

METODOLOGIE STATISTICHE E PROCEDURE INFORMATICHE PER L'ANALISI ESPLORATIVA DI DATI ARCHEOLOGICI A CONNOTAZIONE SPAZIALE

PREMESSA

La zona oggetto di studio è quella definita *Sabina Tiberina* ed è situata in un'area delimitata a nord dal fiume Nera, a sud dal fiume Aniene, a ovest dal fiume Tevere e a est dai monti Sabini. La superficie è di circa 700 km² e il suo territorio è caratterizzato da tracce notevoli di insediamenti di antichi popoli italici, che vanno dall'età del ferro al medioevo. Sebbene gli storici antichi abbiano lasciato nelle loro opere tutta una serie di notizie sull'origine e sull'espansione dei Sabini nell'Italia antica, alcune delle vicende di questi popoli meritano un ripensamento o quantomeno un approfondimento, anche a causa di recenti importanti acquisizioni.

In ragione di ciò, è stato sviluppato un GIS inserendo sul data base cartografico innanzitutto la cartografia di base della zona. Più precisamente sette tavolette in scala 1:25.000 dell'I.G.M. sono state digitalizzate in formato raster per riferimento topografico su video, mentre le infrastrutture e l'idrografia principale presenti nelle zone di interesse sono state digitalizzate per costituire un riferimento topografico nella stampa di mappe tematiche. Successivamente è stata inserita l'orografia numerica in scala 1:25.000 dell'I.G.M. e la copertura aerea del comune di Magliano Sabina, sempre riferita alla stessa scala. Inoltre è stata georeferenziata e inserita nel GIS un'immagine acquisita da piattaforma orbitante LANDSAT di tutta l'area della Sabina Tiberina, nonché la geologia regionale in scala 1:100.000. Sono stati infine inseriti nel Sistema Informativo dati geometrici e descrittivi relativi alla tavoletta I.G.M. di Passo Corese (*Ager Eretanus*) desunti da letteratura (OGILVIE 1965) ovvero, per quanto riguarda *Eretum*, sono stati studiati e classificati più di 100 siti tra sabini-arcaici e romani.

1. ASPETTI METODOLOGICI

Dopo una classificazione per epoca e per tipologia di sito, la prima operazione effettuata è stata quella di trasformare la distribuzione spaziale dei siti selezionati per epoca storica in pattern di punti distribuiti in uno spazio omogeneo (ESPA *et al.* 1995). Occorre precisare che in questo lavoro si è fatto riferimento ai dati relativi ai siti di epoca sabina desunti da letteratura (OGILVIE 1965) poiché, oltre all'interesse attuale degli archeologi per studi riguardanti i Sabini di epoca storica, essi costituiscono una raccolta ampia e ben organizzata di informazioni. Si è quindi proceduto all'analisi preliminare

della mappa di punti così ottenuta (Fig. 5) attraverso la funzione $K^{\wedge}(d)$ di Ripley (RIPLEY 1977) la quale assume la seguente espressione:

$$K^{\wedge}(d) = A \sum_{i=1}^N \sum_{j=1}^N K_{ij}(d) / N^2 \quad [1]$$

o, più precisamente, attraverso una sua trasformata (la funzione $L^{\wedge}(d)$ dovuta a Besag (BESAG 1977) ovvero:

$$L^{\wedge}(d) = [K^{\wedge}(d) / \pi]^{1/2} \quad [2]$$

In termini estremamente sintetici, i risultati ottenuti sono i seguenti. Analizzando la Fig. 4 si può osservare che:

1. il grafico della $L^{\wedge}(d)$ mostra una marcata tendenza alla clusterizzazione nell'intervallo $487 \text{ m} < d < 1625 \text{ m}$ con un livello di significatività pari al 5%;
2. $L^{\wedge}(d) = 0$ nell'intervallo $0 \leq d \leq 130$ significa che non si hanno siti in uno spazio di 130 m di raggio con centro in uno qualunque dei siti considerati (distanza dal vicino più prossimo);
3. $d = 487 \text{ m}$ è la soglia al di là della quale la clusterizzazione è statisticamente significativa;
4. $d = 1625 \text{ m}$ è la soglia di massima clusterizzazione.

Tutte queste informazioni sono estremamente importanti in quanto consentono, almeno sul piano teorico, il confronto tra varie configurazioni spaziali di siti di interesse storico. Esse costituiscono altresì la base per la costruzione e la verifica di modelli previsivi.

Utilizzando infine un'altra trasformata della funzione $K(d)$ di Ripley ovvero la $L_i^{\wedge}(d)$ (GETIS, FRANKLIN 1987)

$$L_i^{\wedge}(d) = \left[A \sum_{j=1}^n K_{ij} / \pi (n-1) \right]^{1/2} \quad [3]$$

si sono ottenute delle rappresentazioni ad isolinee e in falsi colori che evidenziano zone di intensità di sito variabile (Figg. 7, 8 e Tav. 1). Informazioni di questa natura rivestono un'importanza fondamentale in quanto consentono di quantificare fenomeni di clusterizzazione a differenti livelli di scala.

Fin qui si sono mostrati i risultati relativi all'utilizzo di una statistica funzionale (funzione $L^{\wedge}(d)$) che rappresenta, in un certo senso, la media delle relazioni che intercorrono fra le distanze tra i siti archeologici osservati. Ma per quantificare i fenomeni di attrazione e repulsione che interessano le configurazioni spaziali in esame, si può procedere costruendo la funzione $L_i^{\wedge}(d)$ per ogni singolo sito archeologico. In tal modo si può stabilire a quale livello di distanza un sito appartenga o meno a un cluster spaziale.

Lo strumento statistico adottato è, fra l'altro, di grande utilità qualora il pattern totale non si discosti dalla casualità completa per studiare fenomeni locali di inibizione e clusterizzazione che possono comunque essere presenti. In particolare, nel caso in esame non è stato affrontato il problema di condurre un'analisi del secondo ordine via $L_i^{\wedge}(d)$ per alcuni siti particolarmente significativi e rappresentativi. L'attenzione è stata però incentrata su una rappresentazione grafica di immediato riscontro pratico. Prendendo in considerazione la sola soglia $d = 487$ m (distanza alla quale, come detto al precedente punto 3., la clusterizzazione diviene statisticamente significativa), sono stati calcolati i valori $L_i^{\wedge}(d = 487)$, $i = 1, 2, \dots, 25$.

Una rappresentazione per isolinee di $L_i^{\wedge}(d = 487)$, $i = 1, 2, \dots, 25$, è riportata in Fig. 7; una lettura più semplice della Fig. 7 è consentita dall'esame di Tav. 1, che mostra una rappresentazione di $L^{\wedge}(d = 487)$ in falsi colori. Le porzioni di figura più chiare corrispondono alle zone interessate da più intensi fenomeni di clusterizzazione. La Fig. 7 riporta i siti oggetto di studio e le isolinee per le quali $L_i^{\wedge}(d)$ supera il valore atteso nell'ipotesi di casualità completa ($L_i^{\wedge}(d = 487) = 487$) (area indicata dal tratteggio) ed evidenzia in modo più preciso le zone di maggiore intensità di sito archeologico. In questo lavoro ci si è limitati, comunque, a presentare un esempio che costituisce una sorta di "fotografia istantanea scattata" a una soglia pari a 487 m.

2. IL PROGRAMMA DI ELABORAZIONE

Per la realizzazione di un programma di analisi statistica di dati a connotazione territoriale, si è scelto di utilizzare l'ambiente operativo MATLAB¹ per le sue caratteristiche di semplicità d'impiego, di potenza di calcolo e di possibilità di restituzioni grafiche. Infatti MATLAB non è un classico linguaggio simbolico di programmazione, ma è un ambiente matematico dove i dati numerici vengono manipolati tramite la scrittura di una sequenza di espressioni matematiche. Tali espressioni hanno una notazione semplice e molto simile a quella correntemente adottata nei testi di analisi matematica.

Nell'ambiente MATLAB, oggetti diversi come segnali, immagini, polinomi, serie temporali, dati statistici multivariati o sistemi lineari, vengono trattati come matrici su cui operano funzioni matematiche della libreria di MATLAB. L'applicazione scritta in questo ambiente si traduce in una serie di espressioni matematiche le quali possono essere modificate in seguito dall'utente per adattarle a nuove esigenze di elaborazione. Una volta realizzata una applicazione, questa diviene una delle funzioni presenti nelle librerie di MATLAB e quindi può essere inclusa in un'altra applicazione. Nelle librerie di base sono presenti numerose funzioni matematiche, scientifiche e di visualizzazione grafica dei dati, le quali consentono di realizzare programmi di

¹ COPYRIGHT The MathWorks, Inc., U.S.A.

elaborazione dati molto sofisticati e sempre aggiornabili. Lanciando il programma MATLAB, si genera una finestra dei comandi nella quale, da tastiera, si possono immettere sia dati che comandi per creare o modificare funzioni matematiche della libreria.

Nel caso proposto il programma di analisi statistica si avvia fornendo il comando *archo*. Questo, oltre ad avviare una serie di comandi di inizializzazione, produrrà una finestra dove verranno in seguito visualizzati i dati, secondo le necessità. Nella parte alta di questa finestra vi sono dei menu che consentono di eseguire delle operazioni utili all'analisi dei dati. Tali menu si attivano selezionandoli con il cursore.

Lo schema a blocchi del programma è riportato in Fig. 1; i vari passi di tale programma verranno illustrati e commentati nel seguente ordine:

- immissione dei dati;
- analisi dei dati: a) analisi di una mappa di punti tramite la funzione $K^{\wedge}(d)$;
b) analisi di una mappa di punti tramite la funzione $L_1^{\wedge}(d)$;
- visualizzazione e stampa dei risultati.

2.1 Modalità di input dei dati

I dati possono essere immessi nel seguente modo:

- da tastiera, con la routine INGRKEB.M;
- da disco, con la routine INGRESSO.M.

Nel primo caso, oltre a immettere i dati, si genererà un file dove salvare i dati nel formato compatibile alla routine INGRESSO.M. Con la seconda routine, il file che contiene i dati da analizzare viene selezionato tramite un menu simile a quello dell'apertura files del sistema operativo Windows 3.1. Un esempio è riportato in Fig. 2.

Dopo l'immissione dei dati, si deve selezionare il tipo di strumento statistico con cui si intendono analizzare i dati: a) $K^{\wedge}(d)$ di Ripley; b) $L_1^{\wedge}(d)$ di Getis - Franklin.

2.2 Analisi del secondo ordine di una mappa di punti (funzione $K^{\wedge}(d)$)

L'analisi di Ripley prende le mosse con la definizione del diametro massimo di analisi, il quale viene, in genere, posto pari a un quarto del lato maggiore dell'area in esame (supposta rettangolare) e dell'incremento di raggio durante il processo di analisi. Quest'ultimo valore è funzione del numero di cicli stabiliti all'avviamento dell'analisi (*ncycle*). A ogni ciclo dell'analisi di Ripley, dopo aver confrontato la distanza ro_{ij} tra il sito i ed il sito j con la distanza di analisi ds , si computa il valore di $kappa(i,j)$ con la routine CALKAPPA.M (Fig. 3), per il calcolo della funzione $K^{\wedge}(d)$ qui di seguito riportata, unitamente al relativo diagramma di flusso (Fig. 3).

Le distanze considerate nella funzione CALKAPPA sono distanze rela-

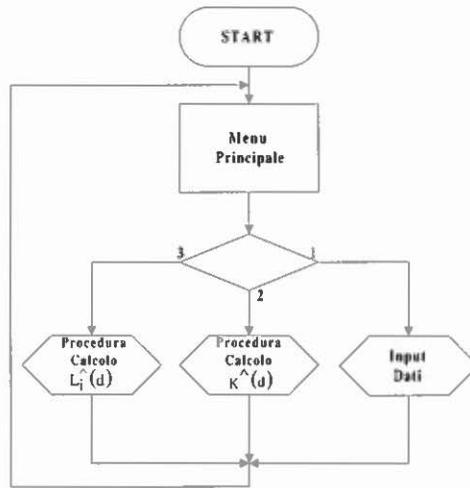


Fig. 1 – Diagramma a blocchi del programma Archeo.

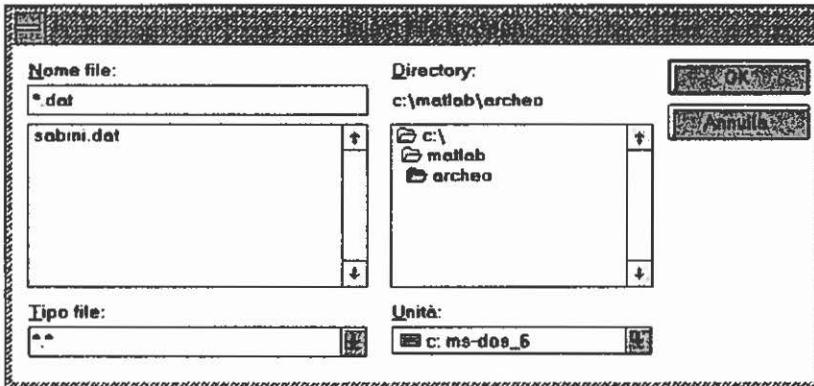


Fig. 2 – Finestra per la selezione dei dati in ingresso.

tive a un'area unitaria. La trasformazione delle coordinate reali dei punti in coordinate normalizzate a un'area di superficie unitaria avviene tramite la seguente routine **COORDREL.M**:

```

function [data2,latirel]=coordrel(data1,lati)
% riscalatura delle coordinate
%
coordmin=min(data1);
data2(:,1)=(data1(:,1)-coordmin(1))./(lati(1).*lati(2)).^0.5;
data2(:,2)=(data1(:,2)-coordmin(2)).*(lati(1).*lati(2)).^0.5;
areasq=sqrt(lati(1).*lati(2));
latirel=lati./areasq;
    
```

dove **lati** è un vettore che contiene la lunghezza e la larghezza dell'area di analisi, la quale deve includere i caratteri di tutti i siti da analizzare; **data1**, **data2** contengono le coordinate dei siti prima e dopo la riscalatura; in **latirel** sono memorizzati i lati dell'area unitaria di analisi. Nel caso in cui l'area sia un rettangolo, l'area unitaria sarà anch'essa un rettangolo con lo stesso rapporto di lunghezza dei lati.

Ovviamente la scelta dell'area di analisi influenza il valore del diametro massimo dell'analisi di Ripley. Qualora si definisca un'area di analisi inferiore a quella minima per il contenimento di tutti i siti immessi, il programma escluderà automaticamente i siti posti all'esterno di tale area. Come si può osservare dal listato della funzione **COORDREL.M**, sono sufficienti quattro espressioni matematiche invece di una serie di loop **FOR NEXT** per la trasformazione del sistema di coordinate.

Tramite la routine **SITIMAP.M** si ottiene una visione topografica dei siti su un diagramma bidimensionale, i cui assi sono i lati dell'area di analisi

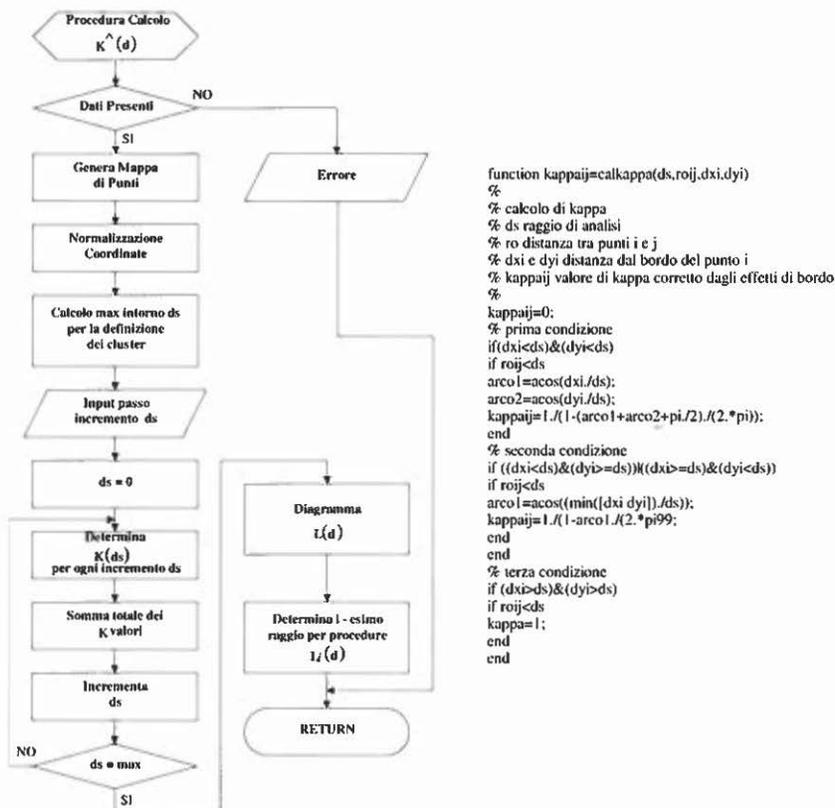


Fig. 3 - Diagramma di flusso della procedura **CALCAPPA** e relativo listato.

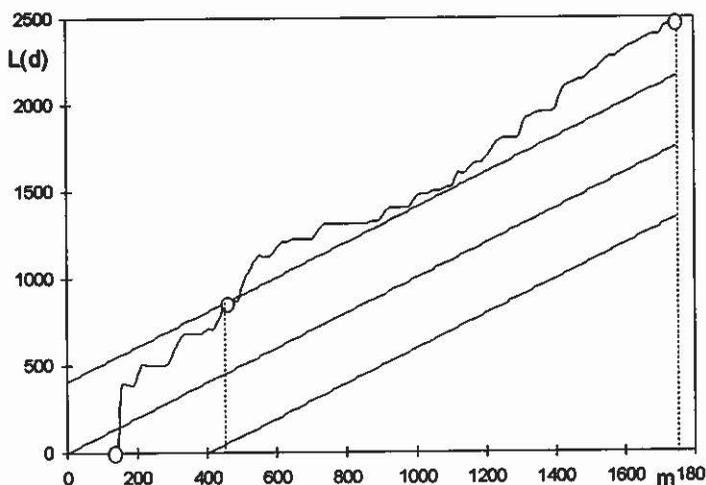


Fig. 4 - Siti di epoca sabina. Grafico della funzione $L^{\wedge}(d)$ calcolata per $N = 25$ siti.

(Fig. 5). Durante lo svolgimento dell'analisi di Ripley viene prodotta una seconda finestra in cui, sulla parte alta, si disporrà di una rappresentazione topografica dei siti, mentre nella parte inferiore, di una serie di rappresentazioni spaziali della funzione $L^{\wedge}(d)$ al variare del raggio di analisi. Alla fine di ogni ciclo si genera un diagramma che fornisce una rappresentazione dell'andamento spaziale della funzione $L^{\wedge}(d)$ per il raggio ds .

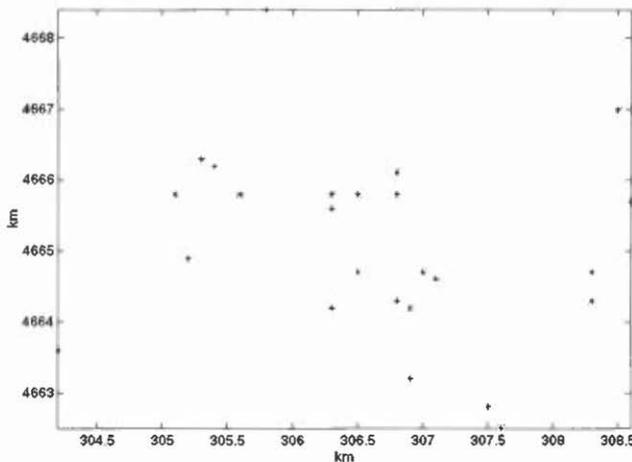
2.3 Analisi di una mappa di punti tramite la funzione $L_i^{\wedge}(d)$

L'analisi condotta con la trasformata di Getis e Franklin, prima descritta, procede in modo analogo a quella di Ripley tranne per il fatto che soltanto per alcuni raggi prestabiliti viene calcolata la funzione e generata una rappresentazione topografica della stessa (Fig. 6). Questa routine si avvale delle medesime funzioni della routine CALKAPPA precedentemente illustrate.

2.4 Stampa della visualizzazione dei dati

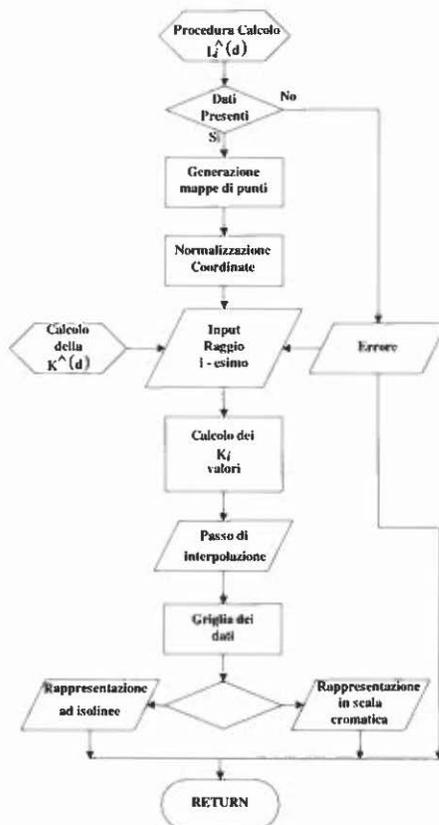
Nel menu EDIT delle figure, è presente il comando *copy* il quale consente di copiare la figura nel clipboard (formati BitMap, Metafile, PostScript e PCX 256 colori) e fornirlo così ad altre applicazioni Windows come, ad esempio, un programma di scrittura testi. In ogni caso, essendo quelli disponibili alcuni tra i formati grafici più diffusi, è possibile esportarli, con opportuni tools, in altri formati grafici.

A conclusione dell'analisi di Ripley condotta con la funzione $K^{\wedge}(d)$ si generano due nuove figure che rappresentano la distribuzione spaziale 3-D della funzione $L_i^{\wedge}(d)$ al diametro massimo di analisi e lo sviluppo della fun-



```
function sitimap(data1)
% mappa della distribuzione dei
%
% sitimap(data1)
%
x=data1(:,1);
y=data1(:,2);
plot(x,y,'*g');
```

Fig. 5 – Mappa dei siti sabini del territorio di *Eretum* (N =25) e routine di generazione di mappe.



```
function cntgetis(data2,l,latirel,entopt)
%
% produce una mappa topografica della funzione L1^d(d)
%
% cntopt(1)=numero di punti nel grigliato n x n
% cntopt(2)=numero di livelli
%
if nargin<4, cntopt=[25 10]; end
ti=0:(1./cntopt(1))-latirel(1);
tj=0:(1./cntopt(1))-latirel(2);
[XXI,YYI]=meshgrid(ti,tj);
ZZI=griddata(data2(:,1),data2(:,2),l,XXI,YYI);
contour(XXI,YYI,ZZI,cntopt(2));
```

Fig. 6 – Diagramma di flusso e routine di calcolo della procedura CNT GETIS.

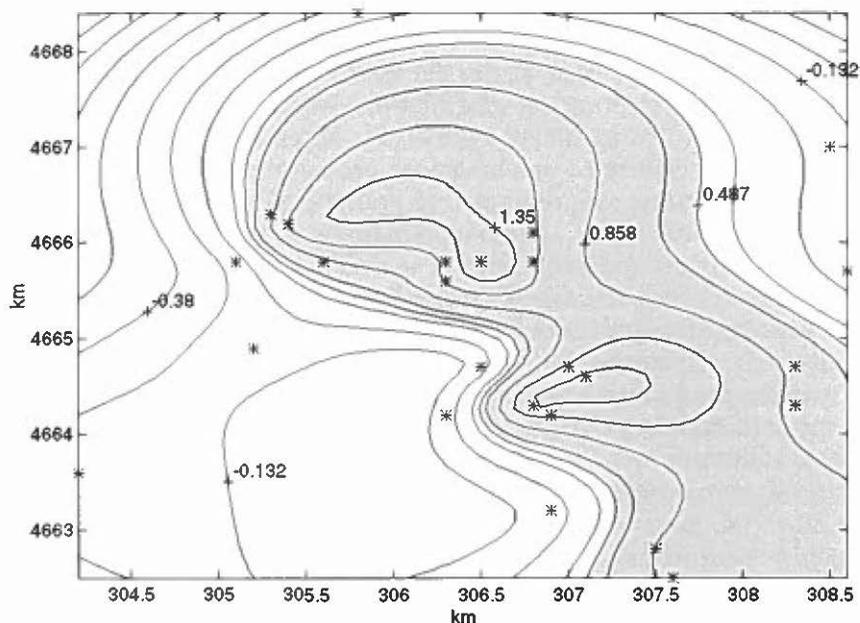
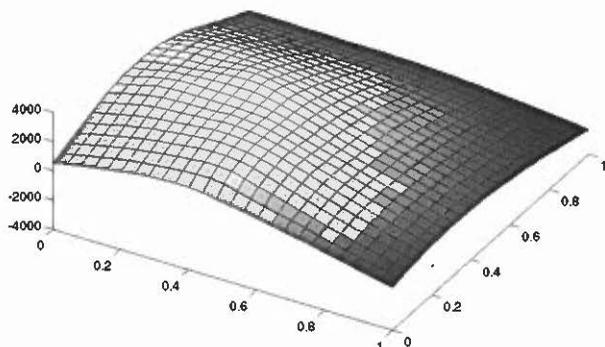


Fig. 7 – Rappresentazione della $L_i^{\wedge}(d)$ ad isolinee di densità per la zona d'indagine.



```
function surfzgs(data2,1,latirel,srfopt)
%
% produce una superficie 3d della funzione di Geti
%
% srfopt=numero di punti nel grigliato n x
%
if nargin<4, srfopt=25;end
ti=0:(1./srfopt(1)):latirel(1);
tj=0:(1./srfopt(1)):latirel(2);
[XXI,YYI]=meshgrid(ti,tj);
ZZI=griddata(data2(:,1),data2(:,2),1,XXI,YYI);
surf(XXI,YYI,ZZI);
```

Fig. 8 – Diagramma in 3-D della funzione $L_i^{\wedge}(d)$ e routine di generazione di grafici.

zione di Ripley con il raggio per il test di casualità. In questo diagramma sono riportate due rette che definiscono i limiti di casualità dei dati secondo la seguente espressione:

$$L^{\wedge}(d) = d \pm \frac{1.45}{N}$$

Attivando il menu POINT nella finestra dove è diagrammato l'andamento della funzione $L^{\wedge}(d)$ (Fig. 4), si possono estrarre le coordinate di punti significativi per l'analisi mediante la funzione $L_i^{\wedge}(d)$, marcandoli con il cursore.

3. CONSIDERAZIONI CONCLUSIVE

Sorvolando su considerazioni di carattere storico e metodologico circa i risultati dell'elaborazione, ciò che si vuole evidenziare con questo lavoro è che, mediante il software messo a punto, si è ottenuta una notevole semplificazione delle procedure ed una maggiore precisione relativa ai calcoli numerici. Inoltre, le ottime rappresentazioni grafiche di immediato riscontro pratico, oltre a costituire nuovi livelli informativi per il GIS, consentono l'impostazione di ulteriori approfondimenti analitici.

È qui appena il caso di ricordare che il software sviluppato è autoconsistente nei limiti dei metodi impiegati. Tuttavia esso costituisce un insieme aperto a nuovi sviluppi, peraltro già in corso, ed è in ogni modo a disposizione per tutti coloro i quali intendessero affrontare i problemi dei beni culturali con il tipo di approccio descritto nel presente lavoro e con gli strumenti adottati.

GIUSEPPE ESPA

Università degli Studi di Trento
Istituto di Statistica e Ricerca Operativa

SALVATORE ESPA, ROBERTO GABRIELLI, UGO RICCI
C.N.R. - Istituto per le Tecnologie Applicate ai Beni Culturali

GUIDO RIGHINI

C.N.R. - Area della Ricerca di Roma - Montelibretti

BIBLIOGRAFIA

- BESAG J. 1977, *Contribution to the discussion of Dr. Ripley's paper*, «Journal of the Royal Statistical Society», B, 39, 193-195.
- ESPA G., ESPA S., GABRIELLI R., PIFFERI A., RICCI U. 1995, *Impiego integrato di metodi statistico-spaziali e tecniche informatiche nell'analisi di aree di interesse storico-archeologico*, «Archeologia e Calcolatori», 6, 35-60.
- GETIS A., FRANKLIN J. 1986, *Second-order neighborhood analysis of mapped point patterns*, «Ecology», 65, 3, 473-477.
- OGILVIE R.M. 1965, *Eretum*, «Papers of the British School at Rome», 32, 70-112.
- RIPLEY B.D. 1977, *Modelling spatial pattern (with discussion)*, «Journal of the Royal Statistical Society», B, 39, 172-212.

ABSTRACT

Recent studies stressed the usefulness of a linked employment, in historical and archaeological research works, of the most advanced computer technologies with statistical methods. In this work, as an application of a Geographical Information System (GIS), we present a Matlab language program for the statistical analysis of spatial data related to ancient human settlements. The program explanation is preceded by some notes about the GIS, the zone from which data have been collected, and the statistical methods applied to analyse data reported as examples.