluster* da individuare,  $x_i$  è il vettore *i*-esimo del set di dati (nel nostro caso il vettore costituito dalla pendenza e dalla quota associati alla cella *i*-esima del DEM) all'interno di un gruppo e  $c_j$  sono i vettori dei centri dei gruppi. Il modulo  $\|x_i^{(j)} - c_j\|$  denota una misura di distanza tra il centro di un gruppo ed un vettore di variabili. Nonostante sia possibile impiegare diversi tipi di misure di distanza, si è scelto di utilizzare quella euclidea, che produce gruppi di forma sferica nello spazio multivariato.

L'algoritmo necessita che sia determinato a priori il numero di gruppi nel quale suddividere il set di dati. Secondo il numero di gruppi scelto, il DEM risulta partizionato in K *cluster* (corrispondenti sul *raster* da insiemi di celle più o meno contigui spazialmente), costituiti da celle aventi valori simili di quota e pendenza.

Tramite il comando *r.clump* del pacchetto GIS è stato possibile effettuare lo *spatial clumping*, ossia individuare

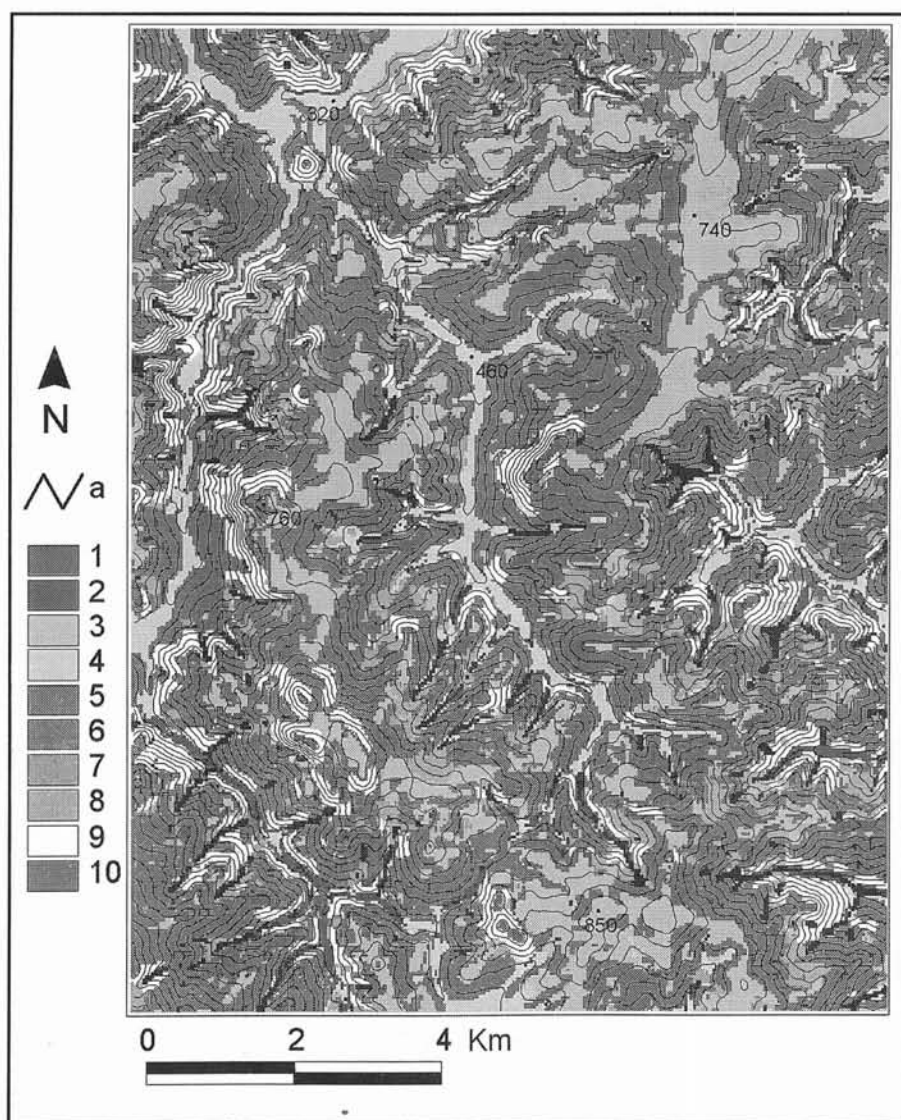


FIG. 3 - Risultato della classificazione non supervisionata. 1) Creste arrotondate; 2) Fondovalle a V; 3) Glacis e superfici di erosione sommitali; 4) Fondovalle; 5) Fianchi vallivi attuali; 6) Creste affilate; 7) Testate vallive acclivi; 8) Rami fluviali di ordine elevato; 9) Elementi a controllo strutturale; 10) Faccette di interfluvio, a) Linee di livello con equidistanza di 25 m.

aggregati di celle contigue. Il DEM risulta quindi suddiviso in un gran numero di *patche* (dipendente dalla contiguità delle *K* classi precedentemente individuate) su cui è possibile effettuare un'ulteriore classificazione. Gli elementi derivati dallo *spatial clumping*, essendo stati ottenuti a partire da proprietà geomorfologiche di rilevante importanza, dovrebbero costituire insieme di celle i cui domini corrispondono, almeno parzialmente, a forme del rilievo caratteristiche dell'area in esame. Il passo successivo consiste quindi nell'analisi delle proprietà dell'insieme delle celle associate a tali elementi. A tal fine si sono calcolati i valori medi di pendenza, curvatura del profilo e curvatura tangenziale tramite il comando *r.fpstat* che crea altrettante matrici per ogni parametro calcolato. Le matrici sono state quindi utilizzate per la classificazione finale tramite l'algoritmo *k-means*. In base a diverse prove basate su una valutazione visiva dei risultati ottenuti con diverse soluzioni (in termini di numero di *cluster*), si è scelto di suddividere il *raster* del bacino del Vella in 10 *cluster*.

## RISULTATI E DISCUSSIONE

Nella fig. 3 è possibile osservare il risultato della classificazione automatizzata. Essa evidenzia elementi che sono per la maggior parte riconducibili a quelle forme del rilievo che anche l'analisi visuale tradizionale evidenzia nell'area. Per l'esattezza i *cluster*, corrispondenti ad altrettante classi individuate, corrispondono in buona misura alle forme riportate manualmente sulla carta geomorfologica dell'area realizzata con tecniche tradizionali (vedi Aucelli, 1999 ed Aucelli & alii, 2001). Le classi di corrispondenza sono illustrate in fig. 3.

Malgrado il metodo abbia qualche limite ad evidenziare con chiarezza elementi di piccola estensione, quali creste molto affilate, esso riesce ad evidenziare bene macroforme di interesse quali ad esempio le paleosuperfici sommitali e GVS.

Questo risultato è molto importante per la ricostruzione di scenari morfoevolutivi in aree molto ampie. Nell'area di studio analizzata, data la mancanza di terrazzi fluviali, le superfici di erosione sommitale ed i GVS costituiscono gli unici *marker* utili a tale scopo. L'individuazione di tali elementi in automatico risulta di grande utilità soprattutto se si intende effettuare uno studio della distribuzione delle superfici a scala regionale. Ad esempio, le superfici individuate possono essere classificate automaticamente secondo l'ordine di generazione in base a considerazioni di carattere geometrico, analisi che risulterebbe sicuramente proibitiva con il solo ausilio delle carte topografiche. Nell'area analizzata risulta evidente la discriminazione automatica tra superfici sommitali e fondovalle, elementi che sono pure caratterizzati da ampie sovrapposizioni di molti parametri tra cui i valori di pendenza. C'è da sottolineare come tale discriminazione non sia dovuta alla differente posizione altimetrica poiché la quota non è stata usata tra i parametri della classificazio-

ne finale ma solo per l'individuazione di *patche* sulle quali poi effettuare l'effettiva classificazione. La distinzione dipende infatti dalle caratteristiche morfometriche dell'intera superficie rendendo il metodo applicato superiore a quelli di tipo tradizionale basati sulla sola analisi delle pendenze o di altri parametri, che si rivela insufficiente a discriminare forme peculiari del paesaggio che presentano parametri morfometrici analoghi ma che hanno genesi e/o età diverse. Anche l'uso diretto di metodi di analisi multivariata fallisce in tale obiettivo. Applicando, infatti, il *clustering* iterativo ai valori di pendenza e curvature associate alle singole celle del *raster*, le superfici di fondovalle risultano in buona parte classificate assieme alle superfici sommitali (fig. 4).

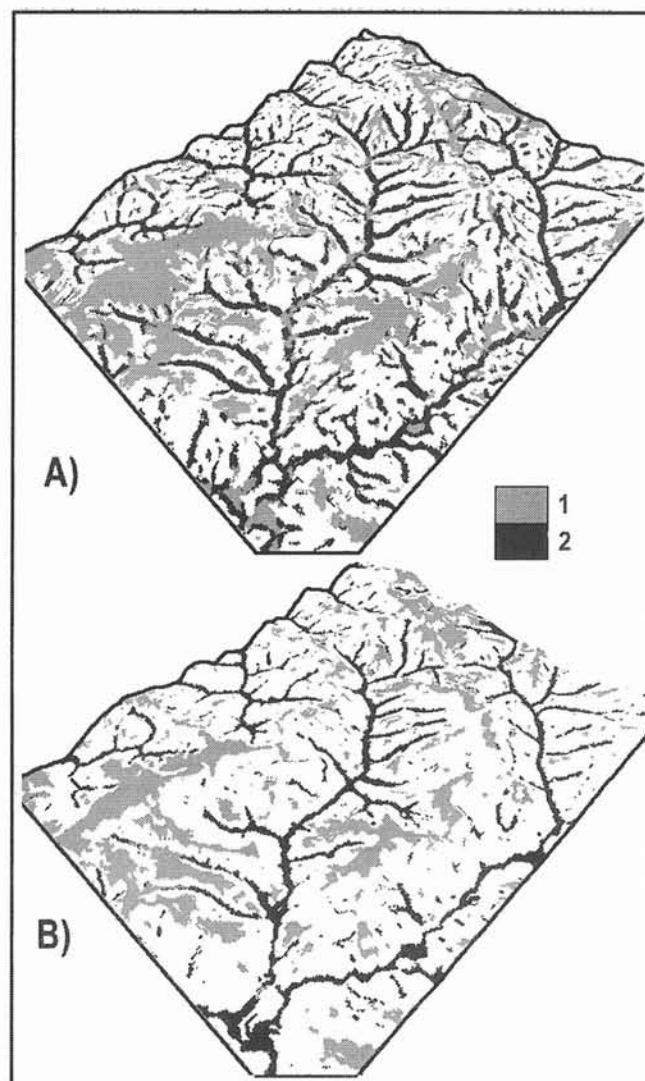


FIG. 4 - Risultato della classificazione non supervisionata impiegando. A) una classificazione non supervisionata utilizzando i valori delle singole celle del DEM e B) il metodo proposto; 1) Glacis e superfici di erosione sommitali; 2) Fondovalle e rami di basso ordine.

Il metodo proposto mostra come i modelli digitali del terreno costituiscono uno strumento di grande utilità per l'analisi geomorfologica, sia essa volta alla mappatura automatica di forme, che alla semplice analisi di elementi individuati tramite approcci più tradizionali in cui Giovanni Battista Castiglioni è stato per noi tutti un grande maestro.

I DEM sono utili per estrarre automaticamente attributi morfometrici e forme semplici del rilievo, ma per la analisi di forme complesse o per la mappatura automatica di paesaggi policiclici e/o poligenici i dati di elevazione che i DEM forniscono devono essere elaborati con metodi di selezione adeguati alla complessità morfologica dell'area in esame. L'approccio metodologico qui presentato mostra come l'uso appropriato del *clustering* consenta di estrarre strutture utili in tal senso. Nel presente lavoro tali strutture, corrispondenti ad insiemi di celle o *patche* del raster, sono state impiegate per estrarre informazioni utili a migliorare i risultati della classificazione non supervisionata delle forme comunemente cartografabili nell'area di studio. Benché tale metodo comporti una certa soggettività dovuta alla scelta del numero di *cluster*, la rapidità con la quale è possibile ottenere tali elaborati cartografici rispetto alle tecniche tradizionali consente di sperimentare diverse combinazioni di parametri. Le tecniche di classificazione non supervisionata possono inoltre essere proficuamente impiegate in congiunzione con tecniche di classificazione supervisionata per migliorare i risultati della classificazione; ad esempio impiegando gli elementi individuati dallo *spatial clumping* come *input* per algoritmi per la classificazione supervisionata.

- AUCELLI P.P.C. (1999) - *Analisi morfodinamica del bacino del fiume Trigno: distribuzione entità e cause dei fenomeni di alveo e di versante*. Tesi di Dottorato, XII Ciclo, Università degli Studi del Molise, 306 pp. (Inedita).
- AUCELLI P.P.C., CINQUE A. & ESPOSITO A. (2002) - *I GIS nella ricerca geomorfologica: L'evoluzione di lungo termine delle valli a deflusso adriatico del Molise indagata con l'ausilio del GIS GRASS*. Atti del Convegno «I sistemi di informazione geografica (GIS) nella gestione e lo sviluppo dell'ambiente e del territorio», Università degli Studi del Molise, Isernia, 20 Novembre 2002, 5-9.
- AUCELLI P.P.C., CINQUE A. & ROSSKOPF C. (2001) - *Geomorphological map of the Trigno river (Southern Italy): explanation notes*. Geogr. Fis. Dinam. Quat., 24, 3-12.
- BIVAND R.S. (1999) - *Integrating GRASS 5.0 and R: GIS and modern statistics for data analysis*. Proceedings of 7<sup>th</sup> Scandinavian Research Conference on Geographical Information Science, Aalborg, Denmark, 111-127.
- BROWN D.G., LUSCH D.P. & KENNETH A.D. (1998) - *Supervised classification of glaciated landscapes using digital elevation data*. Geomorphology, 21, 233-250.
- BURROUGH P.A., VAN GAANS P.F.M. & McMILLAN R.A. (2000) - *High resolution landform classification using fuzzy k-means*. Fuzzy sets and systems, 113, 1, 37-52.
- DIKAU R., BRABB E.E, MARK R.K. & PIKE R.J. (1995) - *Morphometric landform analysis of New Mexico*. Zeit. Geomorph., Suppl. Bd. 101, 109-126.
- IRVIN B.J., VENTURA S.J. & SLATER B.K. (1997) - *Fuzzy and isodata classification of landform elements from digital terrain data in Pleasant Valley Wisconsin*. Geoderma, 77, 2-4, 137-154.
- JAIN A.K., MURTY M.N. & FLYNN P.J. (1999) - *Data clustering: a review*. ACM Comp. Surveys, 31, 3, 264-323.
- MITASOVA H. & HOFIERKA J. (1993) - *Interpolation by regularized spline with tension: II. Application to terrain modelling and surface geometry analysis*. Math. Geol., 25, 657-669.
- NOGAMI M. (1995) - *Geomorphometric measures for digital elevation models*. Zeit. Geomorph., Suppl. Bd. 101, 53-67.
- WEBB A. (2002) - *Statistical Pattern Recognition*. Wiley, London, 496 pp.