



## **UNA METODOLOGIA DI CLASSIFICAZIONE DEGLI INCIDENTI IN AMBITO AUTOSTRADALE**

### **Gianluca Dell'Acqua**

Dipartimento di Ingegneria dei Trasporti "Luigi Tocchetti"  
Università degli Studi di Napoli "Federico II"  
Via Claudio 21- 80125 - Napoli - Italy  
Tel: +39.081.7683748  
Fax: +39.081.2390366  
E-mail: gdellacq@unina.it

### **Tommaso Esposito**

Dipartimento di Ingegneria dei Trasporti "Luigi Tocchetti"  
Università degli Studi di Napoli "Federico II"  
Via Claudio 21- 80125 - Napoli - Italy  
Tel: +39.081.7683370  
Fax: +39.081.2390366

### **Renato Lamberti**

Dipartimento di Ingegneria dei Trasporti "Luigi Tocchetti"  
Università degli Studi di Napoli "Federico II"  
Via Claudio 21- 80125 - Napoli - Italy  
Tel: +39.081.7683355  
Fax: +39.081.2390366  
E-mail: rlambert@unina.it

# UNA METODOLOGIA DI CLASSIFICAZIONE DEGLI INCIDENTI IN AMBITO AUTOSTRADALE

GIANLUCA DELL'ACQUA - D.I.T. Università degli Studi di Napoli "Federico II"

TOMMASO ESPOSITO - D.I.T. Università degli Studi di Napoli "Federico II"

RENATO LAMBERTI - D.I.T. Università degli Studi di Napoli "Federico II"

## SOMMARIO

Nelle scienze applicate l'incertezza viene in genere considerata come un fattore di disturbo da contenere o da eliminare, nel tentativo di pervenire, quando possibile, alla formulazione di modelli deterministici dei fenomeni considerati. Ciononostante, in numerosi problemi di ingegneria stradale quest'approccio non è sempre praticabile giacché i fenomeni in studio sono, tra l'altro, strettamente dipendenti dagli effetti del comportamento umano. I modelli che non ignorano di detti effetti sono evidentemente più realistici di quelli che li escludono.

L'incertezza nell'attività modellistica (formulazione, taratura e validazione di modelli) può essere efficacemente trattata impiegando metodologie fondate su logiche polivalenti.

Il fenomeno incidentalità stradale è affetto da un particolare tipo di incertezza intrinseca degli eventi che ne costituiscono le realizzazioni, connessa alla mancanza di informazioni riguardanti alcune interazioni esistenti tra le componenti del sistema (uomo, veicolo ed ambiente) che produce il fenomeno stesso ed alla quantità e qualità dei dati generalmente disponibili. Tuttavia la formazione di teorie e la costruzione di modelli interpretativi dell'incidentalità stradale devono essere necessariamente fondati sulla classificazione dei dati sperimentali, ma d'altra parte la natura degli eventi da analizzare è tale da rendere quanto mai improbabile e scarsamente significativo il raggruppamento di questi in classi disgiunte.

L'obiettivo dello studio è la formalizzazione di un metodo di validità generale per la ripartizione di eventi incidentali in classi continue, che consenta cioè di attribuire ciascun evento a più di una classe in misura diversa ed in relazione alle cause che lo hanno generato. I criteri impiegati consentono inoltre di riconoscere gli eventi cosiddetti "anomali" (extragrades) che se presenti all'interno dello spazio campionario devono essere riconosciuti, collocati in una classe non ordinaria e trattati differentemente dalle altre realizzazioni del fenomeno (intragrades), che possono essere invece opportunamente attribuite a classi ordinarie, al fine di evitare eventuali distorsioni dell'intero processo di raggruppamento.

La procedura è stata impiegata per analizzare 836 incidenti rilevati lungo l'Autostrada A3 Napoli-Salerno (di lunghezza pari a circa 50 Km) e verificatisi negli anni 1998, 1999 e 2000. La metodologia proposta può essere inquadrata nella particolare categoria di studi denominati "indagini generali". Tali studi non mirano in realtà, a valutare l'efficacia di un provvedimento; si tratta piuttosto di identificare dei fattori ricorrenti, sulla base di un campione di incidenti, e di dedurre delle contromisure possibili, suscettibili di formare ulteriore oggetto di uno studio sperimentale.

## ABSTRACT

In this paper preliminary results of accidents classification using fuzzy cluster algorithms are reported. The object of clustering is to divide a given data set into homogeneous groups, that is all accidents in the same cluster share similar attributes and they do not share similar attributes with accidents in other clusters. The classification process was carried out using the fuzzy c-means method with extragrades. The basic idea is to classify accidents according to their patterns and causes into one or a combination of classes recognizing the complex interaction among accident factors and the uncertainties associated with accident data. Owing to their nature, continuous classes should provide better representations of outliers or atypical individuals than discontinuous classes. This is especially the case with outliers located between clusters in property space (intragrades). However, outliers outside the main body of data points (extragrades) are still not suitably represented by traditional fuzzy cluster algorithms. It is necessary to set the outlier data points in a special class to decrease the effect of them on classification. The members of this particular class are not concentrated in a fuzzy hypersphere around a defined class centre, as with regular classes. Instead, they are spread across and over regions of larger distances between an individual and the class centres. The object of the study are the accidents happened along the highway A3, Naples-Salerno; the data base is composed by 836 accident cards collected in the period August 1998 - March 2000.

## 1. INTRODUZIONE

I dati riguardanti la incidentalità stradale sono generalmente vaghi, insufficienti, poco affidabili e in alcuni casi rappresentati da stime e non da rilievi. Da ciò scaturisce l'esigenza di impiegare algoritmi numerici in logiche non tradizionali che, dalla schematizzazione degli incidenti come i punti di uno spazio di parametri, quali ad esempio le caratteristiche della strada (curvatura, pendenza longitudinale, pavimentazione, ecc.), le condizioni meteorologiche, di traffico, ecc. consentisse di ripartirli in insiemi fuzzy costituiti da elementi simili. Le tecniche impiegate, estensioni di quelle già proposte in un lavoro precedente [3], producono risultati tecnicamente più validi.

## 2. RICHIAMI METODOLOGICI

Le classi di una partizione possono essere discontinue (crisp) o continue (fuzzy). La classe continua è una generalizzazione di quella discontinua [1]. L'obiettivo di ogni processo di classificazione è la ricerca di una famiglia di insiemi  $\{\underline{A}_i, i = 1, 2, \dots, k\}$  ovvero la definizione di una  $k$ -partizione dello spazio campionario. La partizione di un insieme di  $n$  elementi in  $k$  classi discontinue produce l'assegnazione di ciascun individuo ad una ed una sola classe. Una simile partizione è formalizzata mediante la matrice delle appartenenze  $\mathbf{M} = (m_{ic})$  di dimensioni  $[n \times k]$ . Il singolo elemento  $m_{ic}$  della matrice  $\mathbf{M}$  vale 1 se l'individuo  $i$ -esimo appartiene alla  $c$ -esima classe e 0 se non vi appartiene, ossia la funzione caratteristica della  $c$ -esima classe ha condominio  $m_{ic} \in \{0,1\}$ . Le classi di una partizione crisp sono dunque mutuamente esclusive, congiuntamente esaustive e non vuote. Le matrici  $\mathbf{M} [n \times k]$  che soddisfano le relazioni seguenti rappresentano invece una partizione fuzzy di  $n$  individui in  $k$  classi continue:

$$\sum_{c=1}^k m_{ic} = 1 \quad i=1 \dots n \quad (1)$$

$$\sum_{i=1}^n m_{ic} > 0 \quad c=1 \dots k \quad (2)$$

$$m_{ic} \in [0,1] \quad i=1 \dots n \quad c=1 \dots k \quad (3)$$

ciascun individuo del campione appartiene a più di una classe in misura diversa.

## 2.1 Metodi delle k medie

La generica matrice  $\mathbf{M}$  definisce una partizione dell'insieme dei dati. L'espressione della funzione obiettivo per la ricerca della partizione ottima caratterizza i diversi metodi di analisi dei gruppi reperibili in letteratura. La funzione obiettivo del metodo fuzzy delle k medie (FkM), generalizzazione del metodo hard delle k medie (HkM), minimizza in ogni classe la somma dei quadrati delle distanze del singolo elemento dal centro della classe supposte valide le condizioni (1), (2), (3). Sia  $\mathbf{X}_i = (x_{i1}, \dots, x_{ip})$  il vettore che individua la posizione dell'elemento campionario i-esimo all'interno dello spazio delle caratteristiche, la forma analitica della funzione obiettivo del metodo HkM è così esplicitata:

$$J(\mathbf{M}, \mathbf{C}) = \sum_{i=1}^n \sum_{c=1}^k m_{ic} * d^2(x_i, c_c) \quad (4)$$

in cui  $\mathbf{C} = (c_{cv})$  è una matrice  $[k * p]$  il cui generico elemento  $c_{cv}$  è il valore che la variabile  $v$ -esima attinge in corrispondenza del centro della classe  $c$ -esima, ossia è la coordinata  $v$  del centro medesimo individuato dal vettore  $c_c = (c_{c1}, \dots, c_{cp})$ . La grandezza  $d^2(x_i, c_i)$  è il valore del quadrato della distanza misurata tra i vettori  $\mathbf{X}_i$  e  $\mathbf{C}_i$ . La funzione obiettivo del metodo FkM è una generalizzazione della (4) opportunamente modificata attribuendo all'esponente del termine  $m_{ic}$  un valore  $\phi$  diverso dall'unità [7]:

$$J_B(\mathbf{M}, \mathbf{C}) = \sum_{i=1}^n \sum_{c=1}^k m_{ic}^\phi * d_{ic}^2 \quad (4a)$$

L'esponente  $\phi$  teoricamente può assumere valori compresi nell'intervallo  $]1, \infty]$  ed è una misura dell'incertezza ("fuzziness") della soluzione. Per valori prossimi all'unità di  $\phi$  [ $\phi \cong 1$ ] le classi sono quasi disgiunte e la partizione, seppur fuzzy, si approssima ad una partizione crisp; d'altro canto un valore del parametro  $\phi$  elevato [ $\phi \rightarrow \infty$ ] produce classi sovrapposte, una partizione molto incerta e valori della funzione di appartenenza  $m_{ic} = 1/k$ . Inoltre, con riferimento agli aspetti computazionali, occorre precisare che il valore attinto dalla funzione obiettivo in corrispondenza della partizione ottima decresce monotonicamente all'aumentare del numero delle classi  $k$  e del valore dell'esponente  $\phi$ . La ricerca della migliore partizione presuppone quindi la scelta dei valori dei parametri  $\phi$  e  $K$  e del criterio di similarità più adeguato (tipo di distanza tra i punti dello spazio) di cui si riferirà di seguito. Il valore dell'esponente fuzzy  $\phi$  può essere individuato massimizzando rispetto a  $\phi$  la funzione:

$$W(\phi) = -[(dJ/d\phi) * v(k)]$$

si noti la complessità del problema essendo la funzione  $W(\phi)$  [5] [6] dipendente dal numero delle classi  $k$  direttamente ed indirettamente (attraverso la funzione derivata della funzione obiettivo). La scelta del valore da attribuire al parametro  $\phi$  è dunque vincolata alla definizione del numero delle classi  $k$  per questo il problema è risolvibile solo per tentativi. Con riferimento alla scelta del criterio di similarità occorre precisare che la distanza tra i punti del medesimo spazio vettoriale può essere misurata in modi diversi. Un criterio di similarità frequentemente adottato quando si operano analisi dei gruppi è fondato sull'impiego della distanza *Euclidea*:

$$d_{ic}^2 = \sum_{v=1}^p (x_{iv} - c_{cv})^2 = (x_i - c_c)' * (x_i - c_c) \quad (5)$$

in tale caso alle variabili è attribuito lo stesso peso nel processo di raggruppamento quindi le variabili a varianza più elevata influenzano il risultato più di quanto accada per le altre. Quest'effetto indesiderato può essere eliminato mediante un processo di standardizzazione delle variabili stesse o in alternativa utilizzando un criterio di similarità differente fondato ad esempio sulla distanza *diagonale* o sulla distanza di *Mahalanobis* che controlla anche la dipendenza statistica tra variabili differenti:

$$d_{ic}^2 = (x_i - c_c)' * \Sigma^{-1} * (x_i - c_c) \quad (5a)$$

dove  $\Sigma$  è la varianza campionaria dello spazio vettoriale formalizzato mediante la matrice  $M$ . Il valore minimo della funzione (4a) deve essere ricercato mediante un algoritmo di ottimizzazione iterativa fondato sulla risoluzione delle relazioni seguenti:

$$m_{ic} = d_{ic}^{-2/(\phi-1)} / \sum_{j=1}^k d_{ij}^{-2/(\phi-1)} \quad i=1, \dots, n \quad c=1, \dots, k \quad (6)$$

$$c_c = \sum_{i=1}^n m_{ic}^\phi * x_i / \sum_{i=1}^n m_{ic}^\phi \quad c=1, \dots, n. \quad (7)$$

La successione delle fasi che compongono l'intero processo di analisi sopra descritto sono sintetizzate nell'elenco seguente:

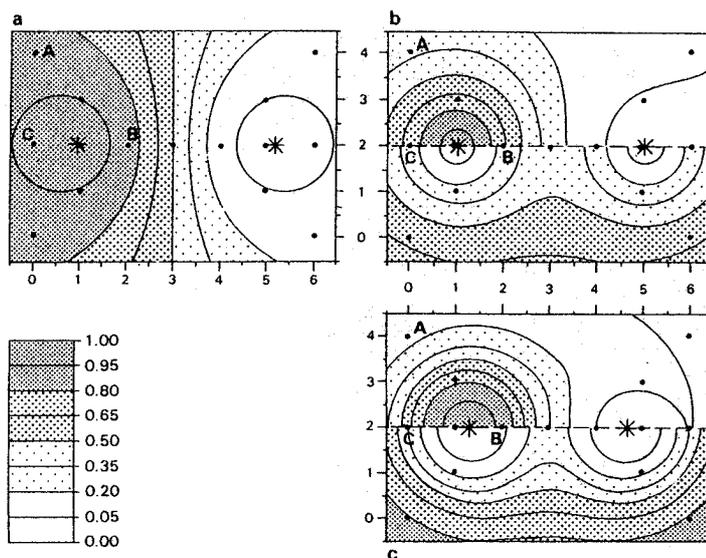
- (a) scegliere il numero delle classi  $k$  ( $1 < k < n$ );
- (b) individuare il valore dell'esponente fuzzy  $\phi$  ( $\phi > 1$ );
- (c) scegliere un criterio di similarità e quindi una misura della distanza tra i punti campionari (Euclidea, diagonale, di Mahalanobis, ecc.);
- (d) assegnare un valore per la tolleranza  $\epsilon$  della soluzione ( $\epsilon=0.001$  fornisce un risultato accettabile);
- (e) ipotizzare una partizione casuale di primo tentativo  $M=M^{(0)}$ ;
- (f) individuare la posizione dei centroidi ossia le coordinate dei vettori  $C=C^{(i-1)}$  applicando la relazione (7) alla matrice  $M^{(0)}$  (in generale,  $M^{(i-1)}$ );

- (g) definire la matrice  $M=M^{(1)}$  (in generale,  $M^i$ ) mediante la (6) in base alle coordinate dei vettori  $C^{(1)}$ . La relazione (6) non può essere utilizzata quando la distanza  $d_{ic}$  è prossima a zero, in tale caso si assume valore unitario al grado d'appartenenza;
- (h) confrontare  $M^{(i)}$  con  $M^{(i-1)}$ , se  $\|M^{(i)} - M^{(i-1)}\| \leq \epsilon$  allora arrestare il procedimento, altrimenti ripercorrere il processo dalla fase (f).

## 2.2 Il problema degli extragrades

I metodi k-means tuttavia presentano degli inconvenienti ed in particolare tendendo alla formazione di gruppi sferoidali possono attribuire ad elementi localizzati nelle aree marginali dello spazio campione (extragrades), e dunque prossimi al “guscio” dei gruppi stessi, valori del grado di appartenenza paragonabili a quelli assegnati ad elementi prossimi al nocciolo dei gruppi (intragrades).

La limitazione che caratterizza il metodo è dunque non trascurabile ed influenza negativamente il processo di raggruppamento e di conseguenza la successiva fase di riconoscimento della struttura (pattern) presente all'interno del database. Si consideri a titolo di esempio uno spazio campionario bidimensionale da ripartire in due classi imponendo  $\phi = 2$ . Lo spazio campione (vedi figura 1a) presenta una forma strutturale nota come configurazione a “farfalla”.



**Figura 1 – bipartizioni di uno spazio bidimensionale ( $k = 2, \phi = 2$ )**

Dalla figura (1a) si evince chiaramente che la bipartizione prodotta applicando il metodo FkM attribuisce ai punti A e B valori del grado di appartenenza quasi uguali. Il metodo dunque non riconosce il punto A, collocato ai margini dello spazio campione, come un punto singolare (“anomalo”) rispetto agli altri punti del campione producendo una rappresentazione distorta della struttura (forma) del database. Allo scopo di fornire una soluzione del problema è stata proposta una modifica del metodo FkM variando l’espressione della funzione obiettivo [2]:

$$J_0(M, C) = \alpha * \sum_{i=1}^n \sum_{c=1}^k m_{ic}^\phi * d_{ic}^2 + (1 - \alpha) * \sum_{i=1}^n m_{i*}^\phi \quad (8)$$

in cui  $m_{i*}^\phi$  è il valore del grado di appartenenza del punto  $i$ -esimo ad un gruppo non ordinario detto degli "extragrades". La funzione obiettivo può essere minimizzata impiegando il medesimo algoritmo proposto per la risoluzione della (4a), la relazione (7) per determinare i centri dei gruppi e le relazioni di seguito riportate per la determinazione del grado di appartenenza:

$$m_{ic} = \frac{d_{ic}^{-2/(\phi-1)}}{\sum_{j=1}^k d_{ij}^{-2/(\phi-1)} + \left(\frac{1-\alpha}{\alpha}\right)^{-1/(\phi-1)}} \quad i=1, \dots, n \quad c=1, \dots, k$$

$$m_i^\phi = \frac{\left(\frac{1-\alpha}{\alpha}\right)^{-1/(\phi-1)}}{\sum_{j=1}^k d_{ij}^{-2/(\phi-1)} + \left(\frac{1-\alpha}{\alpha}\right)^{-1/(\phi-1)}} \quad i=1, \dots, n.$$

Il gruppo degli "extragrades", a differenza di quelli ordinari, non si configura come un'ipersfera fuzzy ma come una superficie irregolare distante dai centroidi degli altri gruppi (vedi Figura 1b). Il grado di appartenenza dei punti campionari alla classe degli "extragrades" ( $m_{i*}$ ) è una funzione crescente della distanza tra questi ed i centroidi dei gruppi ordinari. La possibilità di distinguere adeguatamente gli extragrades dagli intragrades mediante la funzione di appartenenza  $m_i^*$  dipende dalla localizzazione dei centroidi dei gruppi ordinari. La minimizzazione della (8) richiede di specificare il valore di  $\alpha$  che determina il grado di appartenenza medio dei punti al gruppo degli "extragrades". Nell'esempio proposto in figura 1b, per ragioni di cui si riferirà nel seguito, al coefficiente  $\alpha$  è stato attribuito un valore che producesse un grado di appartenenza medio alla classe degli extragrades uguale a quello relativo ai due gruppi ordinari. Con riferimento alla figura 1b si noti che la (8) consente di differenziare in una certa misura i gradi di appartenenza di A e B pur non riconoscendo sufficientemente al punto A lo status di extragrade quindi il metodo FkM così modificato non risolve in via definitiva il problema per il quale è stato proposto. Un'ulteriore modifica del metodo FkM presuppone la dipendenza diretta della funzione di appartenenza del gruppo degli extragrades dalla distanza misurata tra i punti campionari ed i centroidi [2]:

$$J_{M,G}(M, G) = \alpha * \sum_{i=1}^n \sum_{c=1}^k m_{ic}^\phi * d_{ic}^2 + (1 - \alpha) * \sum_{i=1}^n m_{i*}^\phi * \sum_{c=1}^k d_{ic}^{-2} \quad (10)$$

inoltre si assume convenzionalmente che nel caso in cui simultaneamente  $m_{i*}$  e  $d_{ic}$  siano uguali a zero il contributo del secondo membro è nullo. L'algoritmo di risoluzione della (10) è quello del metodo FkM tradizionale in cui  $m_{ic}$ ,  $m_{i*}$ ,  $c_c$  assumono le espressioni seguenti:

$$m_{ic} = \frac{d_{ic}^{-2/(\phi-1)}}{\sum_{j=1}^k d_{ij}^{-2/(\phi-1)} + \left( \frac{1-\alpha}{\alpha} * \sum_{j=1}^k d_{ij}^{-2} \right)^{-1/(\phi-1)}} \quad i=1, \dots, n \quad c=1, \dots, k \quad (11a)$$

$$m_{i*} = \frac{\left( \frac{1-\alpha}{\alpha} * \sum_{j=1}^k d_{ij}^{-2} \right)^{-1/(\phi-1)}}{\sum_{j=1}^k d_{ij}^{-2/(\phi-1)} + \left( \frac{1-\alpha}{\alpha} * \sum_{j=1}^k d_{ij}^{-2} \right)^{-1/(\phi-1)}} \quad i=1, \dots, n \quad (11b)$$

$$c_c = \frac{\sum_{i=1}^n \{ m_{ic}^\phi - (1-\alpha) * \alpha^{-1} * d_{ic}^{-4} * m_{i*}^\phi \} * x_i}{\sum_{i=1}^n \{ m_{ic}^\phi - (1-\alpha) * \alpha^{-1} * d_{ic}^{-4} * m_{i*}^\phi \}} \quad c=1, \dots, k \quad (11c)$$

Nei casi in cui la distanza  $d_{ic}$  sia prossima o uguale a zero possono insorgere problemi di "overflow numerico" per cui in tali casi al grado di appartenenza ed ai pesi degli elementi si attribuisce valore unitario nei casi in cui sia  $d_{ic} = 0$  e nullo in tutti gli altri casi. Inoltre la (11c) funzione del termine  $d_{ic}^{-4}$ , a sua volta dipendente dalla posizione dei centri delle classi  $c_c$ , non può essere risolta in via diretta ma solo attraverso una valutazione approssimata di  $C^{(1)}$  quantificando quindi  $d_{ic}^{-4}$  ed iterando il procedimento.

In alternativa è possibile evitare l'ottimizzazione della posizione dei centroidi procedendo alla stima diretta della matrice  $M^{(1)}$  atteso che la differenza tra  $C^{(i-1)}$  e  $C^{(1)}$  comunque decresce rapidamente nel corso delle iterazioni. In Figura 1c sono indicati i risultati derivanti dall'applicazione del metodo FkM modificato secondo la (11) dai quali si evince l'efficienza del metodo stesso per la risoluzione del problema degli extragrades. In figura 2 sono consegnati i risultati derivanti dall'applicazione dei tre metodi illustrati per la partizione in quattro gruppi di un database bidimensionale. Lo spazio campione presenta una forma strutturale nota come configurazione a "griglia nidificata".

### 2.3 Indici di prestazione

In letteratura sono reperibili diversi tipi di indicatori di prestazione delle partizioni che possono essere impiegati per la scelta del numero dei gruppi [8] [9]. Gli indicatori sono funzioni della matrice di partizione  $\mathbf{M}$  e del valore minimo che la funzione obiettivo attinge in corrispondenza della partizione stessa. Di seguito sono illustrati gli indicatori impiegati nel lavoro in oggetto:

$$F(\mathbf{M}) = \sum ( \sum ( m_{ik} )^2 / n ) \quad F(\mathbf{M}) \in [1/k; 1] \quad (12)$$

$$H(\mathbf{M}) = \sum [ \sum ( m_{ik} ) \log(a) (\mu_{ik}) ] / n \quad \text{con } 0 < a < 8 \quad H(\mathbf{M}) \in [0; \log.(c)] \quad (13)$$

$$S(\mathbf{M}) = J_{\min} / ( n \cdot \min |v_i - v_j|^2 ) \quad (14)$$

L'indicatore più efficiente tra quelli utilizzati è  $S(\mathbf{M})$  poiché  $F(\mathbf{M})$  ed  $H(\mathbf{M})$  decrescono monotonicamente al crescere del numero dei gruppi.

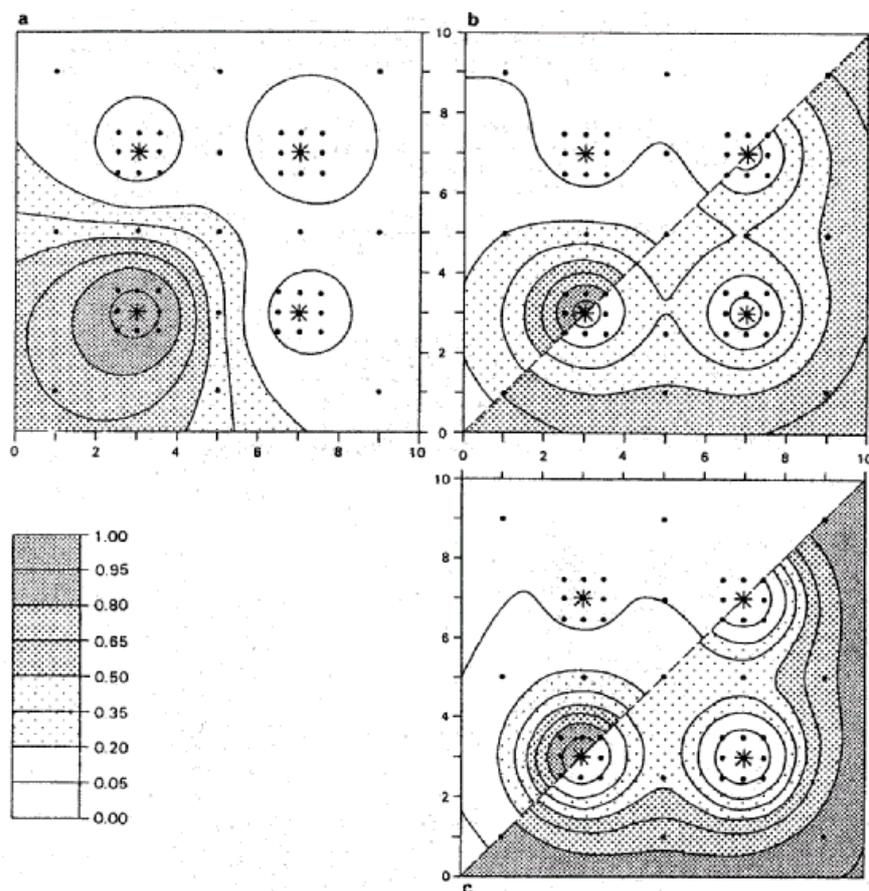


Figura 2 - partizioni di uno spazio bidimensionale ( $k = 4, \phi = 2$ )

### 3. IL CASO APPLICATIVO

Le tecniche illustrate nei paragrafi precedenti sono state impiegate per analizzare l'incidentalità rilevata lungo l'autostrada A3 Napoli-Salerno. L'autostrada A3 ha origine nella zona orientale di Napoli, nei pressi della stazione ferroviaria, e si sviluppa lungo la costa tirrenica fino a Castellammare di Stabia attraversando Ercolano, Torre del Greco e Pompei; successivamente s'inoltra verso l'interno per collegare i centri di Scafati, Angri, Nocera e Cava dei Tirreni, tagliando la penisola sorrentina e riavvicinandosi sulla costa, nei pressi di Vietri sul Mare, per raggiungere Salerno e congiungersi, senza soluzione di continuità, all'autostrada Salerno-Reggio Calabria.

L'autostrada, gestita dalla Società Autostrade Meridionali (SAM), si sviluppa per una lunghezza di 51.7 km. La sezione trasversale presenta una larghezza di 18 m e a doppia carreggiata con due corsie di 3.75 m per senso di marcia e banchine di larghezza 1.00 m, la larghezza dello spartitraffico è pari ad 1.00 m. I dati riguardanti gli incidenti, processati secondo lo schema delle elaborazioni illustrato in Figura 3, sono stati estratti prevalentemente dai rapporti di incidente redatti dalla polizia stradale. Tuttavia le informazioni estratte dai rapporti, come era prevedibile, si sono rivelate insufficienti allo scopo. È stato necessario quindi completare il database con i valori del traffico

giornaliero direzionale, del coefficiente di aderenza trasversale e della tessitura della pavimentazione, i parametri geometrici (curvatura dell'asse, pendenza longitudinale, ecc.), i limiti di velocità e le velocità attuate riferibili a ciascun incidente esaminato.

Il modello di rilevazione degli incidenti SAM è un questionario predisposto dall'ISTAT ed è compilato dalle pattuglie della Polizia Stradale che intervengono sul luogo del sinistro. Le schede d'incidente constano di due fogli, il primo dei quali contiene il questionario in forma chiusa in linea con le esigenze di concisione ed affidabilità delle risposte, mentre nel secondo è riservato uno spazio dedicato alla raccolta di informazioni riguardanti i veicoli coinvolti (tipo, targa, assicurazione, proprietario, conducente) e sulla dinamica dell'incidente. La prima parte del rapporto è stata utilizzata integralmente, la seconda invece è stata consultata unicamente per desumere la presunta dinamica degli incidenti ed il luogo di residenza degli utenti coinvolti nel sinistro a causa dei vincoli imposti dalla legge sulla privacy che tutela i dati personali. Per lo studio in oggetto sono stati analizzati 893 rapporti relativi ad incidenti verificatisi in un arco temporale della durata di 20 mesi (agosto 1998 - marzo 2000), tuttavia a causa di omissioni e/o incoerenze nella compilazione dei questionari, i dati relativi a 57 incidenti (il 6.38 % del totale) si sono rivelati inutilizzabili.

#### **4 ELABORAZIONI**

Il database sopra descritto è stato processato per il riconoscimento di una eventuale struttura interna ad esso che consentisse di analizzare i caratteri principali dell'incidentalità rilevata. A tal fine si è proceduto a frazionare il database in gruppi di incidenti simili per poter sintetizzare le caratteristiche che qualificassero ciascun gruppo distinguendolo dagli altri. Gli 836 incidenti, che costituiscono il campione, sono stati quindi rappresentati mediante punti di uno spazio vettoriale a 14 dimensioni:

$$X = \{x_1, \dots, x_{836}\}: \forall x_i = (x_{i1}, \dots, x_{i14}) \text{ con } i = 1, \dots, 836$$

in cui le coordinate degli  $n$  incidenti sono i livelli che le variabili caratteristiche attingono in corrispondenza di ciascuno di essi. Dunque  $x_i$ ,  $i$ -esimo punto campionario, rappresenta incidente  $i$ -esimo. Il criterio di similarità impiegato è la distanza di Mahalanobis. La matrice di input composta da 836 righe e 14 colonne è la formalizzazione vettoriale del campione di incidenti. Le elaborazioni (vedi Figura 3) sono state sviluppate su due livelli:

1. applicazione del metodo delle  $k$ -medie con extragrades in logica binaria;
2. applicazione del metodo delle  $k$ -medie con extragrades in logica fuzzy.

##### **4.1 Scelta delle variabili d'incidente**

Le informazioni ricavate dai rapporti di incidente e da altre fonti sono state codificate attraverso 14 variabili, continue e discrete, i cui valori (modalità, o livelli) assunti in corrispondenza di ciascun evento, sono rappresentativi delle caratteristiche di questo. Di seguito sono riportate le descrizioni delle variabili menzionate.

###### **4.1.1 Intensità di traffico**

La variabile "Intensità di traffico" è stata ottenuta collegando la collocazione temporale dell'evento, riportata sulle schede come ora d'inizio dell'incidente, ai volumi di traffico mediamente presenti sull'autostrada in funzione della fascia oraria; si è proceduto alla

suddivisione delle 24 ore in fasce orarie, a ciascuna delle quali è stata, poi attribuito uno dei tre seguenti livelli di intensità di traffico:

- Ore diurne di punta = traffico intenso;
- Ore diurne restanti = traffico normale;
- Ore notturne = traffico scarso.

#### **4.1.2 Luminosità**

La variabile “Luminosità” è stata codificata utilizzando l’informazione relativa all’ora d’inizio dell’incidente e quella relativa alla data del sinistro ossia al periodo stagionale in cui questo si è verificato, in tal modo è stato possibile attribuire a ciascun punto campione uno dei tre gradi di luminosità seguenti: pieno giorno, alba/tramonto o notte.

#### **4.1.3 Traffico Giornaliero**

I dati di traffico forniti dall’ente gestore sono relativi al numero di biglietti emessi giornalmente dalla totalità delle stazioni; il dato riguarda, però, solo il numero di veicoli che accedono all’autostrada nel corso della giornata ma non fornisce indicazioni sulle percorrenze. Al fine di pervenire ad una stima seppur approssimativa del traffico giornaliero (TG) per ciascuna direzione di marcia si è correlato il dato suddetto ai valori del traffico giornaliero medio AISCAT forniti su base mensile forniti per entrambe le direzioni.

#### **4.1.4 Limite di velocità**

I livelli che la variabile “Limite di velocità” attinge sono i valori dei limiti imposti dal codice della strada (segnalati e non segnalati) nei casi in cui l’infrastruttura si presenta in condizioni ordinarie e quelli debitamente segnalati in presenza cantieri. I valori sono ricavati dai rapporti d’incidente e dalla planimetria della segnaletica fornita dall’ente gestore. La variabile attinge valori compresi tra i 40 ed i 90 km/h, nei punti in cui è presente apposita segnaletica ed un valore pari a 130 km/h (limite massimo consentito dalla legge) nelle zone prive di specifiche segnalazioni.

#### **4.1.5 Aderenza**

La variabile “Aderenza” è stata quantificata mediante i valori delle misure annuali effettuate dai tecnici della SAM con il SUMMS (survey machine for macrottexture and skid) nell’ambito dei programmi di monitoraggio e manutenzione. La strumentazione rende disponibili come è noto misure dell’aderenza trasversale (CAT) e della macrotessitura (TEX).

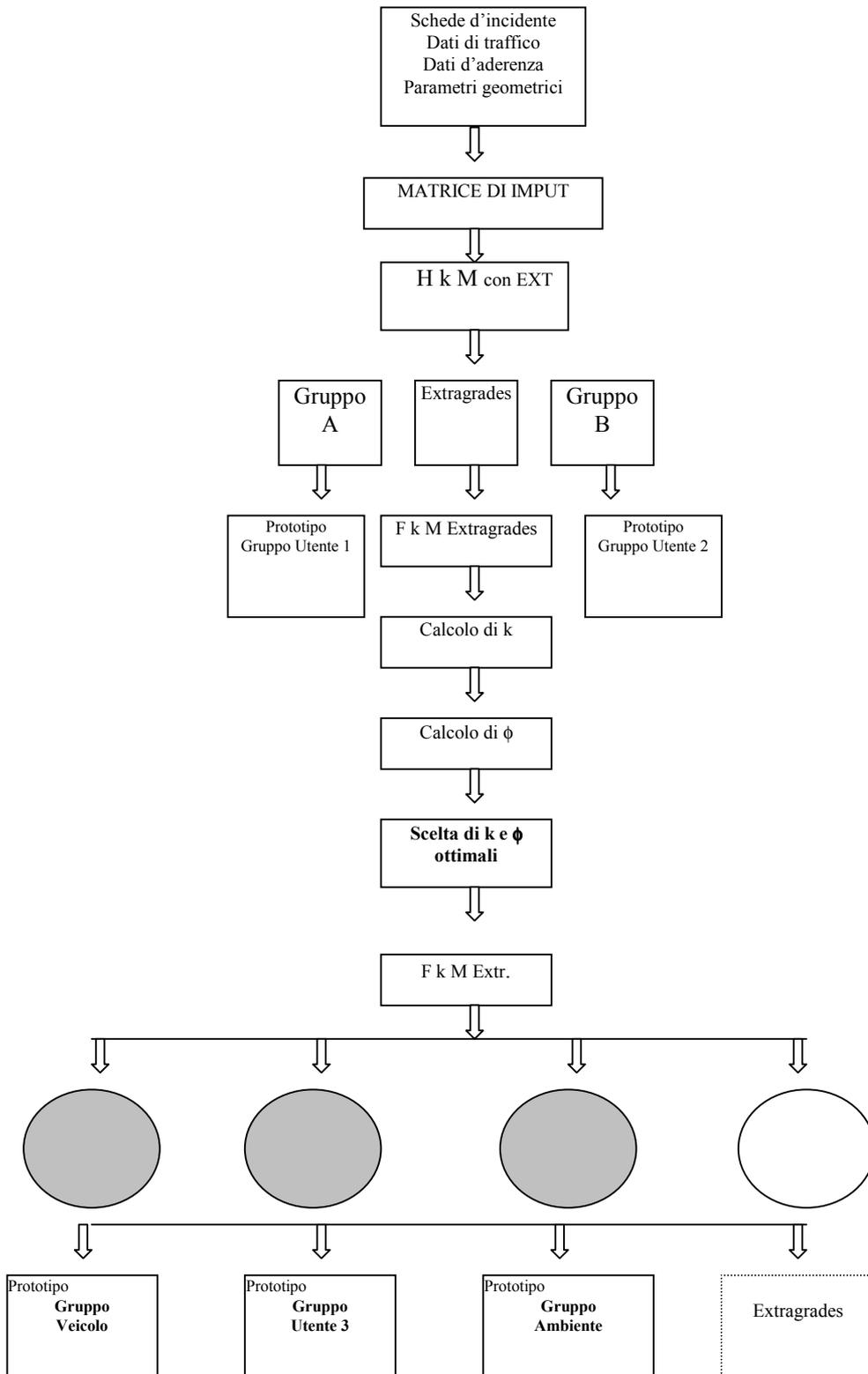
#### **4.1.6 Curvatura**

La variabile “Curvatura” esprime il valore della curvatura dell’elemento su cui si è verificato l’incidente. Le dimensioni dei raggi sono stati ricavati dall’andamento planimetrico dell’autostrada fornito in forma di aerofotogrammetrie (in scala 1:2000) dalla SAM.

#### **4.1.7 Pendenza longitudinale**

La variabile “Pendenza longitudinale” esprime il valore, in termini percentuali, della pendenza del tronco sul quale è localizzato l’incidente. I valori delle pendenze

longitudinali sono stati ricavati approssimativamente dalle quote riportate sulle aerofotogrammetrie (in scala 1:2000) fornite dalla SAM.



**Figura 3 – diagramma di flusso delle elaborazioni**

#### **4.1.8 Ottantacinquesimo percentile della distribuzione delle velocità**

La variabile “ $V_{85}$ ” è definita come il valore della velocità che non è superato dall’85% degli utenti in condizioni di flusso libero su strada asciutta. I valori, nel caso in esame, sono stati calcolati mediante una relazione di tipo regressivo opportunamente modificata per il caso di studio:  $V_{85}=(130 - 1.8 \times |i \%|) \times (1 - 34.61 / R)$ . La velocità in piano ed in rettilineo è stata posta pari 130 km/h in quanto la A3 presenta è caratterizzata dalla presenza di numerose immissioni, banchina e spartitraffico di dimensioni ridotte che la rendono per certi versi simile ad arterie di categoria inferiore.

#### **4.1.9 Localizzazione**

La variabile indica la localizzazione dell’incidente: corsia di marcia o di sorpasso, svincoli, stazioni di servizio, ecc.

#### **4.1.10 Fondo stradale**

La variabile indica le condizioni del fondo stradale al momento del sinistro: asciutto, bagnato, presenza di neve e/o ghiaccio, ecc.

#### **4.1.11 Condizioni meteorologiche**

La variabile indica le condizioni meteorologiche al momento del sinistro: tempo sereno o nuvoloso, pioggia ecc.

#### **4.1.12 Tipo di Veicolo**

La variabile indica la tipologia del veicolo che presumibilmente ha provocato l’incidente: auto di cilindrata inferiore superiore ai 1000 cc, autocarro, autoarticolato, ecc.

#### **4.1.13 Avaria**

La variabile, di tipo dicotomico (sì/no), indica l’eventuale avaria occorsa al veicolo che presumibilmente ha provocato l’incidente.

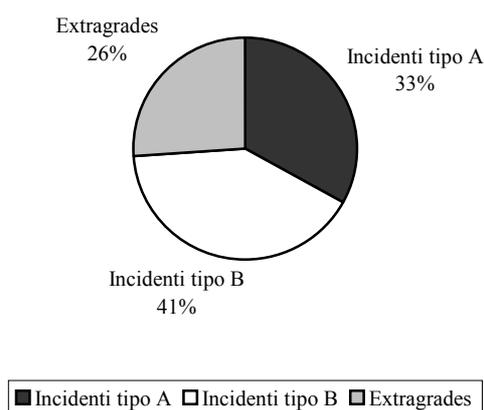
#### **4.1.14 Tipo di utente**

La variabile, di tipo dicotomico, qualifica il conducente del veicolo che presumibilmente ha provocato l’incidente: abituale o non abituale.

### **5. ANALISI DEI RISULTATI E CONCLUSIONI**

Le rilevazioni degli incidenti, impiegate nelle analisi, sono corredate come detto dalle deduzioni del rilevatore sulla dinamica dell’evento fondate spesso sulle dichiarazioni degli utenti coinvolti nel sinistro. Nella maggior parte dei casi esaminati, inoltre, l’esito delle diagnosi imputa esclusivamente al comportamento scorretto dell’utente il verificarsi dell’incidente stradale. E’ noto invece che, la condotta di marcia è l’effetto delle sollecitazioni che l’ambiente autostradale produce sull’utente per cui ad esso deve essere attribuita solo una quota, spesso prevalente, delle responsabilità del verificarsi del sinistro. Tali considerazioni sono confermate dai risultati prodotti e di seguito illustrati. Inizialmente si è provveduto a ripartire il database in tre gruppi (vedi Figura 4) impiegando il metodo HkM modificato con extragrades ed illustrato nei paragrafi precedenti.

Le caratteristiche degli incidenti appartenenti ai gruppi A e B sono indicate in Tabella 1. Il gruppo degli extragrades (26%) non è, per definizione, qualificabile giacché contiene incidenti che possono essere accomunati unicamente dal fatto di non potere essere inseriti nei gruppi A e B. Gli incidenti ricadenti nel gruppo A (33%) avvengono in condizioni di luminosità favorevoli, di traffico normale e sono localizzati lungo la corsia di marcia. Gli incidenti del gruppo B (41%) si verificano in condizioni di luminosità sfavorevoli, di traffico intenso e sono localizzati lungo la corsia di sorpasso. Dalla verifica dei livelli attinti dalle variabili in corrispondenza degli incidenti che ricadono nei gruppi A e B non si desume in definitiva un'influenza rilevante delle componenti ambiente stradale e/o veicolo. La conferma di questo assunto deriva dalla rilettura a valle delle elaborazioni delle schede degli incidenti dalle quali si evince la responsabilità quasi esclusiva dell'utente in particolare addebitabile a sonnolenza, malore, distrazione, mancato rispetto delle distanze di sicurezza, velocità eccessiva, manovre pericolose ed inosservanza delle norme di sorpasso. Successivamente il gruppo degli extragrades, ottenuto dalla elaborazione di primo livello, è stato decomposto in ulteriori quattro gruppi impiegando il metodo FkM con extragrades. L'elaborazione di secondo livello ha condotto al riconoscimento di un primo gruppo di incidenti attribuibile in misura rilevante alla componente veicolo, un secondo attribuibile in prevalenza alla componente ambiente stradale ed un terzo nel quale ancora una volta ricadono incidenti causati prevalentemente dalla componente utente. L'ultimo gruppo è invece costituito da incidenti "atipici", rispetto a tutti gli altri presenti nel database, e che anche in seguito alle elaborazioni di secondo livello ricadono nel gruppo degli extragrades. Le caratteristiche dei prototipi dei nuovi gruppi sono indicate in tabella 2 attraverso i livelli che le variabili utilizzate nelle analisi attingono in corrispondenza degli incidenti prototipo di ciascun gruppo.



**Figura 4 – ripartizione preliminare dello spazio campionario**

La ripartizione in questo caso non è però binaria ma fuzzy, di conseguenza l'appartenenza degli incidenti ai gruppi identificati non è totale ma parziale. In definitiva non è possibile attribuire univocamente ciascun incidente ad uno solo dei gruppi se non introducendo una soglia, ossia definendo un valore del grado di appartenenza di ogni incidente ad un gruppo oltre il quale questo sia attribuibile in

maniera prevalente al gruppo stesso. Nel caso in esame alla soglia è stato attribuito il valore 0,75 ottenendo la ripartizione del database originario indicata in Tabella 2.

VARIABILI	INCIDENTE TIPO A “UTENTE 1”	INCIDENTE TIPO B “UTENTE 2”
Luminosità	pieno giorno	Alba / tramonto
Intensità	normale	intenso
Traffico giornaliero	40404 veicoli/giorno	37838 veicoli/giorno
Localizzazione	corsia di marcia	Corsia di sorpasso
Curvatura	rettilineo	rettilineo
Pendenza longitudinale	discesa	salita
Limite di velocità	90 km/h	90 km/h
V <sub>85</sub>	126 km/h	113 km/h
Aderenza	soddisfacente	soddisfacente
Fondo stradale	asciutto	asciutto
Condizioni del tempo	sereno	sereno
Tipo di veicolo	autovettura	autovettura
Veicolo	nessun'avaria	nessun'avaria
Tipo di utente	abituale	abituale

**Tabella 1 – caratterizzazione dei gruppi dall’elaborazione di primo livello con HkM**

GRUPPO	PERCENTUALE DEL CAMPIONE
Utente	72 %
Ambiente	6 %
Veicolo	7%
Extragrades	6 %
Utente/Ambiente	2.4 %
Utente/ Veicolo	0.7 %
Veicolo/Ambiente	1.7 %
Utente/Veicolo/Ambiente	2.7 %
Utente/Veicolo/Ambiente/Extragrades	1.5 %

**Tabella 2 - ripartizione finale dello spazio campionario**

VARIABILI	INCIDENTE TIPO C “VEICOLO”	INCIDENTE TIPO D “UTENTE 3”	INCIDENTE TIPO E “AMBIENTE”
Luminosità	pieno giorno	pieno giorno	notte
Intensità	normale	normale	scarso
Traffico giornaliero	38857 veicoli/giorno	37778 veicoli/giorno	40825 veicoli/giorno
Localizzazione	corsia marcia	svincolo	stazione entrata
Curvatura	rettilineo	rettilineo	curva
Pendenza longitudinale	discesa	salita	pianura
Limite di velocità	60 km/h	40 km/h	90 km/h
V <sub>85</sub>	127.8 km/h	127.8 km/h	117.4 km/h
Aderenza	discreta	soddisfacente	soddisfacenti
Fondo stradale	asciutto	asciutto	bagnato
Condizioni del tempo	sereno/nuvoloso	sereno/nuvoloso	nuvoloso/pioggia
Tipo di veicolo	Auto infer. 1000 cc	Auto sup.1000 cc	Auto sup.1000 cc
Veicolo	Avaria veicolo	Nessun'avaria	Nessun'avaria
Tipo di utente	abituale	Non abituale	abituale

**Tabella 3 – caratterizzazione dei gruppi dall’elaborazione di secondo livello con FkM**

Il metodo FkM con extragrades riconosce le componenti Veicolo ed Ambiente (vedi Tabella 3). Gli incidenti provocati in prevalenza dalla componente Veicolo (7 %) sono eventi causati da avarie dello sterzo, delle sospensioni o dei freni, dall’esplosione dei pneumatici, da perdite del carico trasportato, dalla rottura dei ganci di traino e da altri guasti occorsi ai mezzi coinvolti nel sinistro.

Gli incidenti causati in prevalenza dalla componente Ambiente (6 %) coinvolgono in genere veicoli in buone condizioni e guidati utenti abituali. Gli eventi si verificano in condizioni meteorologiche non favorevoli (pioggia o altro) e frequentemente in orari notturni. Gli incidenti sono localizzati lungo tronchi curvilinei con raggi di curvatura

oscillanti tra i 450 ed i 500 metri, in discesa e su fondo stradale bagnato. La localizzazione prevalente degli incidenti del gruppo Ambiente, in termini di progressive chilometriche, è in carreggiata nord (in direzione Napoli) nella tratta Angri-Torre del Greco, dal km 25 al km 45, entro il quale sono ubicati gli svincoli di Pompei e Castellamare di Stabia.

La componente Ambiente incide inoltre in misura significativa, unitamente ad una o due delle altre componenti, su un'ulteriore aliquota di incidenti (8,3 %) verificatisi lungo l'autostrada oggetto dello studio. Una frazione di questi (1,3 % del totale), attribuibili equamente al gruppo Ambiente ed al gruppo Utente, sono presumibilmente causati da difetti di percezione del tracciato ed in generale da difficoltà di varia natura che insorgono nella lettura dello spazio stradale da parte dell'utenza. In definitiva all'ambiente stradale è riconducibile in misura sostanziale il 14,3 % degli incidenti verificatisi lungo la A3 nel corso del periodo di studio, il risultato è estremamente indicativo soprattutto se si tiene conto del fatto che i livelli sicurezza offerti da un'autostrada sono certamente più elevati di quelli che caratterizzano tutte le altre infrastrutture stradali. Una conferma di ciò si evince dalla lettura dei risultati conseguiti in precedenza a seguito dell'applicazione dei metodi di analisi dei gruppi a database di incidenti rilevati lungo alcune strade statali e riportati in uno studio precedente [4].

## **BIBLIOGRAFIA**

- [1] Bezdek, J.C., 1981, Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York.
- [2] de Gruijter, J.J., McBratney, A.B., 1988, A modified fuzzy k means for predictive classification. In: Bock,H.H.(ed) Classification and Related Methods of Data Analysis. pp. 97-104. Elsevier Science, Amsterdam.
- [3] Dell'Acqua G., Lamberti R., De Guglielmo M. L., 2001, Highway Accidents Analysis Using Fuzzy Cluster Algorithms, Proceedings VIII° Congress SIGEF "New Logics for the New Economy", Napoli;
- [4] Dell'Acqua G., 2000, Modelli e metodi in logica fuzzy di classificazione degli incidenti stradali e scelta degli interventi di ripristino delle condizioni di sicurezza, Tesi di Dottorato in Infrastrutture Viarie e Sistemi di Trasporto, Università degli Studi di Napoli "Federico II", Italy.
- [5] McBratney, A.B., Moore, A.W., 1985, Application of fuzzy sets to climatic classification. Agricultural and Forest Meteorology 35, 165-185.
- [6] McBratney, A.B., de Gruijter, J.J., 1992, A continuum approach to soil classification by modified fuzzy k-means with extragrades. Journal of Soil Science 43, 159-175.
- [7] Pal, N.R., Bezdek, J.C., 1995, On cluster validity for the fuzzy c-means model. IEEE Transactions on Fuzzy Systems 3, 370-379.
- [8] Roubens, M., 1982, Fuzzy clustering algorithms and their cluster validity. European Journal of Operational Research 10, 294-301.
- [9] Xie,X. L., Beni, G., 1991, A validity measure for fuzzy clustering. IEEE Transactions of Pattern Analysis and Machine Intelligence 13, 841-847.