

Il consolidarsi della posizione dell'Europa come meta di cospicui flussi di immigrazione internazionale provenienti dal Sud, ma anche da un Est sempre più lontano, ha avuto e continua ad esercitare un forte condizionamento sul ruolo dell'Italia nel contesto di tali processi. Proprio per la sua conformazione fisica e per la sua specifica collocazione nel bacino mediterraneo e in uno spazio di congiunzione di diverse vie di comunicazione – elementi squisitamente geografici, quindi! – si può affermare che lo Stivale rappresenti sempre più una delle porte d'accesso “privilegiate” verso il Vecchio Continente.

Col passare del tempo, d'altronde, l'Italia ha integrato la sua naturale vocazione geografica come terra di transito, divenendo teatro di un'immigrazione non più temporanea e provvisoria, bensì stanziale e prolungata. Tale evoluzione ha chiamato in causa nuovi problemi e fenomeni di trasformazione territoriale molto interessanti, che, tra l'altro, coinvolgono, con dinamiche e soluzioni diverse, sia il mondo rurale che quello strettamente urbano.

Tali processi sono leggibili, innanzi tutto, nelle trame che emergono da un primo vaglio critico del dato quantitativo e del suo divenire. In tal senso i contributi contenuti nella I sezione di questo volume ci permettono di apprezzare appieno il valore conoscitivo e descrittivo di tale tipo di indagine, che si pone, appunto, come attenta e dettagliata lettura della situazione attuale. Questi studi rappresentano, infatti, con adeguata chiarezza, quelle che sono le caratteristiche principali dei diversi modelli di flussi di immigrati e soprattutto della loro distribuzione spaziale nonché dei loro

modelli di insediamento ed organizzazione sul territorio.

Bisogna osservare, poi, come il processo di insediamento di diverse comunità sia ormai abbastanza avanzato e, per certi versi, anche maturo. La stabilizzazione degli immigrati, che passa per lo più attraverso la regolarizzazione della posizione occupazionale, il ricongiungimento familiare o comunque la formazione di un nucleo familiare e/o abitativo non più precario, si esprime, poi, attraverso forme di vita e di relazione con l'ambiente sociale circostante sempre più complesse e “dense”.

È in questo momento che, a fronte di tale processo di complessificazione, si avverte maggiormente l'esigenza di ulteriori approfondimenti e studi capaci di cogliere i mutamenti strutturali e qualitativi in atto nelle diverse comunità immigrate. Questa esigenza di acquisizione di una maggiore e differente conoscenza sui fenomeni legati all'immigrazione è direttamente connessa con/e strumentale all'implementazione di misure e pratiche politiche generali, ma anche e soprattutto operative miranti, a favorire il più possibile la riduzione di conflitti sociali e il reperimento di forme di convivenza pacifica al di dentro delle diverse anime che oramai coesistono nel nostro Paese.

È, quindi, in questo ambito che si avverte anche forte il bisogno di elaborare nuovi strumenti di studio ed analisi, adeguati alle mutate ed eterogenee condizioni che è possibile identificare nelle differenti realtà geografiche italiane. In realtà, la continua attenzione all'innovazione metodologica ed all'affinamento dei mezzi di indagine a nostra disposizione è un'aspirazione che deve e, di

fatto, abbraccia anche il momento della prima e preliminare indagine conoscitiva sul campo.

I contributi delle sezione che segue ci offrono

un buon esempio degli sforzi compiuti in tal senso da alcuni studiosi della nostra disciplina, impegnati su fronti e situazioni molto eterogenee.

*(F. K.)*



## Metodi di analisi e visualizzazione di fenomeni immigratori \*

### 1. Introduzione

Nel presente contributo l'attenzione viene posta su alcune delle caratteristiche relative alla popolazione cinese a livello urbano. In particolare si focalizza l'attenzione sulla distribuzione spaziale degli immigrati cinesi all'interno del Comune di Trieste, in base alle registrazioni anagrafiche. I dati relativi ai cinesi residenti a Trieste non presentano cifre particolarmente elevate in assoluto (555 iscritti al 31/12/2003), soprattutto se confrontate con la popolazione complessiva residente nel Comune (200.000 abitanti circa); essi rappresentano tuttavia il quarto gruppo etnico sul territorio, dopo Jugoslavi, Croati e Albanesi. I cinesi quindi risultano interessanti in quanto primo gruppo etnico presente in misura rilevante subito dopo quelli storicamente legati alla città di Trieste, quali sono le etnie provenienti dalla ex Jugoslavia o dall'area balcanica più in generale. La velocità d'insediamento cinese in Italia ha caratterizzato soprattutto gli anni a cavallo tra la fine del XX e l'inizio del XXI secolo. All'inizio degli anni '90, infatti, i numeri relativi ai residenti cinesi trasmettono l'immagine di un gruppo alquanto ridotto. Negli ultimi anni la presenza è visibile e in sensibile aumento di anno in anno. Come ricorda Romei (2004) generalmente la scelta localizzativa prediletta riguarda soprattutto l'ambito urbano di città capoluogo di provincia, soprattutto nei centri storici, o nell'immediata periferia. La preferenza per uno stile di vita urbano è legata alla consuetudine, allo svolgimento di attività commerciali e la tendenza all'agglomerazione spaziale. Per quanto riguarda le scelte occupazionali, commercio e ri-

storazione risultano tra le principali attività svolte dalla comunità cinese in Italia. Trieste non fa eccezione relativamente alle caratteristiche principali dell'immigrazione cinese. Ciò vale soprattutto per quanto concerne la predilezione di attività commerciali e la concentrazione spaziale nei centri storici.

Nell'affrontare i fenomeni legati alla popolazione s'incontrano delle difficoltà legate agli ambiti territoriali cui i dati sono riferiti. Solitamente questi dati sono aggregati a livello di unità amministrative, quali province e comuni, o sezioni di censimento. Questo tipo di aggregazioni presenta tuttavia una serie di problemi sia per quanto riguarda l'analisi che per la visualizzazione dei dati. Ciò deriva dalle diverse forme ed estensioni territoriali che le stesse unità amministrative o statistiche hanno; tali caratteristiche possono influenzare i risultati dell'analisi e portare a conclusioni viziate.

Una delle parziali soluzioni consiste nel rappresentare la popolazione per mezzo di una superficie, nella quale i dati sono elaborati e presentati come un campo continuo, non dipendente dal tipo di partizione dello spazio arbitraria per mezzo di unità di tipo areale. La considerazione di fondo riporta al considerare la densità di popolazione come una funzione continua, anche se la popolazione stessa è un fenomeno discreto.

Nel presente lavoro, dopo aver considerato alcuni aspetti relativi alla geografia della popolazione nell'analisi spaziale, si sono presentati due approcci relativi alla rappresentazione della popolazione come distribuzione di punti nello spazio. In particolare si sono considerate come metodolo-

gie, il KDE (*Kernel Density Estimation*) per la stima della densità e due algoritmi di simulazione di punti su superfici URDA (*Uniform Random Distribution Algorithm*) e NURDA (*Non Uniform Random Distribution Algorithm*).

## 2. Alcuni aspetti relativi alla Geografia della popolazione e all'Analisi Spaziale

### 2.1. La popolazione come dato spaziale

Uno dei limiti cui ci si trova davanti quando si esaminano i fenomeni riguardanti la popolazione riguarda la disponibilità dei dati e il loro riferimento spaziale.

Nell'ambito della scienza dell'informazione geografica si tende oggi a distinguere tra due grandi visioni della realtà geografica: *oggetti e campi* (Goodchild, 1992). Tale prospettiva non è legata al tipo d'immagazzinamento delle informazioni, che tradizionalmente vede una distinzione tra dati in formato raster e vettoriale, preferendo adottare una visione discreta (*object view*) ed una continua (*field view*) dello spazio. Nel primo caso gli elementi geografici sono astratti per mezzo di elementi discreti, semplificati per mezzo di punti, linee, aree (o poligoni). All'interno di questa visione la popolazione può essere rappresentata per aree (comuni o altre unità amministrative, sezioni di censimento, codici di avviamento postale), o alternativamente per punti (indirizzi basati su numeri civici, centroidi di unità areali o coppie di coordinate più in generale). Nel secondo caso non si ricorre a suddivisioni nette, discrete della realtà geografica, ma la si considera come un continuo, superfici che variano a seconda dell'intensità dei valori che vengono pesati nelle singole unità minime. Questa visione è consistente con una partizione dello spazio in unità minime omogenee, quali possono essere le celle (quadrate, esagonali, ecc.) di una griglia fine sovrapposta alla regione studiata. La *field view* viene quindi definita utilizzando i due concetti chiave di "continuità nello spazio" e "auto definizione": in un campo c'è un valore in ogni luogo (al limite lo zero) ed insiemi di valori presi assieme definiscono un campo (O'Sullivan e Unwin, 2003).

Diverse metodologie sono state utilizzate nel corso degli ultimi anni per ridurre il problema legato allo studio di fenomeni relativi a suddivisioni di tipo areale e rappresentare nel modo più efficace possibile la popolazione distribuita su di un determinato territorio. Ciò è avvenuto parallelamente alla diffusione della cartografia automati-

ca, dei GIS e all'aumento delle potenze di calcolo dei calcolatori, nonché agli sviluppi nel campo della statistica spaziale. Tra i diversi autori, Martin (1989) propone di attribuire il dato relativo alla popolazione ai centroidi (cioè ai baricentri dei poligoni) delle sezioni di censimento, che diventano così coppie di coordinate pesate in base al numero di individui, e successivamente ottenere una superficie di densità, tenendo conto della distribuzione della popolazione intorno al centroide secondo una certa funzione di distanza. Langford e Unwin (1994) propongono da un lato di utilizzare immagini satellitari per una distinzione tra spazio 'abitato' e 'non abitato', al fine di ridurre la superficie di una regione cui attribuire la popolazione, e dall'altro lato riaggregare i dati così ottenuti per mezzo di una griglia di celle e procedere alla loro interpolazione. Queste basi sono state portate avanti e affinate, con il denominatore comune costituito sia dalla delimitazione più corretta delle aree effettivamente occupate od occupabili dalla popolazione, sia dalla sua rappresentazione la più 'realistica' possibile. In seguito Dykes e Unwin (1998) riflettono sulla difficoltà e sui problemi legati alla rappresentazione areale dei dati di popolazione, con proposte di utilizzo di strumenti alternativi, quali cartogrammi basati su carte dasimetriche (*dasymeric mapping*), che utilizzano metodologie di interpolazione areale con l'utilizzo di dati ancillari esterni, mentre Mesev e Longley (2000) e Mennis (2003) sfruttano l'effetto combinato dell'elaborazione dell'immagine satellitare per ottenere delle classificazioni dell'uso del suolo tali da garantire l'individuazione dell'ambiente costruito 'residenziale' e dei GIS per ridistribuire i dati demografici sul suolo occupato e ottenere superfici di densità.

L'utilizzo di dati puntuali, basati quindi su numeri civici o, più in generale, coppie di coordinate aiuta a limitare il problema connesso ai dati 'spalmati' su raggruppamenti areali.

La visualizzazione puntuale di un fenomeno può fornire delle prime indicazioni relative alle sue caratteristiche. Può tuttavia risultare fuorviante se utilizzata da sola per la rappresentazione del fenomeno, mentre risulta ancora difficoltosa, anche all'interno dei GIS, la gestione di dati multi-attributo: ad esempio, a un solo numero civico (punto sulla carta) corrispondono più residenti di sesso, età, razza, condizioni economiche diverse. Tali caratteristiche sono difficilmente supportate all'interno di pacchetti GIS standard e si rendono necessarie metodologie avanzate di analisi spaziale.



L'analisi statistica spaziale, o *spatial statistical analysis*, è una disciplina che impiega metodi statistici applicati a dati spaziali al fine di determinare dei modelli utilizzabili per fare previsioni (O'Sullivan e Unwin, 2003). Definire in maniera precisa questa materia è difficile a causa del gran numero di campi in cui essa viene impiegata. Di fatto la Statistica Spaziale copre ogni ramo dell'analisi dei dati che si occupa di dati spaziali. I dati possono essere, ad esempio, posizioni d'alberi in una foresta o indirizzi di persone che risiedono in una determinata regione (Rowlingson, 2003).

Una volta chiarito il tipo di rappresentazione dei dati spaziali si pone il problema di analisi statistica degli stessi. Mentre nell'analisi statistica classica si descrive la distribuzione dei valori di dati osservabili attraverso misure descrittive quali la media, la varianza e così via, nell'analisi statistica spaziale si è interessati a studiare anche la distribuzione nello spazio dei dati stessi. Questa distribuzione spaziale viene descritta attraverso relazioni tra entità, le quali, essendo collocate nello spazio, coinvolgono concetti quali *distanza*, *adiacenza*, *interazione* o *vicinanza*.

La *distanza* usata in questo campo nella maggior parte dei casi è quella euclidea; esistono però altri tipi di distanze quali quelle di Manhattan, Mahalanobis, a minima varianza, etc. che vengono utilizzate qualora il problema in analisi lo richieda. L'*adiacenza* può essere pensata come l'equivalente nominale o binario della distanza: due entità spaziali sono adiacenti oppure no. Una semplice formulazione può essere quella di decidere che due entità sono adiacenti se la loro distanza è minore di un certo valore prefissato; oppure si può definire adiacente il "più prossimo tra i vicini" (*nearest neighbour*). L'*interazione* (*interaction*) è una combinazione fra le nozioni di distanza e adiacenza tra due entità spaziali: essa è definita per ogni entità come un numero che varia tra 0 e 1 dove 0 sta a significare che non c'è alcuna interazione, mentre 1 indica il grado massimo di interazione. Attraverso queste nozioni si arriva a quella di *vicinanza*, che può essere definita ricorrendo alla distanza, all'adiacenza o all'interazione.

Questi concetti, una volta applicati all'insieme di dati, possono venire convenientemente rappresentati sotto forma di matrice: se si considera infatti un insieme di entità spaziali e si calcola la distanza o l'adiacenza di ognuno di essi da tutti gli altri, si ottengono dei valori che formano una matrice quadrata, la cui diagonale principale è composta

da zeri, simmetrica e che comprende tutte le informazioni riguardo la misura considerata sull'insieme di entità.

Un'altra operazione molto generale usata nello specificare proprietà spaziali di un insieme di oggetti è la partizione dello spazio, ossia la suddivisione esaustiva della regione di studio in zone di area minore. Quella maggiormente usata è la partizione in *poligoni di prossimità*, conosciuti anche come *poligoni di Thiessen* o *Voronoi*.

Sia  $S$  un insieme di punti del piano. Per ciascun punto  $P \in S$  si definisce come suo poligono di Voronoi l'insieme di tutti i punti del piano più vicini a  $P$  rispetto a qualsiasi altro punto dell'insieme  $S$ . L'insieme di tutti i poligoni di Voronoi dell'insieme di punti  $S$  prende il nome di *diagramma di Voronoi di  $S$* . Per dati che rappresentano oggetti puntuali, i lati dei poligoni si costruiscono tracciando gli assi dei segmenti che uniscono coppie di punti, mentre la costruzione dei poligoni per oggetti di tipo lineare o areale è più complessa. Una costruzione simile può essere pensata anche nelle tre dimensioni, in questo caso non ci saranno poligoni risultanti ma zone di forma simile a bolle irregolari (la forma dipenderà dal tipo di distanza definita). In ogni caso i poligoni riempiranno l'intera regione di studio (non ci saranno cioè pezzi di regione non ricoperti da un poligono). Congiungendo i punti  $P \in S$  tra loro se e solo se i loro poligoni di Voronoi hanno un lato in comune, si ottiene una nuova partizione dello spazio in poligoni questa volta di forma triangolare, chiamata *triangolazione di Delaunay*. La Figura 1 riporta un esempio di entrambe le partizioni. Mentre una triangolazione di un insieme di punti  $S$  è definita come una partizione del più piccolo insieme convesso contenente  $S$  (*convex hull*), la triangolazione di Delaunay è quella in cui non ci sono altri punti di  $S$  all'interno di ogni cerchio circoscritto ai suoi triangoli.

Passando ad analizzare i principali problemi collegati all'analisi statistica di dati spaziali si deve tener presente innanzitutto il problema dell'auto-correlazione spaziale, esso sta ad indicare che "...è più probabile che dati provenienti da posizioni vicine nello spazio siano simili tra loro rispetto a dati che provengono da posizioni lontane." (O'Sullivan e Unwin, 2003). Questo comportamento si verifica normalmente nei dati spaziali; una posizione di altitudine elevata sarà probabilmente vicina ad un'altra di altitudine di poco differente, e lontana da una di altitudine molto inferiore (si vedano ad esempio dati di territori che hanno al loro interno montagne e valli). Molti fenomeni geografici possono essere descritti in

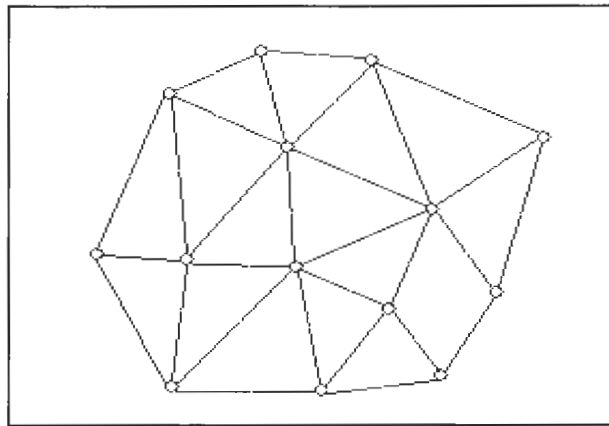


Fig. 1. Esempio di poligoni di Voronoi (in grigio) e triangoli di Delaunay (in nero).

questi termini: le città sono concentrazioni di individui, le tempeste sono concentrazioni locali di particolari condizioni atmosferiche, e così via. Purtroppo ciò impedisce di applicare alcuni strumenti della statistica convenzionale (Bao, 2004).

Il problema dell'autocorrelazione spaziale non è l'unico che s'incontra; un'altra complicazione che si ritrova spesso in questa fase deriva dal fatto che la maggior parte dei dati sono in realtà aggregati. Il problema è che le unità di aggregazione usate sono arbitrarie, non sono definite considerando le caratteristiche del fenomeno in analisi e ciò porta ad errori nel calcolo di statistiche o stime dei parametri. Questo fatto, noto come Problema dell'Unità Areale Modificabile (*Modifiable areal unit problem* (MAUP)), non è rilevante solo a livello teorico ma anche dal punto di vista pratico. Si tratta della situazione in cui variando i confini o la scala a cui sono aggregati i dati l'analisi degli stessi ne viene influenzata. Di conseguenza spesso non è chiaro se i risultati dell'analisi riguardano effettivamente la realtà degli individui che vivono in una determinata regione o se siano piuttosto una funzione del particolare tipo di suddivisione areale utilizzata nell'analisi. Analogamente, nella visualizzazione dei dati demografici associati a unità areali si corre il rischio di considerare la popolazione come omogeneamente distribuita all'interno di ogni unità areale, anche nel caso in cui parti della regione considerata siano di fatto non abitate (Dorling, 1994).

Un altro tipo di errore strettamente collegato al MAUP si presenta quando si osserva una relazione tra dati aggregati, e la stessa relazione viene estesa ad un livello di aggregazione superiore senza ulteriori indagini. Questo tipo di errore è molto più diffuso di quanto si possa pensare; si ritrova specialmente nelle notizie di ogni giorno divulgate

dai canali d'informazione, che cercano la spiegazione più semplice per alcuni fenomeni i quali invece hanno ragioni ben più complesse di quanto possa apparire a prima vista.

Sempre in questo contesto si possono collocare i problemi inerenti alla scala geografica; molto spesso s'individua una relazione tra i dati se si utilizza una certa scala, ma questa relazione scompare non appena se ne utilizza un'altra (O'Sullivan e Unwin, 2003).

Un'ultima serie di problemi che s'incontra nell'analisi dei dati spaziali è sostanzialmente di tipo operativo. Un argomento che distingue la statistica spaziale dalla statistica convenzionale può essere riassunto in questa frase: "lo spazio non è uniforme". Se si studiano fenomeni connessi alla geografia dell'uomo, ad esempio, si scopre che ci sono elementi sul territorio che non sempre si riescono a controllare teoricamente. Una città è fatta di case abitate, ma anche di zone lavorative, di distretti industriali, di incroci di strade, di parchi pubblici. Un'analisi condotta all'interno del territorio di una città potrebbe produrre dei risultati che non hanno alcun senso non appena si considera anche la geografia urbana sottostante i dati.

Un particolare tipo di problema legato alla non uniformità dello spazio è dovuto agli *edge effects*, che sorgono quando per svariati motivi una regione viene suddivisa in aree più piccole. Le aree interne non risentiranno di questo effetto, poiché saranno confinanti in tutte le direzioni con altre aree, mentre quelle poste sul confine avranno aree adiacenti solamente verso la direzione rivolta verso il centro. Questo potrebbe produrre delle asimmetrie nei risultati dell'analisi: per com'è definita la zona principale di studio infatti, i dati delle aree che fuoriescono da quest'ultima non



sono disponibili, e quindi le zone confinanti non possono essere confrontate con altre zone al di fuori della regione (O'Sullivan e Unwin, 2003).

### 3. La popolazione come distribuzione di punti nello spazio: due approcci

#### 3.1. KDE e la costruzione di superfici di densità

Grazie alle banche dati oggi disponibili da parte dei principali enti locali territoriali (regioni, province, comuni) è possibile attribuire dati di tipo demografico e anagrafico a porzioni di territorio molto ridotte. È frequente quindi il ricorso a sezioni di censimento o numeri civici per le analisi relative alla distribuzione della popolazione in ambito urbano. Entrambe le tipologie geografiche di dato possono essere ricondotte a insiemi di punti nello spazio: nel caso delle sezioni di censimento ciò implica la conversione dell'unità poligono in punto considerando le coordinate del suo centroide. In questo modo è possibile lavorare con distribuzioni di punti che semplificano il lavoro di stima delle densità e di realizzazione di superfici continue.

Una metodologia che ha rivelato la sua efficienza sia dal punto di vista dell'analisi sia della visualizzazione dei risultati, è costituita dalla funzione KDE (*Kernel Density Estimation*), che fornisce una stima di densità di un fenomeno puntuale, producendo una superficie a tre dimensioni a partire da un insieme di punti distribuiti su di una regione dello spazio. Nelle parole di Gatrell et al., (1996), ciò significa una 'funzione mobile a tre dimensioni che pesa gli eventi entro la sua sfera di influenza, a seconda della loro distanza dal punto

dal quale viene stimata l'intensità', che si può indicare:

$$\hat{\lambda}(s) = \sum_{i=1}^n \frac{1}{\tau^2} k\left(\frac{s-s_i}{\tau}\right)$$

in cui  $\hat{\lambda}(s)$  rappresenta la stima dell'intensità della distribuzione di punti, misurata nel punto  $s$ ;  $s_i$  è l' $i$ -esimo evento, e  $k(\cdot)$  rappresenta la funzione di *kernel* mentre  $\tau$  la soglia. Tale soglia consente di ottenere superfici più o meno arrotondate del fenomeno, visualizzando la sua distribuzione a diverse scale.

Il metodo si compone di diversi passaggi dal punto di vista realizzativo:

1. una griglia fine viene sovrapposta alla regione di studio;
2. la funzione *kernel* a tre dimensioni, per la quale viene specificato un determinato raggio, o soglia, passa su ogni cella della griglia e calcola il peso di ogni punto all'interno del raggio. Generalmente, punti più vicini al centro della cella riceveranno un valore più elevato, attribuendo quindi alla cella un valore di densità più elevato;
3. il valore finale di ogni cella viene calcolato sommando tutti i valori delle funzioni *kernel* calcolate in ogni cella vengono calcolati sommandoli.

Il funzionamento della procedura è esemplificato nella Figura 2, dove la funzione *kernel* viene presentata in due dimensioni. Su ogni singolo punto viene calcolata la funzione di densità, generalmente di tipo normale, basata sul *kernel*. Successivamente la somma delle funzioni singole produce la stima di densità complessiva.

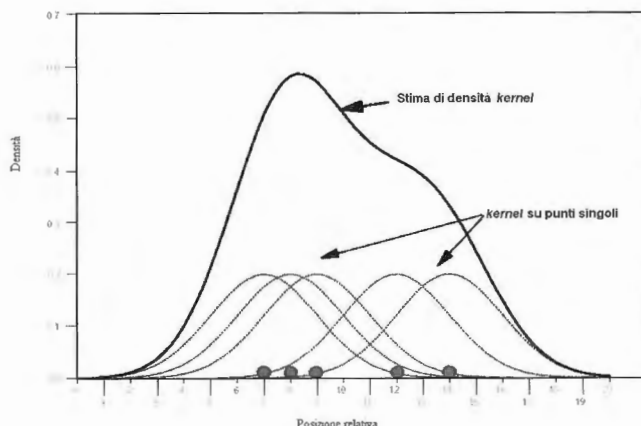


Fig. 2. La funzione kernel in '2 dimensioni' (elaborazione da Levine, 2002).

Molte discipline, quali la meteorologia, la geologia, la biologia etc., prendono in considerazione dati che giacciono su di una superficie non piana; si è cioè in presenza di un gradiente, che rappresenta in un certo modo la pendenza della superficie per ogni punto oggetto d'analisi. In questo contesto sorgono diversi problemi poiché la maggior parte dei metodi ipotizza la complanarità dei dati, o quanto meno considera trascurabile la loro differenza di quota. Si pensi, ad esempio, alla distribuzione di alberi in una foresta in una zona collinare: si vuole analizzare la quantità d'ossigeno proveniente dagli alberi di una certa specie. Questa variabile dipenderà certamente dall'età degli alberi, ma potrebbe essere influenzata anche dalla distanza tra le piante: questa ipotesi discende dalla considerazione che se due alberi sono troppo vicini, si fanno ombra l'uno con l'altro e quindi si tolgono a vicenda la luce del sole, componente essenziale nel processo di produzione d'ossigeno da parte dei vegetali. La quantità d'ossigeno prodotta, però, potrebbe anche dipendere dalla posizione delle piante: una zona esposta a sud è favorita in questo senso rispetto ad una esposta a nord. La pendenza stessa della zona oppure alcune caratteristiche del suolo, come la presenza o meno d'acqua o di sostanze minerali disciolte nel terreno, possono influire sulla variabile d'interesse. Ammettendo che gli alberi siano distribuiti sulla regione di studio in maniera casuale, (ad esempio siano distribuiti in modo uniforme per unità di superficie) le distanze reciproche sono variabili: per stimare la quantità d'ossigeno prodotta si possono applicare tecniche Monte Carlo simulando le posizioni delle piante sul terreno attraverso un opportuno modello di distribuzione casuale.

Recentemente sono stati presentati due algoritmi per la generazione di punti casuali su superfici (Melfi e Schoier, 2004). Il procedimento comune a entrambi gli algoritmi è quello di generare prima dei punti indipendenti, identicamente ed uniformemente distribuiti sulla proiezione della superficie su di un piano; quindi, a seconda delle ipotesi, costruire l'insieme finale di punti accettando solo quelli che soddisfano a particolari condizioni.

Dati un insieme  $D$  compatto tale che  $D \subset \mathbf{R}^2$ , una funzione  $f$  differenziabile, definita in  $D$ , una superficie d'interesse  $S$  definita come:

$$S = \{(x, y, f(x, y)) \in \mathbf{R}^3 : (x, y) \in D\},$$

e avendo a disposizione un generatore di numeri

pseudo-casuali uniformi, in grado di produrre cioè una sequenza

$$\{u_h\}_{h \in \mathbf{N}}, u_h \in [0, 1],$$

i due algoritmi permettono di generare punti casuali sulla superficie  $S$  tenendo conto della sua pendenza. Il primo consente di generare punti sulla superficie  $S$ , che provengono da una distribuzione casuale uniforme, indipendenti tra loro, e viene indicato con il nome di *Uniform Random Distribution Algorithm* (URDA), il secondo è l'estensione del primo ad una distribuzione non uniforme in cui la densità dei punti simulati sulla superficie  $S$  non dipende solo dalla pendenza di tale superficie, fattore codificato in una funzione  $m$ , ma anche da altre componenti rappresentabili attraverso un'appropriata funzione  $t$  positiva, il prodotto delle due da luogo alla funzione  $m_t$ . Esso prende il nome di *Non Uniform Random Distribution Algorithm* (NURDA).

Questo può essere il caso di una specie di pianta che cresce solo al di sotto di una certa altitudine, o in presenza di acqua: alle coordinate corrispondenti alle zone favorite vengono assegnate probabilità maggiori di selezione (cioè valori più alti della funzione  $t$ ) dell'eventuale punto simulato che vi cade, rispetto a zone sfavorite. (Per una descrizione dei due algoritmi si rimanda a Melfi e Schoier, 2004).

## 4. Applicazione

### 4.1. I dati

I dati analizzati fanno riferimento alla popolazione residente nel Comune di Trieste di nazionalità cinese. Questi dati derivano dall'anagrafe comunale e sono aggregati a livello di sezione di censimento. In altre analisi sulla distribuzione e la densità della popolazione (Borruso e Donato, 2003) la base di partenza era costituita da numeri civici cui veniva attribuito come peso la popolazione residente. La necessità di omologare le basi di dati per l'analisi tramite KDE e le simulazioni con l'algoritmo NURDA hanno suggerito il ricorso a sezioni di censimento quale ambito di partenza<sup>1</sup>.

La scelta di tale aggregazione presenta dei vantaggi a fronte di limitati svantaggi. I vantaggi riguardano soprattutto la possibilità di replicare le analisi in altri ambiti, urbani e non, con relativa facilità senza adattare in modo particolare gli algoritmi utilizzati. Tra gli svantaggi va ricordata so-





prattutto la struttura areale delle sezioni di censimento che potrebbe portare a problemi simili al MAUP. Tuttavia tale svantaggio risulta alquanto limitato soprattutto con riferimento alle aree centrali urbane caratterizzate da sezioni di censimento ad alta densità insediativa e piuttosto omogenee quanto a dimensioni.

I dati sono raccolti in tabelle in cui viene riportato il codice che rappresenta la sezione di censimento del Comune di Trieste, il numero complessivo di residenti di nazionalità cinese in ogni sezione, nonché le coordinate metriche del centroide della sezione espresse nel sistema di riferimento locale "Gauss Boaga Roma 40 fuso Est". I dati disponibili sono riferiti al 31 dicembre degli anni 2000, 2001, 2002 e 2003.

#### 4.2. Superfici di densità con il KDE

Un'analisi di densità basata sullo stimatore KDE è stata quindi effettuata sulle distribuzioni di dati relative ai diversi anni presi in considerazione. Ci si è concentrati soprattutto sugli anni 2001 e 2003, per osservare le variazioni nella densità e nella distribuzione della popolazione cinese intervenute in un breve periodo di tempo all'interno del tessuto urbano di Trieste. Le osservazioni riguardano la popolazione residente per sezione di censimento.

L'elaborazione con il KDE è stata sviluppata utilizzando un'ampiezza di soglia  $\tau$  di 300 metri, e ipotizzando una distribuzione normale. Al fine di evidenziare con maggiore efficacia i risultati ottenuti, sia per i dati reali, e soprattutto con riferimento alle stime ottenute per mezzo dell'algoritmo NURDA, il peso, in termini di residenti per sezione di censimento, è stato moltiplicato per 100.

L'analisi di densità mostra la tendenza alla concentrazione degli immigrati cinesi nella parte centrale della città in prossimità della linea costa. Con particolare riferimento al 2001 (Figura 3), notiamo come, a fronte di un nucleo centrale a elevata concentrazione si possono notare dei picchi meno elevati sia in corrispondenza delle linee di comunicazione radiali in uscita dal centro, sia un'area di densità nella parte ovest, sempre lungo la linea della costa. Il 2003 presenta delle differenze rispetto al 2001, che si traducono soprattutto in una minore diffusione spaziale ma in una maggiore concentrazione<sup>2</sup>. Notiamo, infatti, che la distribuzione lungo le linee radiali tende a diminuire, a fronte di un consolidamento della concentrazione in un'area centrale, a sua volta suddivisibile in due nuclei più consistenti, rispettivamente verso nord e sud della stessa area. L'area di picco localizzata a ovest tende altresì a diminuire notevolmente, contribuendo a sua volta alla diminuzione della dispersione.

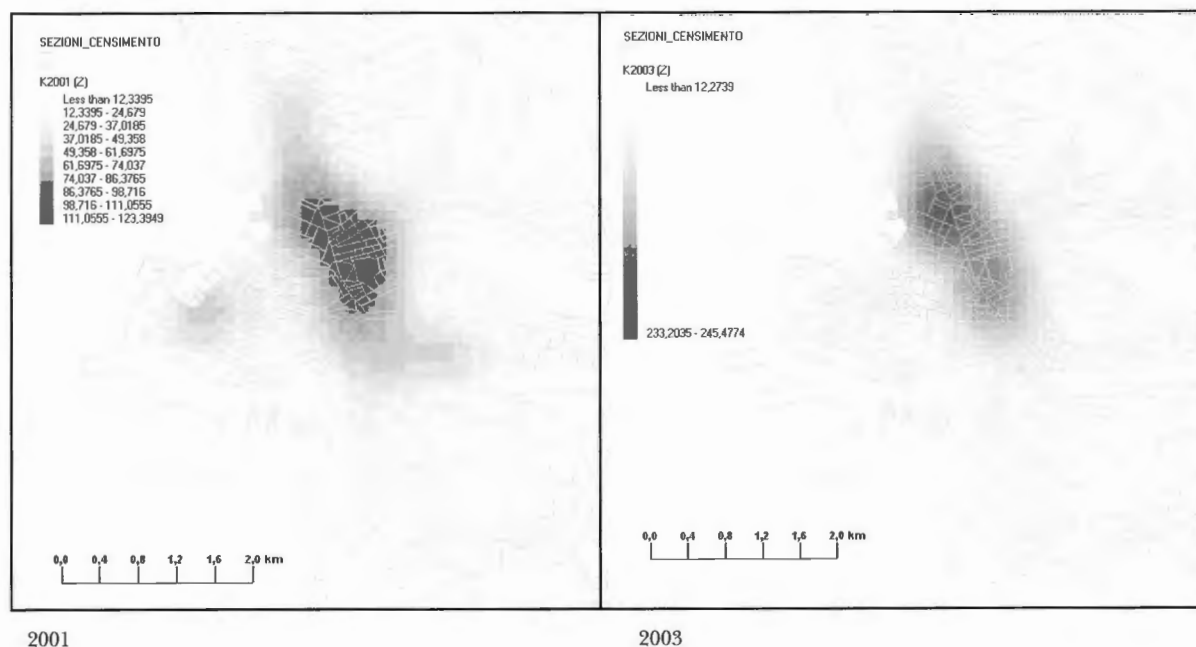


Fig. 3. KDE calcolato sulla popolazione cinese residente. Soglia  $\tau = 300$  m.

Per effettuare la costruzione delle superfici di densità, si è scelto l'Algoritmo NURDA (*Non Uniform Random Distribution Algorithm*) perché presenta maggior aderenza alle ipotesi del problema ed ai risultati che si vogliono ottenere. L'Algoritmo URDA infatti genera punti casuali provenienti da una distribuzione uniforme, ma l'ipotesi di uniformità in questo caso non sembra realisticamente essere verificata.

Nell'applicare l'algoritmo NURDA, in cui la funzione  $t$  esprime l'informazione relativa alla densità di popolazione di nazionalità cinese già presente<sup>3</sup>, si utilizza una funzione  $f$  costante: di fatto la superficie sulla quale si simulano i nuovi punti si riduce ad un piano<sup>4</sup>.

La parte più complessa per quanto riguarda l'applicazione dell'algoritmo è stata la costruzione della funzione  $t$ ; infatti i dati a disposizione per ogni anno rappresentano punti nello spazio tridimensionale, non una funzione. Si è quindi pensato di far passare per tre punti adiacenti un piano e poi definire la funzione  $t$  come unione di questi triangoli nello spazio, ma il problema principale riguardava la scelta dei punti. Seguendo le metodologie usate in questi casi, specialmente per ciò che riguarda il campo dell'analisi spaziale, ci si è attenuti alle regole della *triangolazione di Delaunay* per partizionare il piano e quindi scegliere i triangoli sui quali costruire i pezzi di piano.

Questa operazione è stata fatta per ogni anno in analisi, in quanto ogni anno cambia la densità di popolazione e la sua disposizione sul territorio.

Di seguito (Figura 4) si presentano le superfici di densità riguardanti gli anni considerati a partire dai dati stimati.

Anche in questo caso si può osservare la tendenza alla concentrazione dei residenti cinesi in un'area centrale della città di Trieste, a sua volta suddivisibile in due picchi principali, rispettivamente a nord e a sud-est. È chiara inoltre la riduzione di densità in altre zone della città, in particolare nella parte sud-occidentale lungo la costa, e lungo le vie di comunicazione radiali dal centro storico. I valori di densità stimati per il 2001 e per il 2003 non si discostano di molto da quanto osservato per i dati reali.

La variazione nell'assetto distributivo della popolazione nei due anni considerati si può osservare nella seguente Figura 5. In essa sono riportate le carte di densità delle variazioni rispettivamente reali e stimate per mezzo dell'algoritmo NURDA. Le due carte confermano quanto già osservato: sono evidenti le aree di aumento della densità di residenti cinesi (aree più scure), quelle in cui vi sia stata una diminuzione (aree bianche) o altre di sostanziale stabilità (aree chiare). Mentre l'analisi delle differenze elaborata sui dati reali mostra in modo evidente la formazione di picchi di densità e di concentrazione di popolazione, oltre che la diminuzione delle aree di residenza cinese, la rappresentazione basata sui dati simulati, pur eviden-

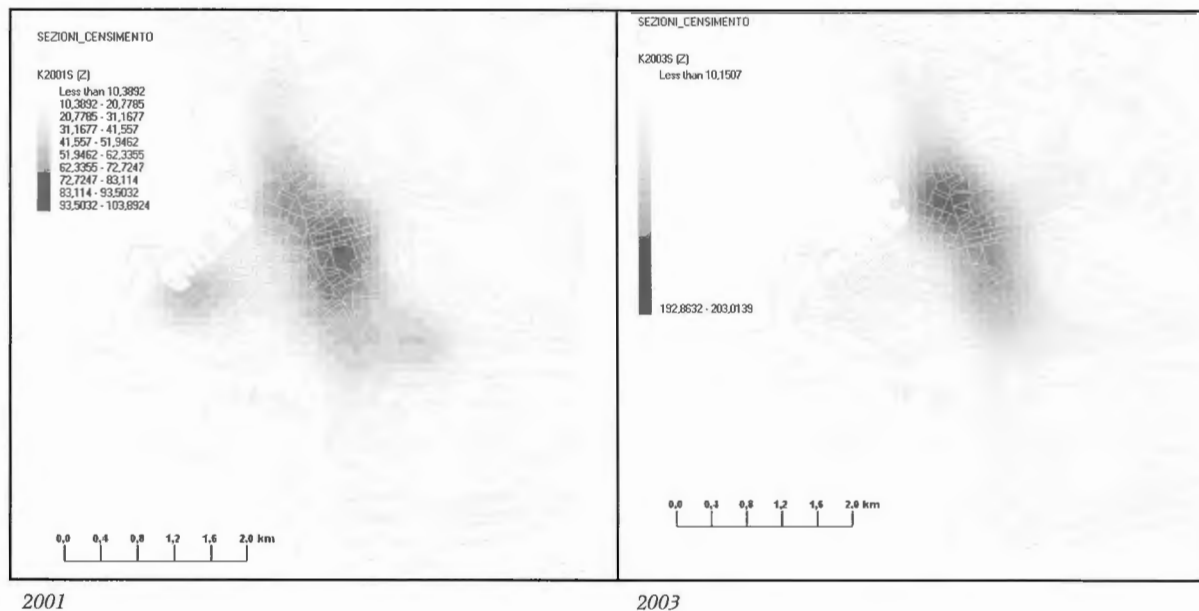
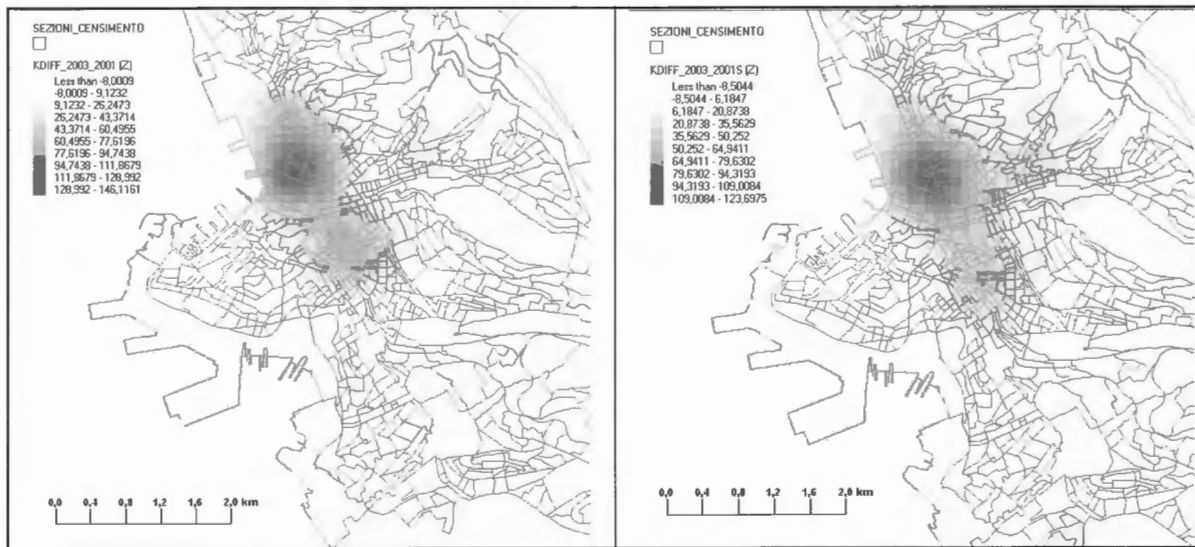


Fig. 4. KDE calcolato sulla popolazione cinese residente stimata con l'algoritmo NURDA. Soglia  $\tau = 300$  m.



Differenza 2003 - 2001

Differenza 2003 - 2001 (stimata)

Fig. 5. KDE calcolato sulla differenza di popolazione cinese residente reale e stimata con l'algoritmo NURDA. Soglia  $\tau = 300$  m.

ziando lo stesso andamento evolutivo, presenta delle 'inerzie' localizzative in aree in cui in realtà si è assistito a una riduzione della popolazione residente cinese.

## 5. Conclusioni

In questo lavoro si è analizzata la distribuzione spaziale in ambito urbano della popolazione di residenza cinese con riferimento in particolare a due momenti temporali, vale a dire il 31 dicembre 2001 e 31 dicembre 2003. Da un lato l'analisi ha riguardato l'esame della densità della popolazione per mezzo di uno strumento, quale il KDE, che permette di evidenziare in modo chiaro aree di maggiore concentrazione della popolazione, nonché osservarne le variazioni nell'assetto distributivo che intervengono nel corso del tempo. Dall'altro lato si è applicato un algoritmo di simulazione di punti su superfici che fornisce una stima dell'andamento della popolazione. L'ipotesi comune alle due procedure utilizzate consiste nell'utilizzare dei dati di tipo puntuale come base di partenza da trasformare in un fenomeno continuo rappresentabile mediante una superficie.

Si rendono tuttavia necessarie alcune considerazioni relative ai due algoritmi utilizzati. Il KDE è stato utilizzato sia con riferimento alla sua caratteristica di trasformare dati puntuali in superfici di densità continue, a partire da osservazioni reali, sia per effettuare elaborazione da dati stimati. Il

KDE è stato quindi utilizzato su dati reali e su risultati della stima dell'algoritmo NURDA.

Diversamente dal KDE, l'algoritmo NURDA fornisce stime di eventi puntuali a partire da una certa distribuzione di dati reali nello spazio. Per tali dati si suppone che siano distribuiti su di una superficie a tre dimensioni. Anche in questo caso si interpreta la popolazione come fenomeno continuo, da cui originano nuovi eventi, in questo caso residenti.

## Note

\* Sebbene la struttura generale del lavoro rifletta il comune intento dei due autori, il Dr. Borruso ha curato i paragrafi 1, 2.1, 3.1, 4.2; la Prof.ssa Schoier ha curato i paragrafi 2.2, 3.2, 4.1, 4.3, 5.

<sup>1</sup> Elaborazioni di confronto effettuate tra la popolazione cinese rapportata al numero civico e alla sezione di censimento non hanno mostrato differenze apprezzabili per quanto riguarda la distribuzione del fenomeno e le densità ottenute. Oltre che un riscontro positivo ai fini del presente studio, tale dato indica come sia possibile ottenere delle superfici di densità realistiche anche in presenza delle 'sole' sezioni di censimento, che generalmente rappresentano dati geografici più facilmente disponibili da fonti statistiche ufficiali rispetto ai numeri civici. Il risultato vale tuttavia soprattutto per le aree urbane, in cui le sezioni di censimento presentano forme e superfici piuttosto regolari e relativamente omogenee, quindi facilmente confrontabili tra di loro.

<sup>2</sup> Un'analisi svolta sulle sezioni di censimento del Comune di Trieste mostra come a fronte di un aumento della popolazione residente cinese non vi sia un corrispondente aumento del numero delle sezioni ove questa è registrata. 103 sezioni di

censimento ospitavano 399 residenti cinesi nel 2001 mentre nel 2003 questi passavano a 555. Le sezioni coinvolte non rimangono costanti nel tempo ma mostrano un abbandono di quelle più periferiche a favore delle più centrali.

<sup>4</sup> Si osservi che gli algoritmi permettono non solo di costruire le superfici di densità ma anzi il loro scopo principale è la previsione in quanto permettono di simulare, in questo caso, il numero di nuovi residenti di nazionalità cinese nel Comune di Trieste.

<sup>5</sup> Si sarebbe potuto pensare di invertire le due funzioni, vale a dire, usare la superficie di densità come  $f$  e definire una  $t$  costante, decidendo cioè una quota (ad esempio il 10%) oltre la quale il nuovo punto simulato sarebbe stato favorito per la scelta di un'abitazione; in questo caso però sorgono problemi di interpretazione della funzione  $m_1$ . Mentre su di una superficie reale, infatti, ha senso richiedere l'uniformità nella distribuzione di punti tra zone di pendenza diversa (la funzione  $m_1$  ha proprio questo scopo), nelle superfici fittizie, come questa di densità, questa scelta non sembra altrettanto giustificata; inoltre la funzione  $t$  nel caso in analisi non dipende dalla quota dei punti da simulare quindi non avrebbe senso considerare una quota variabile nelle coordinate geografiche della popolazione. Nell'analizzare però zone più estese, ad esempio un territorio nazionale o addirittura una regione che comprenda più nazioni, si può considerare una superficie  $S$  non pianeggiante; pensando all'Italia come regione di studio ed avendo a disposizione dati aggregati per Comune, si possono tenere in considerazione le altitudini dei Comuni stessi e costruire quindi la superficie  $S$ .

## Bibliografia

- Bailey T. C. e Gatrell A. C., *Interactive Spatial Data Analysis*, London, Prentice Hall, 1995.
- Bao S., *Literature Review of Spatial Statistics and Models*, China Data Center, <http://141.211.136.209/cdc/docs/review.pdf>, 2004.
- Borruso G. e Donato C., *L'immigrazione straniera a Trieste - I principali impatti sulla situazione socio-economica e sul tessuto urbano*, in Quaderni del Centro studi economico-politici "Ezio Vanoni", N.ro 3-4 Luglio-Dicembre 2003, Trieste.
- Burrough P. A. e McDonnell R. A., *Principles of Geographical Information Systems*, Oxford, Oxford University Press, 1998.
- Dorling D., *Cartograms for visualising human geography*, in Heamshaw H. e Unwin D. (ed.), *Visualisation and GIS*, London, Belhaven press, 1994.
- Dykes J. - Unwin D., *AGOCG. Maps of Census: a rough guide*. <http://www.geog.le.ac.uk/jad7/AGOCG>, 1998.
- Gatrell A. C., Bailey T. C., Diggle P. J. e Rowlingson B. S., *Spatial point pattern analysis and its applications in geographical epidemiology*, in "Transactions of the Institute of British Geographers", 21, pp. 256-274, 1996.
- Goodchild M.F., *Geographical Data Modelling*, in "Computer and Geosciences", 18 (4): 401-408, 1992.
- Kaplan W., *Advanced Calculus, 4th ed.*, Reading, Massachusetts, Addison-Wesley Edition, 1992.
- Langford M., Unwin D., *Generating and Mapping Population Density Surfaces within a Geographical Information System*. In "The Cartographic Journal", 31, 1994: 21-6.
- Levine N., *CrimeStat II: A Spatial Statistics Program for the Analysis of Crime Incident Locations (version 2.0)*. Ned Levine & Associates, Houston, TX, and the National Institute of Justice, Washington, DC., 2002.
- Martin D. (1989), *Mapping population data from zone centroid locations*, in "Transactions of the Institute of British Geographers", 14: pp. 90-97.
- Melfi G. e Schoier G., *Simulation of random distributions on surfaces*, in "Atti della XLII Riunione Scientifica della SIS", Bari, 2004.
- Mennis J. (2003), *Generating Surface Models of Population Using Dasymeric Mapping*, in "The Professional Geographer", 55: pp. 31-42.
- Mesev V., Longley P., *The Role of Classified Imagery in Urban Spatial Analysis*. In Atkinson P. M. and Tate N. J. (ed.), *Advances in Remote Sensing and GIS Analysis*. Chichester, Wiley, 2000.
- O' Sullivan D. e Unwin D. J., *Geographic Information Analysis*, Hoboken, New Jersey, J. Wiley & Sons, inc., 2003.
- Openshaw S., *The Modifiable Area Unit Problem*, in "Concepts and Techniques in Modern Geography No. 38", Norwich: Geo Books., 1984.
- Romei P., *L'immigrazione cinese a Prato tra agglomerazione e diffusione*, in Donato C., Nodari P., Panjek A. (ed.), *Oltre l'Italia e l'Europa - Beyond Italy and Europe. Ricerche sui movimenti migratori e sullo spazio multiculturale*, (Atti del Convegno), Università degli Studi di Trieste, Dipartimento di Scienze Geografiche e Storiche, Trieste, 2004, pp. 237-244.
- Rowlingson B., Baddeley A., Turner R. e Diggle P., *A Package For Spatial Statistics* in "Proceedings of the 3rd International Workshop on Distributed Statistical Computing" (DSC 2003), sito: <http://www.ci.tuwien.ac.at/Conferences/DSC-2003/Drafts/RowlingsonEtAl.pdf>, 2003.

