UNIVERSITA' DEGLI STUDI DI PADOVA

FACOLTA' DI SCIENZE STATISTICHE

CORSO DI LAUREA TRIENNALE
IN STATISTICA E GESTIONE DELLE IMPRESE

GIS, GEOMARKETING E STATISTICA: CONCETTI TEORICI E STUDIO DI FATTIBILITA'

RELATORE: Prof. ssa Susi Dulli
CORRELATORE: Dott. ssa Silvia Rita Sedita

LAUREANDA: Silvia Gambato

ANNO ACCADEMICO 2005-2006

...alla mia famiglia

Ringrazio la prof. ssa Dulli per la disponibilità accordatami in questi mesi, spero che anche in futuro avrò la possibilità di trovare in Lei una fonte così preziosa di consigli.
Ringrazio la dott. ssa Sedita per la collaborazione nella ricerca delle associazioni culturali.
Il ringraziamento più grande va alla mia famiglia, compreso Andrea, che in tutti questi anni hanno saputo sostenermi nel mio cammino.

INDICE

INDICE..... 5

PREFAZIONE..... 7

CAPITOLO 1

SITEMI INFORMATIVI TERRIOTIRALI

1. Che cos'è un sistema informativo territoriale..... 9

1.1 Informazione e sistema informativo..... 9

1.2 Informazione territoriale..... 13

1.3 Sistema Informativo Territoriale..... 16

1.4 Breve storia..... 18

2. Dentro un GIS..... 20

2.1 Modellazione dello spazio: modello vector e modello raster... 20

2.2 Topologia..... 27

2.3 Dbms e query..... 30

2.4 Overlay..... 33

2.5 Buffer..... 36

2.6 Map algebra..... 37

CAPITOLO 2

GEOMARKETING E STATISTICA

1.Geomarketing..... 39

2. Disponibilità di dati statistici..... 41

3. Disponibilità di una metodologia statistica adeguata..... 43

3.1 Definizione di disegni campionari di dati spaziali..... 44

3.2 Interpolazione areale per la stima di informazioni

a livello di disaggregazione inferiore di quello disponibile..... 47

3.2.1 Metodi basati sulla teoria dei campioni classica..... 48

3.2.2 L'approccio top-down..... 49
3.2.3 L'approccio bottom-up..... 50
3.3 Modello di regressione con dati areali..... 51
3.3.1 Rimozione delle ipotesi sul modello
di campionamento..... 55
- L'approccio basato sulla rispecificazione..... 56
- L'approccio basato sulla autocorrelazione..... 57
3.3.2 Rimozione delle ipotesi sul modello
probabilistico..... 58
4. Disponibilità di un supporto informativo adeguato..... 60

CAPITOLO 3
STUDIO DI FATTIBILITA'
1. Introduzione..... 61
2. Ricerca dei dati..... 62
3. Memorizzazione dei dati..... 68
4. Analisi dei dati..... 70
5. Conclusioni..... 73

APPENDICE
IMMAGINI RELATIVE A CAP. 1..... 75

BIBLIOGRAFIA..... 85

PREFAZIONE

L'idea dello sviluppo del presente elaborato nasce all'interno del corso di Sistemi Informativi Aziendali tenuto dalla prof. ssa Dulli alla Facoltà di Scienze Statistiche dell'Università degli Studi di Padova.

Il testo è articolato in tre capitoli dei quali si dà una breve presentazione nel seguito.
Il primo capitolo ha la funzione di introdurre i Sistemi Informativi Territoriali. Nella prima parte, in particolare, vengono descritte alcune idee chiave legate all'informazione territoriale e ai sitemi informativi; nella seconda parte, invece, si entra nel dettaglio del software cercando di spiegare quali sono i concetti alla base delle molteplici operazioni che un GIS può eseguire.
Nel secondo capitolo si parla di geomarketing e statistica. Dopo una breve introduzione del termine geomarketing, infatti, si analizzano in dettaglio gli strumenti, che secondo Arbia, tale disciplina dovrebbe usare: adeguati dati statistici, metodologia affidabile di statistica spaziale e software GIS. Per quanto riguarda la statistica spaziale, in particolare, si è cercato di analizzare il campionamento areale, la stima per piccoli campioni e la regressione con dati spaziali.
Nell'ultimo capitolo, infine, si è cercato di descrivere l'esperienza personale maturata nello studio delle associazioni musicali della provincia di Padova eseguita in collaborazione con la dott. ssa Sedita. Si è riporta, in particolare, la fase di ricerca di un elenco attendibile delle unità di interesse. Ricerca lunga e complessa che ha portato alla scoperta di una serie di registri di organizzazioni non profit. Si sono, poi, valutate le informazioni trovate ai fini di una memorizzazione in tabelle excel per essere elaborate con software statistici o GIS. Infine si sono proposte delle possibili analisi da eseguire con i dati trovati.

La decisione di articolare in tale modo l'elaborato è dovuta al fatto che dopo aver presentato le potenzialità dei GIS si è cercato di mettere in evidenza come

queste possano essere di supporto alle decisioni. Si è scelto in particolare di analizzare il legame delle informazioni territoriali con il marketing. All'interno del geomarketing, quindi, si è cercato di studiare in dettaglio quali sono le metodologie di statistica spaziale che hanno la funzione di supportare l'attività decisionale del management. L'ultima parte, infine ha lo scopo di studiare la fattibilità dell'applicazione pratica dei concetti descritti nei capitoli precedenti.

Le difficoltà incontrate nella stesura di questo elaborato erano dovute alla mancanza di libri recenti che mi permettessero di avere una visione d'insieme degli argomenti trattati e dai quali partire per ulteriori approfondimenti. Fortunatamente l'entusiasmo comunicatomi dalla prof. ssa Dulli per questo campo di ricerca mi hanno permesso di portare a compimento questo lavoro.

CAPITOLO 1
SISTEMA INFORMATIVO TERRITORIALE

1. CHE COS'E' UN SISTEMA INFORMATIVO TERRITORIALE

Prima di dare una definizione di sistema informativo territoriale, vengono richiamati, in forma schematica, alcuni concetti fondamentali per una comprensione approfondita dell'argomento.

1.1 INFORMAZIONE E SISTEMA INFORMATIVO

Dal dizionario della lingua italiana Devoto-Oli, si apprende che la definizione di INFORMAZIONE è la seguente:
"Quanto venga accolto o comunicato nell'ambito di un'utilità e funzionalità pratica ed immediata".

In questa frase, implicitamente, vengono esposti i tre elementi essenziali che servono a formare l'informazione:

- il contenuto: il "quanto" che viene accolto o comunicato;
- gli attori: coloro che comunicano;
- il fine: l'utilità per cui avviene lo scambio tra gli attori.

Specificando questi tre elementi al nostro caso particolare, possiamo trarre le seguenti considerazioni:

- Gli attori vengono identificati con i lavoratori di un' ORGANIZZAZIONE. Per organizzazione si intende un'insieme di risorse pianificate per conseguire specifici obiettivi. La definizione, volutamente generale, comprende sia le organizzazioni di tipo industriale che quelle di servizi, come, per esempio la pubblica amministrazione.
- Il fine dello scambio comunicativo è lo svolgimento dei PROCESSI, cioè sequenze di attività che consentono di trasformare i beni che l'organizzazione acquista in prodotti successivamente immessi nel mercato. Si distinguono due tipi di processi, quelli operativi, composti dalle attività produttive, cioè di trasformazione; e quelli decisionali, nel cui ambito si assumono le decisioni che condizionano i processi operativi.

Il ruolo fondamentale svolto dall'informazione nello svolgimento dei processi di un'organizzazione la pone alla stessa importanza degli impianti, del capitale, dei materiali, ecc. Si può considerare, quindi, una RISORSA.

Un concetto similare, ma distinto dall'informazione è quello di DATO. In generale esso viene definito come:
"Uno degli elementi disponibili per descrivere un oggetto o un evento"[Devoto-Oli].
In informatica, i dati sono degli elementi di informazione costituiti da simboli che debbono essere elaborati per ottenere conoscenza. Per esempio trovare scritto su un foglio di carta "Mario 7645689" è un dato. Se questo, però, viene fornito in risposta alla domanda "Chi è il capo del personale e qual è il suo numero?", allora i dati possono essere interpretati e forniscono conoscenza.

Visto il ruolo chiave dell'informazione, emerge l'esigenza di governare tale risorsa attraverso un SISTEMA INFORMATIVO.

Dare una definizione univoca e precisa di sistema informativo è molto difficile, visto che si tratta di un termine entrato nell'uso comune a cui si associano molti significati.

- *"Il sistema informativo è un componente (sottosistema) di una organizzazione che gestisce (acquisisce, elabora, conserva, produce) le informazioni di interesse (cioè utilizzate per il perseguimento degli scopi dell'organizzazione stessa"* [Atzeni, 2000];
- *"Il sistema informativo e quell'insieme di risorse umane e strumenti per la raccolta, la memorizzazione, l'elaborazione e lo scambio di informazioni, nonché di regole organizzative e procedure automatiche e manuali per la gestione delle informazioni stesse. Il Sistema Informativo consente il corretto funzionamento di un sistema organizzativo di una qualsiasi struttura socio/economica."* [Dulli, 2005];
- *"Il Sistema Informativo è costituito dall'informazione di interesse dell'organizzazione e dall'apparato tecnologico ed organizzativo che tale informazione acquisisce, elabora, rende disponibile ed utilizza."*[Rumor,Marovigh];
- *"Il Sistema Informativo è il complesso di uomini, strumenti e procedure (spesso informali) che permettono l'acquisizione e la distribuzione dei dati nell'ambito dell'organizzazione e che li rendono disponibili nel momento in cui sono richiesti a chi ne ha necessità per svolgere una qualsivoglia attività."* [Bonfatti, 1988].

Il sistema informativo è indipendente dall'automazione tecnologica, tanto che esisteva ancora prima dell'avvento dell'informatica. Bisognerebbe parlare, infatti, di Sistema Informatico per indicare il "sottosistema automatico del sistema informativo". Oggigiorno, però, i due termini vengono usati indistintamente.

Dalle definizioni sopra riportate si possono dedurre i seguenti componenti fondamentali di un sistema informativo:

- Insieme di dati e informazioni, che costituiscono il patrimonio informativo dell'azienda e il cui fabbisogno può essere stimato attraverso la seguente funzione:

$$I = f (V, C, U)$$

  Dove:

  V = volume di attività corrispondente alla quantità di beni e servizi prodotti dall'organizzazione;

  C = complessità dei prodotti o servizi. L'informazione richiesta per descrivere un prodotto è proporzionale alla sua complessità;

  U = incertezza in cui opera l'organizzazione. Maggiore è la volatilità degli andamenti di mercato o dell'incertezza insita nelle prospettive future delle condizioni economiche, maggiore è il fabbisogno informativo.
- Tecnologia, cioè l'insieme di procedure e risorse usate per gestire l'informazione. Non si tratta esclusivamente di tecnologia informatica.
- Insieme di persone, o meglio, contesto organizzativo. Risulta di fondamentale importanza studiare la realtà aziendale, la cultura dell'utenza per progettare un adeguato sistema informativo.

Emerge, quindi, il ruolo centrale del sistema informativo all'interno di un'organizzazione, ruolo esemplificato nella figura seguente.

<u>FIGURA 1</u> (vedi Appendice)

Dalla schematizzazione emerge che i processi produttivi incamerano risorse, le trasformano in beni o servizi da immettere successivamente sul mercato. Questi processi, poi, interagiscono con quelli decisionali attraverso il sistema informativo, che ha anche il compito di acquisire informazioni utili alla vita aziendale dall'esterno e, viceversa, di cedere informazioni verso altri enti o al mercato.

Vediamo ora come si pone il sistema informativo nei confronti dello schema della struttura organizzativa proposto da Anthony.

In questa teoria la realtà aziendale è rappresentata come una piramide divisa in tre blocchi corrispondenti; a partire dal basso, con il Livello Operativo, il Livello Tattico e il Livello Strategico.

FIGURA 2 (vedi Appendice)

Scopo del sistema informativo è quello di fornire ad ogni livello solo le informazioni necessarie.
Si avrà, quindi, a livelli più bassi la necessità di un gran numero di dati che richiedono molteplici operazioni di calcolo e correlazione, mentre, salendo nella scala, saranno richiesti dati più sintetici.

1.2 INFORMAZIONE TERRITORIALE

Visto che lo scopo dell'elaborato è definire un sistema informativo territoriale, l'informazione di interesse è quella territoriale. Si specifica, quindi, di seguito, questo concetto attraverso alcune definizioni.

FIGURA 3 (vedi Appendice)

L'INFORMAZIONE SPAZIALE è riferita ad oggetti nello spazio ed alle mutue relazioni tra questi.
Per esempio possiamo dire che una sedia è vicina ad un tavolo, che una lampada è sopra una scrivania e così via.

L'INFORMAZIONE TERRITORIALE è una parte dell'informazione spaziale riferita più specificatamente ad oggetti nel mondo reale (o che possono essere posizionati nel mondo reale) ed alle relazioni tra questi.

Informazione territoriale e INFORMAZIONE GEOGRAFICA hanno lo stesso significato.

Per INFORMAZIONE NON SPAZIALE si intende un'informazione che non è riferita a oggetti nello spazio né a relazioni spaziali tra di essi, né è in qualche modo riferibile ad una posizione nello spazio.

Un INFORMAZIONE SPAZIALE NON TERRITORIALE è, per esempio, un disegno meccanico, una struttura edilizia, una stanza virtuale.

L'informazione spaziale è composta da proprietà geometriche e proprietà topologiche.

La GEOMETRIA riguarda la forma (cerchio, quadrato, sfera), le dimensione (lunghezza, area, volume), le distanze e gli angoli dei componenti dell'informazione spaziale.

La TOPOLOGIA tratta, invece, delle mutue relazioni tra i componenti, come, per esempio, il contenimento, l'adiacenza e la connessione.
Queste proprietà di posizione non variano se variano forma e dimensione dei componenti.

Le COMPONENTI dell'Informazione Territoriale sono le seguenti:

- componente spaziale: posizione, geometria, topologia;
- componente non spaziale: dati descrittivi (attributi) associati alla componente spaziale.

Si può introdurre, inoltre, una diversa scomposizione in :

- informazione territoriale di base;
- informazione territoriale tematica.

L'INFORMAZIONE TERRITORIALE DI BASE è composta dagli attributi e dalla componente spaziale, che assume il peso maggiore.
Comprende le reti geodetiche, gli elementi naturali del territorio, gli elementi artificiali del territorio, i confini e gli elementi per posizionare le informazioni non territoriali (per esempio il numero civico).

Tale informazione è indispensabile per tutti gli usi, è indipendente dall'uso e dagli utilizzatori. È, inoltre, molto costosa da acquisire e da mantenere aggiornata, e svolge il ruolo di infrastruttura informativa.

Qualora si volesse immettere nel proprio sistema informativo delle informazioni territoriali di base, bisogna, come prima cosa, definire il territorio e l'accuratezza delle informazioni ricercate. Si procede poi, con una ricerca sul mercato dei dati che soddisfino le condizioni poste. Se la ricerca ha esito positivo bisogna analizzare il formato delle informazioni perché se cartaceo bisogna procedere all'acquisizione, mentre se digitale bisogna verificarne la compatibilità o procede ad una conversione. Se, invece, la ricerca ha esito negativo, si possono rivedere le condizioni iniziali, oppure si procedere alla costruzione di tale informazione. In tale caso si dovrà ricorrere agli strumenti della cartografia numerica.
I passi descritti possono essere schematizzati nel seguente diagramma di flusso.

FIGURA 4 (vedi Appendice)

L'INFORMAZIONE TERRITORIALE TEMATICA al contrario di quella di base ha come componente principale gli attributi. Comprende, per esempio informazioni di tipo socio-economico oppure ambientale oppure catastale ed altro ancora. È un'informazione, quindi, che si specializza in funzione dell'uso.

Il processo di reperimento di informazione territoriale tematica è molto simile a quello presentato prima, solo che tra i parametri iniziali bisogna aggiungere anche il tema che l'informazione deve soddisfare.
La costruzione, inoltre, avviene con due tecniche il remote sensing (telerilevamento) e le indagini sul campo.

FIGURA 5 (vedi Appendice)

Molti sono gli enti che mettono a disposizione tali tipi di dati ma se non sono legati alla componente spaziale non servono. A causa, quindi, della diversità delle fonti è necessaria un'accurata indagine preliminare dei metadati.

1.3 SISTEMA INFORMATIVO TERRITORIALE

In base alla nomenclatura adottata per l' informazione abbiamo diversi sistemi informativi.

FIGURA 6 (vedi Appendice)

SIT: Sistemi Informativi Territoriali, che gestiscono informazioni territoriali;
CAD: Sistemi Informativi Spaziali non territoriali, comunemente indicati come sistemi di disegno;
MIS: Sistemi Informativi per il Management, che non trattano informazione territoriali.

Si può ora dare una definizione di SITEMA INFORMATIVO TERRITORIALE.
Riferendoci ai sistemi che trattano informazione spaziale e territoriale, esistono diversi acronimi del termine:
S.I.T.: Sistema Informativo Territoriale;
G.I.S.: Sistema Informativo Geografico (*Geographical Information System*);
S.I.S.: Sistema Informativo Spaziale.
A rigore quello più corretto da usare è SIT; perché GIS indica la componente tecnologica del sistema.

Esistono, inoltre, diverse definizioni di SIT:

- "*A GIS is a powerful set of tools for collecting, storing, retrieving at will, transforming, and displaying spatial data from the real world*" [Burrough 1986]. (Un GIS è una potente serie di strumenti per acquisire, memorizzare, estrarre a volontà, trasformare e visualizzare dati spaziali dal mondo reale).

- "*A GIS is a computer system that can hold and use data describing places on earth's surface*" [Rhind 1989]. (Un GIS è un sistema che può mantenere ed usare dati che descrivono posti sulla superficie terrestre).
- "*A Geographical Information System is a group of procedures that provide data input, storage and retrieval, mapping and spatial analysis for both spatial and attribute data to support the decision-making activities of the organisation*" [Grimshaw 1995]. (Un sistema informativo geografico è un gruppo di procedure che consentono input, memorizzazione, accesso, mapping e analisi spaziale sia per dati spaziali che per attributi, per supportare le attività decisionali dell'organizzazione).
- "*A geographical information system is an organized collection of computer hardware, software, geographic data, and personnel designed to efficiently capture, store, update, manipulate, analyze, and display all forms of geographically referenced information*" [ESRI] (Un sistema informativo geografico è un insieme organizzato di hardware, software, dati geografici e persone progettato per raccogliere, immagazzinare, manipolare, analizzare e rappresentare in modo efficiente tutte le forme di informazione geograficamente referenziata).
- *"Un sistema informativo geografico è un sistema composto da banche dati, hardware, software ed organizzazione che gestisce, elabora ed integra informazione su una base spaziale o geografica"* [Barrett-Rumor 1993].

Ognuna di queste definizioni sottolinea un aspetto diverso di questo mondo complesso: alcune di esse sono basate sul concetto di database, altre su quello di struttura costituita da un insieme di strumenti e tecnologie, infine alcune sottolineano gli aspetti organizzativi. Indipendentemente dalla definizione, però, un GIS permette di integrare operazioni comuni ad un database con i vantaggi offerti dalla visualizzazione su mappa.

La fusione di questi due sistemi ha permesso il superamento del compromesso insito in ogni rappresentazione cartografica; perché ogni raffigurazione di entità geografiche è sempre in qualche misura simbolica e in scala, e questo

rappresenta un limite per una conoscenza completa di tutte le informazioni legate ad essa.
Queste possibilità distinguono un GIS da altri sistemi informativi e lo rendono adatto a vari campi di utilizzo sia nelle aziende pubbliche che private per spiegare eventi, pianificare strategie, ecc ...

1.4 BREVE STORIA

La storia della rappresentazione su carta del mondo circostante inizia con metodi basati su percezioni soggettive. In seguito, il problema di rappresentare il globo terrestre su una superficie piana fu studiato dalle scienze matematiche che svilupparono i rigorosi principi delle proiezioni cartografiche.
Nella storia più recente sono comparse le foto aeree e le immagini satellitari che ci hanno permesso di spostare il punto di osservazione lontano dalla superficie terrestre e ci hanno fatto cogliere la terra come non potevamo immaginarla. Nonostante questo, la richiesta di cartografia è sempre molto alta, specialmente per quanto riguarda le carte tematiche, perché attraverso di esse è possibile analizzare uno o più temi specifici legati comunque alle entità geografiche.
Negli ultimi anni, in questo campo, sono entrate anche le scienze informatiche.
Sul finire degli anni '70 si sviluppa, infatti, la cartografia automatica, cioè una semplice traduzione in ambiente numerico della cartografia tradizionale.
Successivamente si sviluppa la cartografia numerica, nella quale l'elemento base è l'insieme delle coordinate che in forma implicita contiene le informazioni per rappresentare sotto forma di disegno le entità geografiche. Nella cartografia numerica, però, a differenza dei GIS, le entità non sono semanticamente consistenti e non viene gestita la topologia; sono, invece, presenti i livelli o strati informativi nei quali vengono rappresentate entità dello stesso tipo. La cartografia numerica, comunque presenta una serie di vantaggi quali, ad esempio:

- univocità del contenuto metrico: si eliminano i problemi connessi alla

soggettività della misura per determinare le coordinate di punti a partire dal disegno della carta e quelli dovuti alla deformazione nel tempo del supporto cartaceo;

- estensione della tipologia della cartografia: con la possibilità di memorizzazione anche le coordinate altimetriche;
- possibilità di elaborazione dei dati: calcolo delle superfici, delle volumetrie, classificazione, ecc.;
- facilità di gestione e aggiornamento dei dati.

Per la cartografia numerica il concetto di scala è superato poiché le coordinate sono assolute, si usa, invece, il concetto di scala nominale, cioè massimo rapporto di scala a cui si possa stampare una carta in modo tale che essa abbia gli stessi requisiti qualitativi e metrici di una carta tradizionale avente la stessa scala.

Nacquero in seguito i GIS. Il termine "sistemi informatici geografici" fu utilizzato per la prima volta nel 1967 da Roger Tomlison per indicare un sistema informativo sviluppato per trattare l'enorme mole di dati necessario per descrivere le estese zone del Canada.

Il GIS, a quell'epoca, era semplicemente uno strumento che consentiva di maneggiare in modo più agevole, attraverso l'utilizzo di mappe, i dati spaziali raccolti. Poiché in una carta tradizionale l'elemento base è la linea, questi primi sistemi si basavano su un modello di dati vettoriale (vedi in seguito).

Solo alla fine degli anni '80, con l'interessamento delle Amministrazioni locali e delle compagnie di servizi, iniziarono a comparire applicazioni che elaboravano le informazioni sulle entità geografiche.

In contemporanea in quegli anni si elaborò, anche un nuovo modello di memorizzazione dei dati: il modello raster (vedi in seguito); il primo sistema ad usarlo fu il Map Analysis Package, sviluppato da Tomlin nel 1975.

2. DENTRO UN GIS

Dopo aver dato una definizione di GIS, si parlerà di alcune sue caratteristiche e di alcune funzionalità. In particolare si cercherà di capire come è possibile rappresentare lo spazio e le sue entità, come questi dati possono essere immagazzinati e gestiti attraverso alcune particolari operazioni.

2.1 MODELLAZIONE DELLO SPAZIO: MODELLO VECTOR E RASTER

Come si è visto il SIT è di ausilio allo studio di entità georeferenziate, ma per studiare il mondo reale dobbiamo darne una rappresentazione. Si definisce, allora, modellazione la rappresentazione tramite strumenti matematici delle entità del mondo reale e dei reciproci meccanismi di interazione.

Vediamo, ora, che la la tecnologia GIS mette a disposizione due strumenti per rappresentare il mondo reale: il modello raster e il modello vector, che si distinguono per il diverso modo di rappresentare lo spazio.

Anche se ognuno di noi ha un'idea ben definita di spazio è difficile darne una definizione.
Un punto di vista è quello di vedere lo spazio come un sistema di localizzazione, non legato alla presenza di oggetti al suo interno; il punto di vista opposto è quello di considerare lo spazio come un insieme di oggetti dotati di certe caratteristiche.
Per chiarire le idee consideriamo due esempi: nel mosaico lo spazio viene preso in considerazione nella sua totalità, indipendentemente dagli oggetti, infatti, esistono tessere anche dove non viene rappresentato nessun oggetto; al contrario, consideriamo un foglio bianco, dove con una penna si tracciano dei segni, lo spazio, in questo caso, è vuoto e viene riempito con degli oggetti.
Nel mondo della grafica troviamo due strumenti di modellazione che

riprendono questi concetti: le primitive vettoriali (vector) e la primitiva raster.
Le primitive vettoriali nascono dai concetti della geometria euclidea ,infatti, esse sono il punto, la linea e l'area. La primitiva raster (il pixel) nasce, invece, dal concetto di una tassellazione sistematica e totale dello spazio.

In conclusione lo spazio geometrico può essere rappresentato pertanto secondo due criteri o modelli: il modello vettoriale e il modello raster. Ciascuno dei due modelli, a sua volta, si esprime attraverso alcune primitive: punto, linea, area per il modello vettoriale ; pixel per il modello raster.

La primitiva PUNTO è costituita da una coppia di coordinate, per quanto riguarda la parte geometrica, più l'insieme degli attributi per la parte non spaziale dell'informazione.
Un oggetto del mondo reale può essere modellato con la primitiva Punto se:

- è semanticamente un punto (tale cioè da non perdere la propria caratteristica geometrica al cambiare della scala a cui si opera) [esempio: vetta di un monte oppure punto trigonometrico];
- è un punto alla scala a cui si opera [esempio: bocca di un pozzo rappresentata in piccole scale];
- è un punto per l'applicazione che stiamo sviluppando [esempio:cittadina rappresentata in una carta turistica].

Quando si usa la primitiva punto, quindi, occorre tener presente il problema di modellazione, consistente nel valutare "quanto bene la primitiva punto modella l'entità del mondo reale che si deve rappresentare".

La primitiva LINEA è costituita da un insieme ordinato di punti, più l'insieme degli attributi; ogni coppia di punti contigui della sequenza identifica un segmento facente parte della primitiva; il punto iniziale e quello terminale sono chiamati estremi; i punti intermedi si chiamano vertici.
Un oggetto del mondo reale è rappresentabile con una linea se è semanticamente una linea, se è una linea alla scala a cui si opera o se è una

linea per l'applicazione che stiamo progettando.

Quando si rappresenta una entità territoriale per mezzo della primitiva linea, purtroppo, si commettono degli errori, vediamo quali.
Nella maggior parte dei casi gli oggetti del mondo reale che descriviamo tramite una linea spezzata sono in realtà linee curve, di conseguenza, stiamo introducendo un errore, detto errore di discretizzazione. Occorre, però, ricordare, che in tutte le misure delle coordinate che si fanno si commettono degli errori; tale fenomeno potrebbe essere rappresentato disegnando una fascia attorno alla linea rappresentata. In conclusione, se l'errore di discretizzazione è tale che le corde (i segmenti della linea) sono mantenute all'interno della fascia, possiamo dire che il nuovo errore introdotto è nei limiti dell'errore (più grossolano) compiuto nell'acquisizione delle coordinate dei vertici, e come tale trascurabile.

Una ulteriore considerazione riguarda gli attributi associati ad una linea.
Supponiamo, per esempio, di descrivere una rete stradale formata da entità elementari che rappresentano tratti di strada. Come ovvio, ad ogni tratto di strada sono associati degli attributi, per esempio il tipo di pavimentazione. Consideriamo il caso di voler asfaltare la parte centrale di un tratto di strada. In questo caso l'unica soluzione per mantenere la coerenza tra l'entità geografica e i suoi attributi è spezzare il tratto di strada in tre parti.
Questa situazione viene indicata come omogeneità degli attributi, cioè quando un attributo, tramite un suo valore, descrive una caratteristica di un'entità, il valore deve ritenersi valido per tutta l'entità. Se tale condizione non è più valida è necessario frazionare l'entità originale in altre entità.

La primitiva AREA è costituita da un insieme ordinato di punti e di attributi; come nel caso della linea, ogni coppia di punti contigui della sequenza identifica un segmento facente parte della primitiva; a differenza della linea, però, il punto iniziale e quello terminale coincidono.
La primitiva Area descrive quella parte di piano che si trova all'interno

dell'insieme dei segmenti formati dalle coppie di punti.
A differenza delle altre primitive non presenta problemi di modellazione, presenta però gli errori di discretizzazione e bisogna prestare attenzione all'omogeneità degli attributi.

Il termine PIXEL deriva da "picture element" e indica una parte di un'immagine che è stata divisa in modo regolare in piccoli elementi.

Questa tecnica di dividere un'immagine in una serie di elementi regolari (tassellazione) in modo da poter misurare quantità associate ai singoli elementi è abbastanza comune. Per esempio la tassellazione è usata nell'immagine televisiva, nel rilevamento per inventari forestali o per l'archeologia, nella rasterizzazione di immagini fotografiche e in molti altri casi. Come si vede, quindi, questa metodica è stata utilizzata anche in settori disciplinari prima dell'avvento delle tecnologie elettroniche e informatiche.

L'idea alla base del modello raster utilizzato per i dati territoriali è che la tassellazione occupa tutto lo spazio e ad ogni pixel è associato un valore. Nella logica del modello vettoriale, invece, esiste uno spazio vuoto in cui sono definiti alcuni oggetti.
In conclusione: punto, linea e area rappresentano oggetti; il pixel non rappresenta un oggetto, ma è il luogo dove una grandezza assume un certo valore.

Tuttavia, i valori della grandezza dipendono dagli oggetti presenti sul territorio, per cui i pixel, riportano in una logica di contesto e non di singolo tassello informazioni relative a tali oggetti.

Riportiamo qui di seguito un esempio che mostrerà come è possibile passare da un'immagine raster a dati vettoriali.

FIGURA 7 (vedi Appendice)

Immaginiamo che la figura rappresenti la zona a cavallo di una linea di costa, dove valori bassi di radianza rappresentano il mare, valori intermedi una zona boscosa e valori alti la spiaggia.
Supponiamo di avere le informazioni necessarie per dire che in questa zona e nelle condizioni in cui è stata acquisita l'immagine, i valori di radianza compresi tra 0 e 15 indichino "mare", quelli compresi tra 20 e 35 "pineta" e infine quelli tra 70 e 100 "spiaggia".
Siamo allora in grado di operare una classificazione dove ad ogni valore di radianza è stato sostituito un valore simbolico che indica mare (con 1), pineta (con 3) e spiaggia (con 2).

FIGURA 8 (vedi Appendice)

Dalla matrice ottenuta possiamo eseguire l'operazione di accorpamento, cioè due o più pixel adiacenti con lo stesso valore simbolico saranno tutti riconducibili allo stesso oggetto sul territorio; si potranno, in questo modo, distinguere (se si evidenziano con una linea) i bordi dell'oggetto che ha generato la prima immagine.
L'ultimo passo da compiere e l'eliminazione della quadrettatura, le linee rimaste descrivono gli oggetti presenti sul territorio in termini di aree, cioè secondo il modello vettoriale.

Entrambi i modelli, ovviamente, presentano dei VANTAGGI e degli SVANTAGGI.
Si può affermare che il mondo raster è più adatto a misurare grandezze che variano con continuità mentre con le primitive del modello vettoriale si rappresentano bene oggetti il cui carattere principale è la discontinuità al bordo.
Di conseguenza il mondo raster rappresenta meglio oggetti "naturali" e il mondo vector oggetto creati dall'uomo. Questo, però, non rappresenta un vincolo alla memorizzazione delle entità territoriali.
Il modello vector, in aggiunta, occupa meno spazio di memoria del modello raster ed è possibile costruire esplicitamente la topologia (vedi in seguito) ma

alcune operazioni risultano più complicate da eseguire, come per esempio l'overlay (vedi in seguito).
Il modello raster, invece, occupa molta memoria ed ha una topologia "implicita", ma le operazioni sono più semplici ed intuitive ed è possibile costruire un'algebra delle mappe (vedi in seguito).

Per ovviare al problema della memorizzazione del modello raster, però, sono stati sviluppati alcuni algoritmi che cercano di compattare le informazioni e quindi occupare meno spazio.
Per esempio nella codifica run-lenght viene sfruttato il fatto che vi possono essere sequenze di valori che si ripetono. Viene, allora, salvato in una tupla il valore numerico di intensità del pixel e il contatore che specifica il numero di pixel consecutivi simili.
Nel block encoding, invece, non si salvano parti di riga, ma blocchi rettangolari uniformi. Per ogni blocco si archivia l'origine (che può essere il centro o un vertice) e il lato.
Entrambe le codifiche sono mono-risoluzione, cioè non è possibile estrarre in modo efficiente dal dato una versione del raster a risoluzione più bassa. La struttura Quadtree, invece, è multi-risoluzione, perché si tratta di un algoritmo ricorsivo che divide l'immagine in quattro parti e si ferma (cioè passa alla memorizzazione) quando il valore in un blocco è uniforme (cioè i pixel contenuti nel quadrante hanno lo stesso valore) oppure quando raggiunge la risoluzione di partenza.

In realtà, al giorno d'oggi, non è più possibile una così netta distinzione poiché in molti software sono implementati entrambi i modelli senza che l'utente risenta eccessivamente delle differenze esistenti.

2. 2 TOPOLOGIA

Si è visto come oltre alle caratteristiche geometriche, le informazioni spaziali debbono anche gestire la topologia delle entità. Si ricorda che si può parlare di topologia solo quando si usano le primitive del modello vettoriale, con il modello raster le seguenti considerazioni non sono valide. Si veda, ora, attraverso una serie di esempi quanto sia importante questa componente.

Supponiamo di dover rappresentare due comuni adiacenti tramite la primitiva area. Non è necessariamente vero che questa adiacenza (mancanza di spazi vuoti o di aree sovrapposte) sia rispettata nel momento in cui vengono acquisite le due aree; infatti nella fase in cui viene creata l'informazione geografica numerica, le coordinate di un vertice di un'area, per quanto acquisite con grande precisione, non saranno esattamente uguali a quelle del corrispondente vertice dell'area adiacente.

Un altro problema sorge quando si rappresenta il bordo di un'area e di questo se ne vede soltanto una parte; senza ulteriori informazioni non abbiamo alcuna capacità di sapere se la parte interna dell'area si trovi a destra o a sinistra della linea.

Per cercare una soluzione analizziamo in profondità il concetto di area e di adiacenza.
Immaginiamo di rappresentare tre aree formate da un certo numero di segmenti collegati tramite vertici. Cerchiamo adesso di semplificare le linee, togliendo i punti ritenuti meno importanti.
Si può giungere (terza immagine) fino al punto in cui si verrebbe a modificare anche la parte strutturale della figura, in quanto, togliendo ulteriori punti, si perderebbero le relazioni di confine tra le aree, ovvero l'adiacenza.

FIGURA 9 (vedi Appendice)

In questo modo si nota che, in una struttura di aree, l'elemento geometrico di base non è la poligonale chiusa che circonda l'area, ma la linea che divide due aree, o ARCO, formato a sua volta da un numero imprecisato di segmenti. Due o più archi convergono in un punto detto nodo. Un'area, di conseguenza, può essere definita come un insieme di archi.
Se usiamo la nozione di arco, in aggiunta, vengono eliminate le incongruenze sui bordi, in quanto la linea che definisce un'area è la stessa linea che definisce l'area contigua.
Se consideriamo, inoltre, come attributi di una linea il "quale area c'è a destra?" e il "quale area c'è a sinistra?" possiamo risolvere il problema di non sapere a priori, da un tratto di bordo, se l'area racchiusa è a destra o a sinistra.

Un'area può essere codificata come un insieme di archi se sono applicate alcune regole: gli archi sono orientati e si è adottata una convenzione sul rapporto tra il verso della linea e il "pieno" dell'area (di solito si sceglie la regola che "il dentro è a destra", cioè camminando sulla linea nel suo verso troviamo l'area a destra). In questo modo è possibile costruire una tabella degli archi che associ ad ogni arco gli attributi Sinistra e Destra, cioè i codici delle aree che si trovano a sinistra e a destra. In aggiunta, prendendo come base la tabella degli archi è possibile costruire la tabella delle aree nella quale un'area è definita dalla somma algebrica degli archi che hanno come attributo il codice dell'area; si usa più se il codice compare nel campo Destra, meno se compare nel campo Sinistra.

<u>FIGURA 10</u> (vedi Appendice).

Attraverso la tabella delle aree è possibile fare un controllo di correttezza dei dati; perché ogni arco deve comparire nella tabella due volte, una volta per ciascuna area che separa.
Se sommiamo, infatti, tutti gli elementi della colonna composizione otteniamo:

-b +c +d +a -c –f -d +e +f = -b+a+e

Per portare tale somma a zero, come sarebbe corretto, dovremmo introdurre un'area **+b-a-e** che descrive tutto lo spazio all'esterno della figura.
Operando, invece, sulla tabella degli archi è possibile eseguire alcune operazioni:
CONTIGUITÀ: due aree sono contigue se esiste almeno un arco che ha nei campi Area-Sinistra e Area-Destra i codici delle due aree ;
UNIONE: l'area unione di due aree è quella formata dagli archi che hanno il codice di una sola delle due aree nei campi Area-Sinistra e Area-Destra; vanno esclusi sia gli archi che hanno nei campi Area-Sinistra e Area-Destra codici diversi, sia quelli che hanno nei campi Area-Sinistra e Area-Destra i codici delle due aree contemporaneamente.

Oltre ai problemi di incongruenza tra aree adiacenti esistono anche quelli che coinvolgono le linee connesse tra di loro. Immaginiamo, per esempio, una rete stradale, nella quale gli incroci risultano incongruenti, cioè non perfettamente chiusi.
Analogamente al caso delle aree ipotizziamo che la connessione tra linee diverse (dette archi) avviene attraverso i nodi.
E' definibile, quindi, una tabella di nodi, dove ogni nodo è associato agli archi che convergono in esso. A differenza della tabella degli archi, però, nella tabella dei nodi ad ogni nodo sono associati archi in numero indefinito.

Dagli esempi trattati si è giunti a capire che la topologia studia le relazioni spaziali tra oggetti geometrici e, nella tecnologia GIS, è una sovrastruttura ridondante che però dà maggiore efficienza nell'elaborazione e garantisce una certa coerenza dei dati.
Tale struttura, però, costa nella preparazione dei dati, nell'importazione dei dati da altre fonti e nell'aggiornamento.

2.3 DBMS E QUERY

Si è detto che un GIS integra al suo interno la tecnologia dei DBMS (DataBase Management System) per la gestione delle informazioni geografiche.
Cerchiamo, qui di seguito, di capire a cosa ci stiamo riferendo.

In una accezione generica, possiamo definire una base di dati (database) nel seguente modo:
"collezione di dati, utilizzati per rappresentare le informazioni di interesse per una o più applicazioni di una organizzazione";
mentre in una accezione più specifica il database è una "collezione di dati gestita da un DBMS".
Un DBMS (DataBase Management System) è:
"un sistema (prodotto software) in grado di gestire collezioni di dati che siano (anche):

- grandi (di dimensioni (molto) maggiori della memoria centrale dei sitemi di calcolo utilizzati);
- persistenti (con un periodo di vita indipendente dalle singole esecuzioni dei programmi che le utilizzano);
- condivise (utilizzate da applicazioni diverse);

garantendo affidabilità (resistenza a malfunzionamenti hardware e software) e privatezza (con una disciplina e un controllo degli accessi). Come ogni prodotto informatico, un DBMS deve essere efficiente (utilizzando al meglio le risorse di spazio e tempo del sistema) ed efficace (rendendo produttive le attività dei suoi utilizzatori)" [Atzeni-Cerri-Paraboschi-Torlone,2000].

Per rendere efficace ed efficiente un database è necessario operare un'attenta progettazione. Negli anni si sono sviluppate diverse tecniche che attraverso alcune linee guida aiutano il progettista in questa complessa fase. In particolare si cerca di sviluppare nell'ordine, prima un modello concettuale, poi un modello logico ed infine un modello fisico dei dati che dovranno essere gestiti dal software. I modelli si distinguono per un diverso livello di astrazione: il

modello concettuale permette di costruire degli schemi molto vicini all'interpretazione naturale dell'uomo, mentre il modello fisico è molto più legato all'organizzazione del computer.
Il modello concettuale più conosciuto è quello entità-relazione (entity-relationship, E-R), mentre il modello logico può essere gerarchico, reticolare, relazionale oppure ad oggetti.
Generalmente in un GIS viene implementato, per la sua estrema flessibilità, un modello relazionale.

Per estrarre informazioni da un sistema informativo si può eseguire l'operazione di query, cioè di interrogazione della base dati. Il risultato della query è un insieme di dati oppure una mappa che possono essere visualizzati attraverso l'operazione display.
La struttura base di una query è formata da tre componenti fondamentali:

- l'entità oggetto della selezione;
- lo strato informativo all'interno del quale si effettua la selezione (Uno strato informativo o layer è un insieme geometricamente e semanticamente uniforme di dati geografici);
- le condizioni che consentono di effettuare la selezione;

Tale struttura si può tradurre attraverso lo standard SQL, nel seguente modo:

SELECT (seleziona) <entità>
FROM (da) <strato informativo>
WHERE (dove) <condizioni>

Generalmente, però, l'utente non è obbligato alla conoscenza del linguaggio poiché vengono predisposte opportune interfacce user-friendly.
Nei sistemi informativi territoriali si possono avere diversi tipi di query: aspaziali, spaziali, oppure miste.

Ogni informazione territoriale, come visto precedentemente, è formata dalla componente spaziale e da quella non spaziale, detta attributi; tale parte è interrogata nella query aspaziale.

Il risultato di dislay, però, può essere comunque visualizzato su una mappa e non solo come dato alfanumerico.

Nella query spaziale, invece, le condizioni di interrogazione coinvolgono la componente spaziale e comportano l'utilizzo di proprietà e relazioni spaziali. Nelle condizioni della query bisogna specificare gli operatori spaziali, le entità di riferimento e il valore delle proprietà. In queste query, oltre ad una interfaccia user-friendly che aiuta nella costruzione dell'interrogazione si lavora avendo sempre a video la mappa delle entità dello strato informativo di interesse. La selezione, perciò, può avvenire in quattro differenti modi: selezione attraverso puntatore, sulla base della distanza, di elementi che contengono o che sono contenuti.

Nel primo caso la selezione viene effettuata attraverso il puntatore e l'oggetto interessato viene evidenziato; in questo modo si può subito vedere se si è eseguita l'operazione corretta, perché se il puntatore ha dimensioni più grandi dell'oggetto, potrebbero venire evidenziate anche delle entità non di interesse. Nella selezione sulla base della distanza vengono evidenziati elementi (puntuali, lineari o areali) entro oppure oltre una certa distanza rispetto ad un elemento dato o tracciato (entità spaziale di riferimento). In questa query è importante definire i termini entro ed oltre per non creare ambiguità; infatti viene aggiunta nella condizione: completamente, parzialmente (e quanto) con contatto o senza contatto. Nella selezione sulla base di elementi che contengono si parte da un elemento dato (punto, linea o area) e si identifica l'elemento areale che lo contiene, mentre nella selezione sulla base di elementi contenuti si evidenziano le entità contenute in un elemento lineare o areale.

Nella query mista, infine, le condizioni riguardano sia la componente spaziale che quella degli attributi. Questa selezione può essere il risultato di una serie di query in successione, cioè una serie di interrogazioni nelle quali il risultato di una query funge da strato informativo per la query seguente.

Tutti questi tipi di interrogazione possono essere implementati anche nel modello raster, prestando, però, attenzioni diverse.

2.4 OVERLAY

L'operazione di overlay (traducibile in italiano con incrocio) è una delle funzionalità più importanti nei GIS perché attraverso essa si rende esplicito il carattere di supporto alle decisioni di questi particolari strumenti software. In particolare, attraverso questa funzionalità, viene esaltato al massimo il carattere di integratore di dati eterogenei dei SIT.
Attraverso l'operazione di incrocio, infatti, si possono integrare le informazioni presenti in strati informativi diversi sfruttando sia la parte geometrica che gli attributi di ciascun strato. Il risultato sarà un nuovo strato informativo nel quale ogni elemento eredita tutte le caratteristiche dei precedenti. In questo modo le informazioni presenti in diversi strati vengono integrate, trasformate e presentate in maniera diversa e possibilmente più leggibile.

Gli strati informativi possono essere rappresentati sia con il modello vettoriale che con quello raster.
Quando si lavora con i pixel l'operazione risulta sempre possibile, salvo capire il significato del risultato, mentre lavorando con le primitive punto, linea, area, non tutti i possibili incroci sono opportuni (per esempio linee con punti o punti con punti) oppure bisogna prestare attenzione alla trattazione degli attributi.

Si riporta qui di seguito l'esempio di overlay più classico e forse più usato che coinvolge le aree.
Supponiamo di avere due strati informativi di una piccola isola: uno descrive la copertura del suolo; l'altro la suddivisione delle proprietà.

FIGURA 11 (vedi Appendice)
FIGURA 12 (vedi Appendice)

Siamo interessati a conoscere a chi appartengono i vari tipi di area e viceversa che tipo di suolo è posseduto dai vari proprietari.
Eseguiamo allora l'operazione di overlay e otteniamo un nuovo strato informa-

tivo composto dalla parte geometrica e dagli attributi. In particolare otterremo un insieme di aree elementari, ognuna delle quali avrà un solo proprietario e sarà di un unico tipo di suolo.

FIGURA 13 (vedi Appendice)

Da un punto di vista concettuale l'operazione geometrica consiste nel spezzare gli archi nei punti dove si incrociano e quindi di ricostruire le nuove aree coi frammenti.
Per quanto riguarda gli attributi, ciascuna nuova area eredita quelli che quella parte di piano possedeva nei due strati informativi di partenza.

Si noti che l'operazione di overlay è reversibile, perché basta eliminare dal risultato il contributo di uno degli strati informativi di partenza, per ottenere l'altro.

Purtroppo, nell'operazione di overlay si può incorre in alcuni inconvenienti che è bene analizzare per capire come superarli attraverso predisposizione di procedure automatiche.
Da un punto di vista geometrico l'operazione di incrocio genera nel risultato un numero di aree non inferiore a quello degli strati di partenza. Non tutte le aree risultate, però, sono significative, perché si possono incontrare linee con geometria diversa che seguono uno stesso percorso, oppure che sono state digitalizzare in momenti diversi. Per esempio, l'asse di un fiume può rappresentare un limite amministrativo oppure può comparire in uno strato come idrografia. Per eliminare tali poligoni si ricorre al "fattore di forma" che mette in relazione superficie e perimetro: più esso è basso, più è probabile che l'area derivi da errori di digitalizzazione.
Nella maggior parte dei casi i bordi dei due strati informativi non coincidono e nel risultato dell'operazione si possono creare delle aree che non possiedono tutti gli attributi.

FIGURA 14 (vedi Appendice)

Per risolvere questi problemi è possibile eseguire tre diversi tipi di incroci:

- Incrocio per unione: si considerano tutte le aree generate dall'intersezione dei due strati informativi. Con riferimento alla figura si avranno in totale 9 aree di cui 4 con due attributi, 2 con gli attributi dello strato circolare e 3 con gli attributi delle strato rettangolare.
- Incrocio per intersezione: si considerano solo le aree che hanno tutti gli attributi. Nell'esempio avremo come risultato solo 4 aree.
- Incrocio sotto condizione: si considerano solo le aree dove sono presenti gli attributi provenienti da uno specifico strato. Se, nell'esempio prendiamo come strato condizionante quello circolare otteniamo come risultato 6 aree.

In una operazione di incrocio bisogna prestare attenzione al tipo di attributi che vengono coinvolti, perché gli attributi qualitativi mantengono il proprio valore, quelli quantitativi no, e occorrerà procedere a successive elaborazioni.
La distinzione tra le due categorie è abbastanza intuitiva, gli attributi qualitativi valgono per tutta la superficie senza dipendere da essa, e mantengono il loro significato in qualunque punto; mentre quelli quantitativi dipendono dalla superficie.
Per avere una classificazione più rigorosa occorre verificare "cosa succede se si spezza l'elemento in due parti e si associa lo stesso valore dell'attributo alle primitive risultato":

- Attributi qualitativi: mantengono la coerenza (per esempio l'attributo "Morfologia");
- Attributi quantitativi: perdono la loro coerenza (per esempio l'attributo "Numero di Abitanti");
- Attributi specifici: mantengono una certa coerenza, ma probabilmente perdono di validità (per esempio l'attributo "Densità di Popolazione").

Ovviamente, quanto detto si applica alle primitive area e punto (le uniche che possono essere spezzate).

2.5 BUFFER

Un'altra funzionalità dei GIS ampiamente usata e molto interessante è il buffering. Questa operazione permette di costruire un'entità areale creata sulla base di una distanza generalmente costante rispetto ad un'entità di riferimento che può essere puntuale, lineare o areale.

Per creare un buffer di solito si eseguono in sequenza le seguenti operazioni:

- selezione della/e entità di riferimento (eseguibile attraverso una query);
- scelta del tipo di buffer (su un lato, su due lati, esterno, in base ad una distanza costante o meno);
- assegnazione della distanza rispetto all'entità di riferimento, in corrispondenza della quale si troverà il limite del buffer.

Tale distanza può essere definita in tre diversi modi:

- diretto: si assegna un valore numerico;
- in funzione di una grandezza: il valore della distanza è espresso come funzione di una grandezza (per esempio l'attenuazione di onde elettromagnetiche nel caso di emettitori) oppure di un attributo dell'entità (per esempio la larghezza della strada);
- indiretto: il valore della distanza si ottiene attraverso un'operazione spaziale rispetto all'entità di riferimento (per esempio si vuole creare un vincolo di costruzione attorno alla linea di presunta alluvione di un fiume).

Il buffering è usato sia per creare zone di rispetto sia nelle analisi spaziali. Per esempio consideriamo la necessità di conoscere la distribuzione dei terreni coltivati entro 50 metri da un fiume. Per prima cosa creeremo un buffer attorno al fiume e poi eseguiremo l'overlay con lo strato delle aree coltivate. Potremmo, inoltre, essere interessati ad eseguire un'analisi di prossimità, cioè identificare le entità “vicine” ad altre. Tale vicinanza viene definita in base alla posizione, alla distanza o in base agli attributi. Per quanto riguarda i primi due modi di definizione il buffer è lo strumento adatto, mentre il terzo modo non tratta una vicinanza spaziale.

2.6 MAP ALGEBRA

Con il termine map algebra si intende l'uso di operatori elementari in sequenza allo scopo di risolvere problemi spaziali complessi, similmente a quanto avviene in un'espressione algebrica in cui la combinazione di grandezze fornisce un risultato.

Le maggiori potenzialità di calcolo si hanno in ambiente raster, ma è possibile lavorare anche con le primitive vector. La predilezione per il modello raster può essere spiegata dalle caratteristiche dei pixel: si tratta di un insieme omogeneo di celle cui sono attribuiti dei valori ben definiti e che può essere rappresentato come una matrice su cui applicare degli algoritmi.
L'unico inconveniente è dovuto al fatto che i set di dati (layer di input) devono essere delle stesse dimensioni, sia in termini di dominio che di risoluzione.

FIGURA 15 (vedi Appendice)

Gli operatori che si usano nella map algebra sono di quattro tipi: operatori locali, zonali, focali e globali. Ci sono, inoltre, gli operatori di utilità.

Gli operatori locali consentono di attribuire ad ogni elemento del layer-risultato un valore funzione dei valori dei corrispondenti elementi dei layer di input.
Un esempio di tale uso può essere trovato nella figura precedente.

Con gli operatori zonali è possibile attribuire alle celle dello strato-risultato funzioni di valori di celle di un layer-input raggruppate secondo zone definite in un altro strato di input.

FIGURA 16 (vedi Appendice)

Oltre alla somma è frequente usare questo operatore per il calcolo della media, in cui a tutte le celle di una zona viene attribuita la media dei valori delle singole celle oppure il calcolo del minimo e massimo.

Con l'operatore focale il valore degli elementi del layer-risultato è funzione dei valori di elementi appartenenti ad un intorno dell'elemento considerato.
A differenza dell'operatore zonale, quindi, il valore calcolato viene attribuito ad una sola cella e non ad una intera zona.

FIGURA 17 (vedi Appendice)

Gli operatori globali vengono utilizzati per determinare layer di output in cui il valore delle celle dipende potenzialmente dai valori di tutti gli elementi dello strato input.

Gli operatori di utilità, infine, sono indispensabili perché permettono di predisporre le mappe per eseguire la map algebra, cioè trasformano i layer di input in modo che abbiano il medesimo formato e le stesse dimensioni.

CAPITOLO 2
GEOMARKETING E STATISTICA

1. GEOMARKETING

In una continua evoluzione degli studi sul marketing e su "come fare impresa" nasce, recentemente, il filone del geomarketing.

Si propongono, qui di seguito, una serie di DEFINIZIONI che aiutano ad inquadrare questa nuova disciplina.

- *"Il geomarketing può essere definito come il complesso di applicazioni di marketing condotte su una serie di dati ottenuti partendo da basi geografiche."* [S.D.A. Bocconi – Area Marketing]
- *"Il geomarketing è la gestione dei dati tramite tecnologia GIS atta a fornire su base georeferenziata informazione sul mercato attuale e potenziale, sul posizionamento aziendale, sui processi di vendita e sui servizi aziendali al cliente etc..."* [Sito Web]
- *"Il geomarketing è un insieme di tecniche di trattamento delle informazioni che permettono di evidenziare la dimensione geografica dei fenomeni sociologici, economici, politici, di marketing e commerciali."* [TT GOUP]
- *"Il geomarketing? E' la coscienza geografica del business!"* [R. Marin – Geko Consulting, Padova]
- *"Il geomarketing è l'utilizzo della geografia per rendere più efficaci ed efficienti le decisioni e le attività relative alla comunicazione, alla vendita, alla distribuzione ed al servizio ai clienti."* [M. Di Dio Roccazzella -Value Lab Srl]

Nella letteratura tradizionale i problemi che deve affrontare un *operatore di marketing* si possono sintetizzare nella formula: "quanto produrre ed a che prezzo e come commercializzare e distribuire il prodotto"; in realtà interessa sapere "Chi + fa cosa + dove". In particolare si può pensare che un'impresa si interroghi su domande quali, ad esempio, dove produrre? quanto produrre in ogni località? che prezzo praticare in differenti mercati? come comunicare con segmenti geografici di clientela diversi?

In conclusione i PROBLEMI che un *operatore di geomarketing* generalmente è tenuto a risolvere si possono sintetizzare in:

- individuare i confini dei mercati locali (il marketing tradizionale considera la domanda come un tutto unico mentre in realtà esiste un insieme di mercati geograficamente delimitati sebbene intercomunicanti);
- stimare i potenziali e le quote di mercato locali (il geomarketing necessita delle stime di vendita nei mercati locali per poter formulare gamme di offerta, cioè quantità produttive e prezzo, differenziate territorialmente);
- localizzare la rete distributiva in maniera "ottima" (attraverso lo studio delle distanze e dell'analisi della rete viaria e del traffico si ricerca la localizzazione ottima);
- sfruttare campioni areali per le ricerche di mercato (applicazione di tecniche che tengano conto della componente territorio e che permettano di ottenere stime più efficienti e cluster spaziali);
- costruire sistemi di *what if* geografico (studio delle conseguenze delle scelte strategiche con evidenza dei diversi impatti in località diverse).

Per attuare un adeguato approccio al geomarketing un'azienda necessita di tre INGREDIENTI:

- informazioni statistiche con livello di disaggregazione territoriale idoneo agli scopi di analisi;
- metodologia affidabile di statistica spaziale;
- software GIS.

2. DISPONIBILITÀ DI DATI STATISTICI

Al giorno d'oggi in azienda si raccolgono in ogni momento enormi quantità di dati, di cui l'80% può essere georiferito. Purtroppo, molte di queste fonti di informazione normalmente non sono predisposte per essere sfruttate appieno nel processo decisionale. In aggiunta, fino a non molti anni fa, non era possibile trovare dati statistici a livello di dettaglio territoriale voluto. Fortunatamente, la situazione si sta modificando ed è possibile trovare, accanto alle fonti ufficiali tradizionalmente disponibili (censimenti, materiale cartografico derivante da archivi militari) ed alle indagini ad hoc, aziende che vendono informazioni spaziali.
Ricordiamo, però, che se le informazioni territoriali tematiche non sono adeguatamente legate alla componente spaziale viene inficiato il loro possibile uso.

In Italia ci sono diversi Enti pubblici che producono sia informazione di base che tematica:

- ISTAT: attraverso una attenta programmazione dell'attività, svolta da pool di esperti appartenenti a diverse discipline, l'Istituto produce sia dati socio-economici, sia cartografia di base;
- REGIONI, PROVINCIE, COMUNI: per obblighi di legge sono incaricati a mantenere una carta tecnica regionale e in aggiunta raccolgono dati relativi all'uso del suolo, alla geologia, ecc...;
- MINISTERO DELLA FINANZA: tale istituto mantiene costantemente aggiornato l'archivio catastale, che pur essendo classificabile come informazione tematica è usato molto spesso anche per l'informazione di base.

In forte espansione, inoltre, è il settore delle imprese private che attraverso consulenze aiutano le aziende ad implementare un sistema che gestisce i dati interni ed li integra con i dati cartografici digitali che permettono la strutturazione dell'informazione in senso geografico.
Arbia [Verso un approccio statistico al geomarketing: alcune linee guida ed un

caso studio, 1995] propone una classificazione delle fonti di dati spaziali in tre categorie omogenee dal punto di vista dell'input in un GIS:

1. censimenti e dati amministrativi o contabili i quali forniscono una copertura completa;
2. indagini campionarie che forniscono una copertura non completa;
3. immagini aeree o da satellite le quali forniscono un'immagine completa.

Si ricorda però che una mappa contenente informazione spaziale non è esente da ERRORI. Essa infatti è affetta da:

- errori comuni alle rilevazioni statistiche (errori di campionamento, oppure dovuti all'imperizia dell'intervistatore, ...);
- errori tipici dei dati territoriali (errori dovuti alla localizzazione geografica);
- errori dovuti al metodo di immagazzinamento (errori di digitalizzazione).

Sempre Arbia [Impiego dei sitemi geografici informativi nelle rilevazioni ufficiali, 1990] propone un modello per l'analisi degli errori spaziali in un GIS:

$$\boldsymbol{Z = H\,S + \mu}$$

dove Z è il vettore degli n valori osservati, S il vettore degli n valori "veri", H è una matrice n*n tale che $\sum_j h_{ij} = 1$ ed $h_{ij} \geq 0$ e μ un vettore di errori.

Possiamo riscrivere il modello di prima esplicitando i vettori e la matrice:

$$\begin{bmatrix} Z_1 \\ \vdots \\ Z_n \end{bmatrix} = H \begin{bmatrix} S_1 \\ \vdots \\ S_n \end{bmatrix} + \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix}$$

In particolare, per ogni i = 1,...,n si ha

$$Z_i = \begin{bmatrix} h_1^i & \cdots & h_n^i \end{bmatrix} \begin{bmatrix} S_1 \\ \cdots \\ S_n \end{bmatrix} + \mu_i \quad \text{e} \quad Z_i = h_1^i S_1 + \cdots + h_n^i S_n + \mu_i \,.$$

Riscrivendo i termini in modo diverso possiamo ottenere la seguente formula

$$Z_i = h_i^i S_i + \sum_{i \neq j} h_j^i S_j + \mu_i$$

$$Z_i = h_i^i S_i + \sum_{i \neq j} h_j^i S_j + \mu_i + S_i - S_i \quad (aggiungo\ e\ tolgo\ S_i)$$

$$\underbrace{(Z_i - S_i)}_{errore\ totale} = \underbrace{(h_i^i - 1) S_i + \sum_{i \neq j} h_j^i S_j}_{errore\ di\ posizione} + \underbrace{\mu_i}_{errore\ di\ misura}$$

- Nel caso di censimenti si avrà H = 1 in quanto i valori sono correttamente assegnati e non esiste errore di posizione, mentre u~N (μ, D) con μ errore sistematico e D matrice diagonale che incorpora la possibilità di errori eteroschedastici proporzionali alla dimensione delle aree.
- Nel caso di indagini campionarie H = 1 e u~N (μ , V) con μ errore sistematico e V debolmente non diagonale per la presenza di correlazione spaziale.
- Nel caso di immagini da satellite si ha H = 1 e u~N (μ , V). In questo caso l'errore di posizione è generato dall'incertezza che si ha su quale pixel debba riferirsi una certa osservazione, in quanto le superfici rilevate sono continue e gli strumenti possono sbagliare a rilevare la rifrazione.

3. DISPONIBILITÀ DI UNA METODOLOGIA STATISTICA ADEGUATA

La letteratura statistica riguardante dati spazialmente distribuiti ha avuto origine circa cinquant'anni fa ed in principio si è ispirata all'analisi delle serie storiche e dell'inferenza da processi stocastici temporali. Poi, a partire dalla seconda metà degli anni '80, la statistica spaziale ha suscitato un interesse crescente, dovuto anche alla diffusione degli elaboratori elettronici i quali hanno reso disponibile una ricca serie di informazioni e nuove capacità di calcolo.

Qui nel seguito vengono elencati alcuni dei problemi legati alla statistica spaziale.

3.1 DEFINIZIONE DI DISEGNI CAMPIONARI DI DATI SPAZIALI

Tradizionalmente per le indagini campionarie relative a fenomeni economici si usa il campionamento a più stadi. Il disegno campionario tipico di questa tecnica prevede una prima estrazione seguendo il criterio casuale semplice di un insieme di unità primarie dal territorio suddiviso in forma esaustiva in aree di varia dimensione e forma ed una seconda estrazione delle unità finali di rilevazione dalle zone geografiche campionate.
In tale disegno campionario si trascura il fatto che si lavora con dati che non sono indipendenti come le biglie in un'urna (esempio principe per spiegare il campionamento casuale semplice) ma presentano un certo grado di similarità connesso con la prossimità geografica, presentano, cioè, una correlazione spaziale; Cohrane, nel 1963, affermò: "*adjancent units are often more alike than units that are far apart*".
Si ricercano, allora, tecniche che producono stime più efficienti, perché, un campionamento casuale semplice può includere unità adiacenti che, di conseguenza, forniscono solo un piccolo contributo aggiuntivo di informazione.

In maniera più formale possiamo scrivere:

- $\{X\}=\{X_{1,}\cdots, X_N\}$ processo stazionario costituito da N variabili aleatorie distribuite su di un territorio suddiviso in N località geografiche $\{s_{1,}\cdots, s_N\}$;
- $E(X_i)=\mu \quad \forall i$;
- $Var(X_i)=\sigma^2 \quad \forall i$;
- $Corr(X_i, X_j)=\begin{cases}\beta>0 & se\, s_i\, è\, vicino\, a\, s_j \\ 0 & altrimenti\end{cases} \qquad \forall i\, e\, j$;

- $\{x_{1,}\cdots, x_n\}, n<N$ campione con media campionaria $m=\frac{\sum_i x_i}{n}$;
- $E(m)=\mu$;
- $V(m)=\frac{n+\sum_i\sum_j \beta}{n^2\sigma^2}=\frac{n+h\beta}{n^2\sigma^2}$ $\quad h=numero\ di\ adiacenze\ tra\ x_i$.

Dall'ultima formula si vede come la varianza campionari delle stime dipende dall'allocazione delle unità campionarie.

Per superare tale problema, Arbia nel 1990, propose la tecnica DUST (*Dependent Areal Units Sequential Techique*) che si fonda sulle seguenti idee chiave:

- attraverso un campionamento si vuole stimare un incognito parametro caratterizzante la distribuzione di una variabile *X* osservabile su un'area suddivisa in modo esaustivo in N segmenti geografici non sovrapponibili;
- si dispone di una seconda variabile *Y* rilevata sui medesimi N segmenti e correlata con la X.

Il campionamento DUST si può esprimere attraverso i passi della seguente procedura:

- Si stima $_Y\hat{\beta}=\frac{N\sum_i\sum_{j\in N(i)}(y_i-\bar{y})(y_j-\bar{y})}{\sum_i(y_i-\bar{y})^2(\sum_i h-i)}$

  dove N(i) corrisponde all'insieme dei vicini dell'area i-esima;

  $\bar{y}$ è la media delle y_i e h_i è il numero dei vicini dell'area i-esima.

  L'ipotesi base è $_Y\hat{\beta}=_X\hat{\beta}=\beta$.
- Si estrae in modo casuale la prima unità campione s_1.
- Si classificano le rimanenti (N-1) unità a seconda del grado di prossimità rispetto a s_1.

Si estrae s_2 assegnando come pesi $(1-\beta^{d_{1j}})$ dove d_{1j} = 0,1,...,D indica l'ordine di prossimità tra s_1 e il generico s_j e D è il massimo ordine di prossimità realizzabile nello spazio considerato.

- Si estrae s_3 assegnando alle rimanenti (N-2) unità i pesi $(1-\beta^{d_{1j}})(1-\beta^{d_{2j}})$, dove d_{1j} rappresenta l'ordine di prossimità tra s_1 e s_j e d_{2j} la vicinanza tra s_2 e s_j.
- Il processo è iterato n volte e l'ultima unità campionaria è estratta con peso $\prod_{j=1}^{h-1}(1-\beta^{d_{nj}})$.

Da tale disegno campionario si capisce che ogni unità ha una probabilità di inclusione che cresce al crescere della distanza con le unità già campionate e decresce con il valore di β il quale misura il grado di similarità tra unità.

La procedura descritta è estremamente pesante da un punto di vista computazionale, ma questo non deve diventare un ostacolo alla sua realizzazione visto che oggigiorno sono disponibili potenti elaboratori dotati di software adeguati.

Per dimostrare che la tecnica DUST è migliore dal campionamento casuale semplice si sono eseguiti alcuni studi Monte Carlo.
Per esempio, utilizzando i dati relativi al censimento del 1981 nei 649 distretti del comune di Bologna, si sono estratti 1000 campioni di dimensione 10 usando entrambi i disegni campionari e si è dimostrato che la tecnica ideata da Arbia porta ad un miglioramento del 15% in termini di errore quadratico medio.
In aggiunta la tecnica DUST si è rivelata più efficiente del campionamento casuale semplice anche sostituendo ${}_x\hat{\beta}$ con una sua stima distorta purché dello stesso segno.
Infine, da 1000 griglie quadrate 10×10 sulle quali sono stati inseriti dati generati attraverso un modello a media mobile, si sono estratti 1000 campioni con numerosità variabile dal 2% a 90% usando ambo le tecniche. L'analisi dei dati ha mostrato che DUST fornisce stime più efficienti per frazioni

campionate variabili dal 2% al 75%, in quanto per frazioni superiori diventa sempre più difficile estrarre unità distanti da quelle già selezionate. Si è dimostrato, inoltre, che il guadagno in efficienza è maggiore dove la correlazione spaziale è più forte.

Ulteriori studi hanno evidenziato possibili varianti alla tecnica DUST base.
Ridolfi, infatti, ha proposto che per l'estrazione della seconda unità si considerino i pesi $\begin{cases} (1-\beta)\, p_i & se\, i\, è\, un\, vicino\, di\, j \\ p_i & altrimenti \end{cases}$ con p_i la generica probabilità di estrazione dell'i-esimo elemento campionario.
In pratica, dopo la prima estrazione, si suddivide la popolazione in due sottogruppi: il primo, composto dai d vicini all'area estratta ha probabilità di essere estratto diminuita di β%; il secondo composto da (N-d-1) vengono estratti con probabilità più elevata. A seguito della seconda estrazione, poi, nel gruppo dei "vicini" si considerano anche le unità poco distanti da s_2.
Anche questa variante ha dimostrato una maggiore efficienza rispetto al campionamento casuale.

3.2 INTERPOLAZIONE AREALE PER LA STIMA DI INFORMAZIONI A LIVELLO DI DISAGGREGAZIONE INFERIORE DI QUELLO DISPONIBILE

Per elaborare una appropriata tecnica di marketing rivolto al territorio diventa fondamentale poter disporre di informazioni relative alla domanda di mercato e alla sua futura evoluzioni in ogni particolare localizzazione geografica di interesse per l'azienda; perché, attraverso tale conoscenza è possibile delineare le zone di vendita ed i traguardi di fatturato, programmare i lanci di nuovi prodotti, di offerte promozionali e di campagne pubblicitarie; in conclusione controllare i risultati delle strategie e programmare azioni future.
Tale necessità conoscitiva si scontra con una carenza di dati disaggregati a livello territoriale idoneo, in quanto le statistiche disponibili sono, purtroppo,

partizionate sulla base delle unità amministrative (comuni, provincie, regioni,...), che molto spesso non rappresentano in modo adeguato l'omogeneità socio-economica di alcune comunità.

Si sviluppa, allora, nelle aziende la credenza che solo con una rilevazione diretta un dato può considerarsi ufficiale; ma qualora tale rilevazione non sia possibile o conveniente è possibile generare conoscenza attraverso tecniche di stima, che in relazione a questo particolare ambiente vengono denominate *small area estimator*.

Tali tecniche, secondo Benedetti, si possono suddividere in tre gruppi:

- metodi basati sulla teoria dei campioni classica;
- approccio top-down;
- approccio bottom-up.

Di seguito si cerca di spiegare sinteticamente le tre metodologie.

3.2.1 METODI BASATI SULLA TEORIA DEI CAMPIONI CLASSICA

Generalmente, con questa tecnica, si considera un campione casuale semplice di dimensione n estratto dalla popolazione P_T (che insiste sul territorio T) di dimensione N.

Lo scopo dell'analisi è, allora, stimare un parametro relativo alla popolazione. Si suppone, per esempio, che il parametro oggetto di stima sia il totale a livello di piccolo dominio $Y_i = \sum_{j=1}^{N_i} y_{ij}$, dove N_i è la dimensione ipotizzata nota della i-esima zona geografica (SA, *small area*), y_{ij} è il valore assunto dalla variabile di interesse nella j-esima unità afferente all' i-esima SA.

Con queste premesse sono disponibili due diversi metodi di stima:

- Metodi di SAE diretti: usano sole le unità campionarie appartenenti all'area di interesse. Producono, perciò, valori elevati di varianza dovuti alla bassa frequenza di osservazioni.

  Per esempio lo stimatore diretto del totale Y_i relativo all'i-esimo piccolo

dominio è: ${}_d\hat{Y}_i = \frac{N}{n}\sum_{J=1}^{n_i} y_{ij}$, dove $n_i > 1$ è l'insieme delle unità che cadono nell' *i*-esima SA.

- Metodi di SAE indiretti: utilizzano sia i dati relativi alla variabile di interesse sia a variabili ausiliarie appartenenti ad un'area che contiene il dominio di interesse e/o desumibili da altre occasioni di indagine oltre a quella corrente. Tali metodi, a loro volta, si basano su modelli impliciti che producono stimatori sintetici o su modelli espliciti.

  Per esempio uno stimatore sintetico può essere: $\hat{Y}_i = \frac{\bar{y}}{\bar{x}} X_i$, dove $\bar{y}$ e $\bar{x}$ sono le medie campionarie nell' *i*-esimo piccolo dominio; mentre un modello esplicito può avere la forma: $y_{ij} = \beta x_{ij} + \epsilon_{ij}\sqrt{x_{ij}}$, $j = 1,2,\cdots,N$, $i = 1,2,\cdots,I$, dove ϵ_{ij} sono errori iid, I è il numero di piccoli domini e x_{ij} è il valore assunto dalla variabile ausiliaria x per la *j*-esima unità nell'area *i*.

3.2.2 L'APPROCCIO TOP-DOWN

In sintesi questo approccio prevede di disaggregare dati noti ad un livello di risoluzione inferiore a quello di interesse per l'analisi.

I dati disponibili si riferiscono alla variabile di interesse Y_i e sono noti per un insieme di aree che costituiscono la partizione sorgente *S* ma incogniti a livello delle zone che compongono la partizione obiettivo *T*, dove *S* e *T* insistono sullo stesso dominio geografico.

Una delle prime procedure usate prende in input una data partizione in aree ed individua un punto significativo per ogni unità. Successivamente si sovrappone una griglia regolare e si stimano i valori dei nodi. Infine, mediando i valori della griglia con la ripartizione più fine si ottengono le stime desiderate.

I problemi di tale procedura sono molteplici in quanto il dato spaziale è legato alle caratteristiche geometriche e strutturali dell'oggetto associato e quindi si possono produrre stime poco accurate.

La crescente disponibilità di informazione e l'evoluzione dei sistemi

informativi geografici in grado di gestire tali data base ha permesso di usufruire di dati ausiliari circa la distribuzione nella partizione sorgente di variabili collegate in qualche modo alla variabile oggetto di studio; e si sono così potuti sviluppare nuovi algoritmi di calcolo.

Flowerdew, per esempio, considera la stima dei valori y_{st} per ogni SA intersezione tra le aree $s \in S$ e $t \in T$ come un problema di dati mancanti da risolvere con l'algoritmo EM.

Nella fase E (expectation) si calcola il valore atteso dei dati mancanti condizionato al modello selezionato ed ai dati osservati. Nella fase M (Maximum likelihood) si adatta il modello scelto alle stime provenienti dalla prima fase, alle variabili ausiliarie e alle superfici delle SA.

Anche questo metodo, però, ha dei problemi perché bisogna trovare delle idonee variabili ausiliari e si possono adattare solo i modelli Poisson e binomiale.

Un metodo più sofisticato e non dipendente da ipotesi così restrittive come il precedente, e il metodo pycnophylactica (pycnophylactico indica una tecnica di stima di superfici di densità che preservano la densità stessa) proposto da Tobler. In questo caso si procede alla costruzione di una superficie direttamente a partire dai dati areali disponibili e la stima dei nuovi valori areali si ha tramite integrazione e rispetto di alcune condizioni.

Benedetti e Palma propongono, invece un metodo denominato BIM (Bayesian Interpolation Method). In tale soluzione si evitano trasformazioni intermedie delle variabili e si pone attenzione ai dati ausiliari relativi alla partizione obiettivo essendo usati come informazioni a priori; in questo modo si supera il limite imposto dall'ipotesi di variazione regolare dei fenomeni sul territorio.

3.3.3 L'APPROCCIO BOTTOM-UP

L'approccio bottom-up, puramente modellistico, che si concretizza in una stima indiretta di dati individuali (livello di risoluzione massimo) da aggregare al livello di dettaglio territoriale desiderato.

L'idea alla base dell'approccio bottom- up e quella di legare attraverso stima di

modelli statistici la variabile di interesse Y rilevata attraverso indagini campionarie ad alcune variabili chiave X (disponibili anche su base censuaria), ed aggregare a livello di dettaglio desiderato.
Il filone nel quale si inserisce questo particolare approccio è quello dedicato alle procedure sintetiche di generazione di microdati, in quanto la tecnica bottom-up dà luogo alla possibilità di aggregare i dati in base al criterio preferito ed a livello di dettaglio territoriale voluto.
Per risolvere il problema della disaggregazione territoriale Benedetti e Palma costruiscono un modello basato sulla tecnica degli alberi di regressione per arrivare ad una partizione in gruppi omogenei delle famiglie campionate, sulla base di un insieme di variabili esplicative rilevate nel censimento generale della popolazione, usano, cioè, una tecnica non parametrica la quale si è rivelata più flessibile ed adeguata del modello regressivo.
Le variabili usate si riferiscono a dati individuali di carattere socio-demografico del capofamiglia, a dati riferiti al nucleo famigliare nel suo insieme , a dati relativi all'abitazione ed a dati geodemografici come regione di residenza o tipologia insediativa del comune. Il passo successivo dell'analisi consiste nell'applicare il modello di segmentazione costruito sui dati campionari all'universo delle famiglie, per poi procedere all'aggregazione in modi diversi per ottenere una stima del reddito e della spesa.
A favore di tale procedura bisogna annoverare la relativa semplicità dello sviluppo metodologico per ottenere le stime per piccoli domini ed il contenuto costo in termini di elaborazione dell'applicazione del modello ad un archivio di grandi dimensioni.

3.3 MODELLO DI REGRESSIONE CON DATI AREALI

Prima di affrontare l'analisi di un modello lineare nel caso di dati provenienti da località distribuite nello spazio, si richiamo alcuni concetti base.

Nella costruzione di un modello econometrico ci si basa su un insieme di

osservazioni $\boldsymbol{x}=(x_{1,}\cdots,x_n)$ che possono essere pensate come il risultato di un esperimento singolo relativo ad un insieme di variabili aleatorie $(X_{1,}\cdots,X_n)$.

Su tali variabili e sul loro legame con i dati osservazionali vengono fatte alcune ipotesi:

- Ipotesi sul modello di probabilità che imponga una forma plausibile per la distribuzione congiunta delle X del tipo $\Phi=\{f(X_{1,}\cdots,X_n,\theta),\theta\in\Theta\}$, dove $f(X_{1,},...,X_n)$ è la funzione di densità congiunta e θ l'insieme dei parametri rilevati.
- Ipotesi sul modello di campionamento, cioè una serie di informazioni circa il criterio di selezione delle osservazioni dal modello di probabilità.

Le tipologie di dati economici possono essere di tre tipi:

- DATI RELATIVI A CROSS-SECTION. Questi misurano una particolare variabile in relazione a singoli agenti economici.. Solitamente il criterio di campionamento è quello casuale semplice che porta a considerare le variabili indipendenti ed identicamente distribuite con funzione di densità $f(X,\theta)$. In questo modo il modello di probabilità può essere scritto come: $\Phi=\{\prod_{i=1}^{n} f(X_i,\theta_i),\theta\in\Theta\}$.
- DATI RELATIVI A SERIE STORICA. Queste misurano variabili relative ad un agente economico in istanti di tempo successivi. In questo caso il modello di probabilità può essere scritto come: $\Phi=\{f(X(t_1),X(t_2),\cdots,X(t_n),\theta),t_i\in I,\theta\in\Theta\}$,
  dove $\{f(X(t_1),...,X(t_n)),\theta\}$ è un processo aleatorio ordinato secondo l'indice $t\in I$ che può essere $t\in\mathbb{R}$ (processi aleatori a parametro continuo) o $t\in I_+$ (processi aleatori a parametro discreto). In questo ultimo caso, però, la funzione di densità congiunta diventa

$$\Phi=\{\prod_{i=1}^{n} f(X(t_i)|X(t_1), X(t_2), \cdots, X(t_{i-1})); t_i \in I\, \theta \in \Theta\}\ .$$

- DATI RELATIVI A SERIE SPAZIALI. Questi misurano variabili relativi ad agenti economici con riferimento alla loro posizione nello spazio geografico. In tal caso il modello probabilistico può essere
  $\Phi=\{f(X(s_1), X(s_2), \cdots, X(s_n), \theta), s \in S, \theta \in \Theta\}$,
  dove $f\{(X(s_1),...,X(s_n), \theta\}$ rappresenta un campo aleatorio ordinato secondo l'indice $s \in S$ che può essere $s \in \mathbb{R}$ (campi aleatori a parametro continuo) oppure $s \in S_+$ (campi aleatori a parametro discreto).
  In questo ultimo caso non è possibile esprimere la funzione di densità congiunta a causa della multilateralità della dipendenza nel caso di dati spaziali.

Quando in econometria si deve eseguire una analisi di dati spaziali, come si è visto, il problema principale consiste nel fatto che il modello di probabilità alla base del meccanismo generatore dei dati deve tenere conto della struttura spaziale. Si introduce, allora, il concetto di processo aleatorio spaziale o campo aleatorio.

Sia (Ω,B,P(.)) uno spazio di probabilità ed S un insieme non vuoto in $\mathbb{R}^2$ e si definisce la funzione $X(.,.)$ con $X(.,.)$ tale che $\Omega \times S \rightarrow \mathbb{R}$ .L'insieme di variabili aleatorie $\{X(.,s), s \in S\} = \{X(s), s \in S\}$ è detta processo aleatorio spaziale o CAMPO ALEATORIO.

Come si vede le caratteristiche di un campo aleatorio sono:

- La natura degli indici dell'insieme *S.* Questa, come abbiamo visto, può essere continua ed allora s rappresenta le coordinate di n punti in $\mathbb{R}^2$ oppure discreta ed in questo caso s indica le coordinate su di una griglia regolare oppure una sequenza ordinata di valori riferiti ad un insieme finito di poligoni (s sarà allora uno scalare relativo al numero d'ordine arbitrariamente assegnato al poligono).

- La struttura di dipendenza spaziale delle variabili aleatorie $\{X(s),\ s \in S\}$ può essere espressa dalla funzione di densità congiunta $F(X(s)) = F(X(s_1), X(s_2), \cdots, X(s_n)) =$ $Prob\{X(s_1) \leqslant x_1, X(s_2) \leqslant x_2, \cdots, X(s_n) \leqslant x_n\}$.

Spesso nelle analisi reali ci si trova a dover analizzare contemporaneamente vari fenomeni spazialmente distribuiti, come per esempio il consumo ed il reddito o il prezzo e la quantità. Si definisce,allora, il campo aleatorio vettoriale di dimensione k.

Una CAMPO ALEATORIO VETTORIALE $\{X(s),\ s \in S\}$è definito come il processo $X(s) = (X_1(s), X_2(s),...,X_k(s))'$ nel quale ciascuna componente $X_i(s)$ definisce un processo aleatorio $\{X_i(s),\ s \in S\}$.

In questo modo si introduce un'ulteriore dimensione di analisi in quanto in ogni punto s_i, i=1,...,n è associato un vettore aleatorio $X(s)$ di dimensione k.

Dato, quindi, un campo aleatorio vettoriale $Z(s_i) = (y_i, x_i)'$ con $\{Z(s),\ s \in S\}$ che genera un insieme di dati osservati in n località (s_1,...,s_n) si vuole costruire un modello che spieghi il comportamento della variabile y_i nella località s_i in termini delle altre variabili aleatorie componenti il processo.
Si impongono, allora, tre blocchi di ipotesi:

- IPOTESI SUL MODELLO DI PROBABILITÀ.

  $Z_i \sim$ MVN $\quad \Phi = \{f(Z_i, \theta \in \Theta\}$ $\quad f(Z_i, \theta_i) \sim$MVN. Da ciò:

  MP1) $D(y_i | x_i, \theta) \sim N$ Normalità delle distribuzioni condizionate

  MP2) $E(y_i | x_i) = \beta' x_i$ Linearità della media

  MP3) $Var(y_i | x_i) = \sigma^2$ Omoschedasticità

  MP4) θ non dipende dalla località s_i Invarianza spaziale dei parametri

- IPOTESI SUL MODELLO STATISTICO.

  MS1) Il modello è lineare: le osservazioni della variabile y sono generate dalla somma di una componente sistematica e da una

componente erratica $y_i = \mu_i + u_i = E(y_i | x_i) + u_i = \beta' x_i + u_i$

MS2) Dalla normalità discende $\theta \equiv (\beta; \sigma^2)$

MS3) x_i è debolmente esogene rispetto a θ

MS4) Non si hanno informazioni a priori su θ

MS5) $X \equiv (X_1,...,X_n)'$ è una matrice a rango pieno per tutti i valori osservabili di X

- IPOTESI SUL MODELLO DI CAMPIONAMENTO.

  MC1) $Z \equiv (Z_1,...,Z_n)'$ è un campione estratto con criterio casuale semplice dalla distribuzione $D(Z_i;\theta)$ $i=1,...,n$

In ALTERNATIVA le ipotesi sul modello di probabilità e sul modello di campionamento possono essere incorporate in un'unica ipotesi sulla componente di errore stocastico.

- $y_i = \beta' x_i + u_i$
- $D(u | X)$ è i.i.d. N $(0, \sigma^2 I_n)$ (MP1-MP4,MC1)
- Non si hanno informazioni a priori su $\theta \equiv (\beta ; \sigma^2)$ (MS4)
- **X** è una matrice a rango pieno (MS5)

Quando si lavora con dati osservazionali spaziali le ipotesi violate sono quelle relative al modello di probabilità ed al modello di campionamento mentre non vi è nulla da dire sulle ipotesi riguardanti il modello statistico generatore dei dati .

3.3.1 RIMOZIONE DELLE IPOTESI SUL MODELLO DI CAMPIONAMENTO

Se viene a mancare l'ipotesi di campionamento casuale semplice non sono più validi i risultati relativi alle stime ed ai test d'ipotesi.

In questo caso, allora, bisogna considerare un modello di campionamento che, sempre con l'ipotesi di processo gaussiano assume la forma più generale:

$$\begin{pmatrix} Z_1 \\ \cdots \\ Z_n \end{pmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \cdots \\ \mu_n \end{bmatrix}, \begin{bmatrix} V_1 & C_{12} & \cdots \\ \cdots & \cdots & \cdots \\ C_{1n} & \cdots & V_n \end{bmatrix} \right),$$

dove $V_i = V(s_i)$ è la matrice di covarianza spaziale e $C_{ij} = C\ (s_i\ s_j)$ quella di cross-covarianza.

Un simile processo, però, non è studiabile perché nell'analisi econometrica si dispone di una singola osservazione per ogni località e il numero di parametri da stimare nel modello è largamente superiore a quello delle osservazioni disponibili. E' necessario, allora, introdurre delle restrizioni che permettano di ridurre il numero di parametri incogniti. In pratica ci si affida ad una serie di campi aleatori dei quali si conoscono a fondo le proprietà.

A seconda di come si decide di specificare il modello di campionamento avremo due diversi approcci: uno basato sulla rispecificazione, l'altro sulla autocorrelazione che va ad agire sul componente non sistematica u_i.

3.3.1.1 L'approccio basato sulla rispecificazione

Attraverso la seguente procedura è possibile rimuovere la violazione sull'ipotesi del modello di campionamento attraverso la rispecificazione del modello (MC1). Tale violazione porta a costruire stime di β e σ^2 distorte, inconsistenti ed inefficienti, ed inoltre i test *t* di significatività dei parametri e il test *F* del modello non sono più validi.

In pratica, allora, consideriamo $\{Z(s),\ s \in S\}$ un campo aleatorio stazionario e gaussiano che obbedisca alla legge autonormale.

Un campo aleatorio $\{Z(s),\ s \in S\}$ è detto autonormale (o condizionale autoregressivo, CAR) se la densità condizionata di ciascuna variabile aleatoria rispetto alle altre è:

$$f(Z(s_i)|Z(s_j), s_i \neq s_j) = f(Z(s_i)|Z(s_j) s_i \neq s_j, s_j \in N(i)) =$$

$$= (2\pi\sigma_i^2)^{\frac{-1}{2}} \exp\{\frac{-1}{2\sigma_i^2}[Z(s_i) - \mu_i - \sum_{i \neq j} k_{ij}(Z(s_j) - \mu_j)]\}$$

dove N(i) corrisponde al vicinato di s_i, $\mu_i = E(Z(s_i))$, $\sigma^2_i = Var\ (Z(s_i))$ e

$$k_{ij}=\begin{cases}\rho & se\, s_j \in N(i),\ \rho\ è\ correlazione\ spaziale \\ 0 & altrimenti\end{cases}.$$

La densità congiunta di tale campo aleatorio, di conseguenza, è :

$Z \sim$ MVN (μ, **V**)

con $\mu = (\mu_1, ..., \mu_n)$ e

$$\mathbf{V} = (\mathbf{I}- \mathbf{K})^{-1}\ \mathbf{M},\ \mathbf{M} = \begin{pmatrix}\sigma_1^2 & \cdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \cdots & \sigma_n^2\end{pmatrix},\ \mathbf{K} = \{\rho\ w_{ij}\},\ \ w_{ij}=\begin{cases}1 & se\, s_j \in N(i) \\ 0 & altrimenti\end{cases}.$$

Se ci si limita ai processi bivariati $Z(s_i) = (y_i , x_i)'$ possiamo esprimere il valore atteso condizionato della y_i come:

E $\{y_i \mid y_j\ j \in N(i)\ ;\ X_i\ ;\ X_j\ j \in N(i)\} = \alpha \sum_{i \neq j} w_{ij} X_j + \beta X_i + \gamma \sum_{i \neq j} w_{ij} y_j$, con α, β e γ parametri.

Da questa formula possiamo definire il modello statistico di generazione dei dati come:

$y_i = \alpha \sum_{i \neq j} w_{ij} X_j + \beta X_i + \gamma \sum_{i \neq j} w_{ij} y_j + u_i$, dove u_i è la componente non stocastica definita come differenza tra y_i e il valore atteso condizionato.

Tale processo ha matrice di varianza e covarianza

$$\mathbf{Q} = \left\{ I^{\circ} - \begin{bmatrix} \gamma W & \alpha W + (\beta - \alpha) I \\ \alpha W + (\beta - \alpha) I & \gamma W \end{bmatrix} \right\}^{-1} M = \mathbf{Q}\ (\alpha, \beta, \gamma),$$

dove I° è una matrice diagonale unitaria di dimensione 2n, $W = \{w_{ij}\}$ e

$M = \text{diag}\ \left({}_Y\sigma_1^2, {}_Y\sigma_2^2, \cdots, {}_Y\sigma_n^2, {}_X\sigma_1^2, \cdots, {}_X\sigma_n^2\right)$.

Con queste informazioni possiamo definire la funzione di verosimiglianza come $L(\theta, Z) = \frac{1}{\sqrt[n]{2\pi\sigma^2}} |Q(\alpha, \beta, \gamma)|^{\frac{-1}{2}} \exp\left\{ \frac{-1}{2\sigma^2} Z' Q(\alpha, \beta, \gamma)^{-1} Z \right\}$, sulla quale in seguito costruire le stime ed i test di ipotesi.

3.3.1.2 L'approccio basato sulla autocorrelazione

Se si considera la formulazione alternativa del modello di regressione

$y_i = \beta' x_i + u_i$ si può rimuovere la violazione dell'ipotesi di campionamento indipendente lavorando sulla componente non sistematica u_i.

La violazione dell'ipotesi, in questo secondo caso, comporta che la stima di σ^2 è distorta ed inconsistente, mentre la stima di β è corretta ma non efficiente; in aggiunta, i test di ipotesi perdono di validità.

In caso di campionamento casuale semplice u_i può essere considerato un campo aleatorio rumore bianco, cioè le variabili aleatorie componenti il processo sono tutte a media nulla $E(u(s)) = 0 \ \forall s$ e linearmente indipendenti a due a due qualunque sia la loro posizione nel piano

$$E(u(s_i),u(s_j)) = \begin{cases} \sigma^2 & per\ s_i = s_j \\ 0 & altrimenti \end{cases}.$$

Nel caso, invece, di campionamento non casuale da dati spaziali possiamo ipotizzare per u_i un modello autoregressivo simultaneo (SAR).

Il modello di regressione diventa allora $y_i = \beta' x_i + e_i$ con $e_i = \rho \sum_{i \neq j} w_{ij} e_j + u_i$,

dove u_i è rumore bianco e $w_{ij} = \begin{cases} 1 & se\ s_j \in N(i) \\ 0 & altrimenti \end{cases}$.

E' necessario, di conseguenza, fare inferenza su tre parametri $\theta \equiv (\beta;\ \sigma^2;\ \rho)$.

Tale problema viene risolto con la funzione di verosimiglianza

$$L(\theta, y) = \frac{1}{\sqrt{(2\pi\sigma^2)^n}} |V(\rho)|^{\frac{-1}{2}} \exp\left\{ \frac{-1}{2\sigma^2} (y - \beta' X)' V(\rho)^{-1} (y - \beta' X) \right\},$$ dove

$\mathbf{V}(\rho) = [(\mathbf{I}-\mathbf{B})'\ (\mathbf{I}-\mathbf{B})]^{-1}$ ($\mathbf{B}=\{\rho w_{ij}\}$) è la matrice di varianze ed autocovarianze spaziali di un processo SAR.

3.3.2 RIMOZIONE DELLE IPOTESI SUL MODELLO PROBABILISTICO

Quando si lavora con dati economici distribuiti nello spazio di solito vengono violate le ipotesi sulla omoschedasticità ($Var(y_i | x_i) = \sigma^2$) e sulla invarianza spaziale dei parametri (θ non dipende dalla località s_i).

L'omoschedasticità deriva dall'ipotesi di normalità del modello di probabilità; ma qualora tale evento non si verifichi lo stimatore dei minimi quadrati ordinari $\boldsymbol{\beta} = (\mathbf{X'X})^{-1}\,\mathbf{X'y}$ rimane corretto e consistente ma diviene inefficiente.

Si usa, allora, lo stimatore ottenuto stimando ai minimi quadrati generali l'espressione $l(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\,\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$, con $\mathbf{V}$ matrice da stimare sulla base delle informazioni campionarie.

Per quanto riguarda la seconda violazione di ipotesi si può affermare che nella pratica non si può ipotizzare che il vettore di parametri di interesse

$\theta = (\beta, \sigma^2)$ sia costante al variare delle località, cioè che località diverse reagiscano allo stesso modo agli stessi condizionamenti.

Si consideri, per esempio, il caso in cui avviene un solo cambiamento strutturale lungo una linea di confine che delimita due zone di piano S_1 ed S_2 ($S_1 \cap S_2 = S$) costituite rispettivamente da n_1 ed n_2 località.

In tal caso il modello diventa $\begin{cases} y_i = \beta_1{}'\boldsymbol{X_i} + \boldsymbol{u_{1,i}} & Se\, i \in S_1 \\ y_i = \beta_2{}'\boldsymbol{x_i} + \boldsymbol{u_{2,1}} & Se\, i \in S_2 \end{cases}$, con $S_1 = \{s_1,\ldots,s_{n1}\}$ e $S_2 = \{s_1,\ldots,s_{n2}\}$, $n = n_1 + n_2$ il numero complessivo di località considerate e con $\theta_1 = (\beta_1, \sigma^2_1)$ e $\theta_2 = (\beta_2, \sigma^2_2)$ i parametri di interesse.

L'ipotesi di assenza di cambiamenti strutturali diventa allora, $H_o : \theta_1 = \theta_2$.

Tale ipotesi può essere verificata con la statistica test di Chow:

$CH = \frac{\sqrt{-SQR1} - SQR2}{SQR1 + SQR2}\,\frac{n - 2k}{k}$, dove SQRT è la somma dei quadrati dei residui totali, SQR1 dei residui relativi a S_1 ed SQR2 dei residui relativi a S_2 e k è la dimensione del vettore θ. Tale statistica è distribuita sotto H_o come una F(k, n-2k); ma presenta dei problemi nel caso in cui sia violata l'ipotesi sul modello di campionamento.

Si usa, allora, il test asintotico di Wald basato sulla statistica

$cost\ [e_o'\Psi^{-1} e_0 - e_1'\,\Psi^{-1} e_1]$, dove e_0 e e_1 sono i residui della regressione stimati attraverso la massima verosimiglianza rispettivamente sotto l'ipotesi nulla e sotto l'ipotesi alternativa ($H_1 : \theta_1 \neq \theta_2$). Tale statistica si distribuisce asintoticamente come una $\chi^2(k)$.

Per la stima di cambiamenti strutturali si può, anche, usare il modello della

regressione a più regimi $y^* = \beta^{*\prime} X^* + e^*$,

dove $y^* = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$; $X^* = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix}$; $\beta^* = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$; $e^* = \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$ e se consideriamo l'approccio basato sull'autocorrelazione $\mathbf{e^*} = \mathbf{B^*e^*} + \mathbf{u}$, con $\mathbf{u}$ rumore bianco con matrice di varianze e covarianze spaziali

$\mathbf{\Sigma_u} = \begin{bmatrix} \sigma_1^2 I_{n1} & 0 \\ 0 & \sigma_2^2 I_{n2} \end{bmatrix}$, I_{ni} matrice unitaria di dimensione n_i e

$\boldsymbol{B} = \begin{bmatrix} \rho_1 W_1 & 0 \\ 0 & \rho_2 W_2 \end{bmatrix}$, con $\mathbf{W_1}$ e $\mathbf{W_2}$ sono le matrici di vicinato relative alle località incluse rispettivamente in S_1 e S_2.

Con tali notazioni possiamo definire la funzione di log-verosimiglianza

$$l(\theta, x) = \frac{-n_1}{2} \log \sigma_1^2 - \frac{n_2}{2} \log \sigma_2^2 + \log|I - B| - \frac{1}{2} \Gamma ,$$

con $\mathbf{\Gamma} = (\mathbf{y^*} - \boldsymbol{\beta}^{*\prime} \mathbf{X^*})' (\mathbf{I} - \mathbf{B})' \mathbf{\Sigma_u} (\mathbf{I} - \mathbf{B}) (\mathbf{y^*} - \boldsymbol{\beta}^{*\prime} \mathbf{X^*})$ e costruire le stime ed i test di ipotesi.

4. DISPONIBILITA' DI UN SUPPORTO INFORMATIVO ADEGUATO

Per affrontare alcuni dei problemi di geomarketing analizzati si è sottolineato, a volte, la necessità di una adeguata strumentazione non solo hardware (potenza di calcolo) ma anche software. Si è così diffusa la tecnologia GIS.

Questa per essere di vero supporto alle decisioni strategiche aziendali dovrebbe unire la facilità dell'immagazzinamento di grosse mole di dati e la pratica restituzione grafica su mappa con una seria base di statistica spaziale. Tale integrazione di capacità può essere facilmente raggiunta da una struttura modulare che permetta ad ogni singola azienda di costruire uno strumento personalizzato sulle proprie problematiche.

CAPITOLO 3
STUDIO DI FATTIBILITA'

1. INTRODUZIONE

Quando ci si sposta dal campo della teoria a quello della pratica il primo problema da affrontare è quello relativo ai dati.
Se lo statistico deve lavorare su una consulenza, molto spesso i dati gli vengono forniti da altri e lui deve solo valutare la bontà di raccolta, ma se lo statistico decide di eseguire una ricerca libera deve affrontare anche il problema del reperimento della "materia prima".
E' questo infatti il primo scoglio che ho dovuto affrontare nella stesura di questa tesi.
In una realtà che ha fatto dell'informazione una requisito indispensabile è difficile trovare "buoni dati", cioè dati attendibili, pronti all'uso, ben documentati, ecc... Con questo voglio sottolineare l'esigenza di corredare ogni tabella con un accurato corpo di metadati, cioè una serie di documenti che descrivano il contenuto del database e la sua origine in modo da permettere di scoprire il potenziale nascosto in molti archivi.
Se la ricerca non ha un obiettivo preciso, oserei dire particolarissimo, si rischia di naufragare in un oceano di informazioni a volte inutili o fuorvianti. Ma se l'obiettivo è troppo restrittivo si rischia di perdere il piacere della scoperta inattesa che a volte capita di trovare sulla decima o ventesima pagina di un motore di ricerca, o sulla bibliografia di un articolo o libro non proprio legato all'argomento richiesto.

Poiché la mia ricerca si è svolta prevalentemente attraverso internet, ho trovato molto utile, anche, redigere durante le fasi di analisi un "diario di bordo": una sorta di traccia delle sequenze dei link alle pagine visitate ed il dato o la frase che mi ha permesso di compiere tale collegamento. Tale strumento mi ha consentito di non "naufragare", ma di apprezzare in pieno la più importante, a mio parere, caratteristica di internet: la ipertestualità, cioè la possibilità di collegare le informazioni interattivamente. A volte infatti mi è capitato di tornare sulla stessa pagina dopo una divagazione attraverso un link che ha portato ad un vicolo cieco e solo attraverso la segnatura dei percorsi già seguiti ho potuto migliorare la ricerca; in aggiunta a volte ho visitato pagine diverse dello stesso sito perché ad esse portavano collegamenti diversi e solo una trascrizione del percorso seguito mi ha permesso di riconoscere in seguito quale era stato il ragionamento che mi aveva condotto a "stanze diverse contenute nella stessa casa".

La mia ricerca è focalizzata sulle ASSOCIAZIONI MUSICALI INTERESSATE ALLA MUSICA CLASSICA, ma poiché molto spesso queste si nascondono sotto nomi improponibili ho dovuto estendere il raggio a tutte le associazioni culturali.
Tale studio è seguito in collaborazione con la dott. ssa Silvia Rita Sedita appartenente al Dipartimento di Economia dell'Università degli Studi di Padova che da alcuni anni si interessa a tale argomento.

2. RICERCA DEI DATI

Il primo traguardo che mi sono posta è stato quello di individuare un ELENCO soddisfacente ed attendibile di tali associazioni, perché sono molti i siti che propongono elenchi di gruppi "culturali" (e non) senza documentare approfonditamente come sono riusciti a costruire tale database (a volte si tratta di siti ai quali ci si può iscrivere gratuitamente per pubblicizzare il complesso musicale allestito con gli amici).

Non sapendo di preciso dove tali circoli sono obbligati ad iscriversi ho deciso di partire da una ricerca libera.

Fortunatamente ho trovato sul sito del comune di Venezia l'elenco delle associazioni e il modulo di iscrizione. Da tale elaborato ho scoperto l'esistenza della legge n.266 della'11 agosto 1991, che nell'articolo n.6 impone alle Regioni e alle provincie autonome di tenere un REGISTRO DELLE ORGANIZZAZIONI DI VOLONTARIATO.
Nell'articolo n.12, inoltre, tale legge istituisce l'osservatorio nazionale per il volontariato che fra i sui compiti ha anche i seguenti:
a) provvede al censimento delle organizzazioni di volontariato ed alla diffusione della conoscenza delle attività da esse svolte;
b) promuove ricerche e studi in Italia e all'estero;
f) pubblica un rapporto biennale sull'andamento del fenomeno e sullo stato di attuazione delle normative nazionali e regionali;
h) pubblica un Bollettino periodico di informazione e promuove altre iniziative finalizzate alla circolazione delle notizie attinenti l'attività di volontariato;
i) promuove, con cadenza triennale, una Conferenza nazionale del volontariato, alla quale partecipano tutti i soggetti istituzionali, i gruppi e gli operatori interessati.

Visto che non sono stata in grado di trovare notizie di questo bollettino ho deciso, allora di spostare la ricerca sulla REGIONE VENETO; ed ho scoperto la legge regionale 30 agosto 1993 n.40 che all'articolo n.4 comma 1 recita: “ E' istituito, presso la Giunta regionale, il registro regionale delle organizzazioni di volontariato che può essere articolato in sezioni con deliberazione della Giunta medesima.” Nell'articolo 5, inoltre, viene istituito l'Osservatorio regionale sul volontariato che fra l'altro “fornisce ogni utile elemento per la promozione e lo sviluppo del volontariato anche tramite proposte di ricerche e studi” e “promuove la diffusione delle attività svolte dalle organizzazioni di volontariato e dalle loro federazioni e la pubblicazione del rapporto regionale

sull'andamento del volontariato nella regione".
Nell'articolo 14, inoltre, vengono istituiti i Centri di Servizi per il volontariato che oltre ad essere di ausilio alle associazioni offrono informazioni, notizie, documentazioni e dati sulle attività di volontariato locale e nazionale.
Sul sito della Regione (www.regione.veneto.it), per quanto abbia cercato, però, non sono stata in grado di trovare un riferimento a questo registro, che si trova, infatti, sul sito WWW.VENETOSOCIALE.ORG, sotto la categoria non-profit ed al link protagonisti.
A questo punto mi sia permesso di sottolineare quanto siano male gestiti siti di alto interesse come possono essere quelli istituzionali; per i quali dovrebbe essere d'obbligo una tensione estrema verso al facilità d'uso.
Il sito della Regione, pur presentando una sezione "Statistica" e un motore di ricerca interno non fa nessuna menzione di questo elenco e neppure del sito nel quale è possibile reperirlo.
Il motore di ricerca interno del sito Veneto Sociale, poi, inserendo le parole "Registro" AND "Organizzazioni" oppure chiavi alternative porta ad un numero elevato di risultati inutili e fuorvianti che non riescono a facilitare la ricerca; solo un analisi attenta delle sezioni (presenti nella parte alta del frame) e dei link contenuti in ogni pagina riesce ad indirizzare con successo la ricerca.
Per ultimo mi sia consentito anche una puntualizzazione sulla grafica che ho trovato eccessivamente "chiassosa" perché piena di colori con link non sempre disposti in modo adeguato.

Tale ricerca ci permette, però, di rappresentare solo una parte delle associazioni culturali: le organizzazioni di volontariato, cioè *"ogni organismo liberamente costituito al fine di svolgere [...] l'attività di volontariato che deve intendersi quella prestata in modo personale, spontaneo e gratuito, tramite l'organizzazione di cui il volontario fa parte, senza fini di lucro anche indiretto ed esclusivamente per fini di solidarietà"*. Inoltre l'iscrizione al registro non è obbligatoria, ma essendo condizione necessaria per accedere ai contributi e ai finanziamenti pubblici, nonché per stipulare le convenzioni e per beneficiare

delle agevolazioni fiscali, si ritiene che tutte le organizzazioni si siano iscritte, quindi tale registro può essere pensato come una buona rappresentazione dell'intero universo delle organizzazioni di volontariato. Restano pertanto escluse da tale elenco molte associazioni che attraverso le loro attività favoriscono e diffondono la cultura musicale.

Con l'aiuto della dott. ssa Sedita ho scoperto, poi, che il COMUNE DI PADOVA presentava sul sito www.padovanet.it (rete civica di Padova) un registro comunale delle associazioni e delle cooperative sociali. In particolare il registro viene definito nel seguente modo: *"Il registro comunale delle libere forme associative e delle cooperative sociali ha la funzione di promuovere e valorizzare le associazioni, mediante la loro partecipazione e il loro coinvolgimento a servizio della collettività"*.

Si possono iscrivere le associazioni e le cooperative sociali che non hanno fini di lucro e che operano effettivamente nel territorio del Comune di Padova (anche se la sede è altrove), con almeno un anno di attività. In particolare:

- le libere forme associative iscritte ai registri regionali previsti dalla normativa vigente in materia di volontariato;
- le cooperative sociali (secondo la normativa vigente);
- le altre libere forme associative e le realtà confessionali locali.

Non possono, invece, chiedere l'iscrizione:

- gli ordini professionali di cui all'art. 2229 e seguenti del Codice civile;
- le forme organizzative (comprese le istituzioni) e associative di diretta derivazione dell'amministrazione comunale e dei Consigli di quartiere o di altri enti o amministrazioni pubbliche;
- i sindacati, i partiti politici, le organizzazioni di categoria.

Anche in questo caso l'iscrizione al registro costituisce titolo preferenziale per la concessione di sostegni economici da parte del Comune. Abbiamo trovato quindi un altro elenco che permette di descrivere in modo esaustivo solo una parte della realtà dell'universo delle associazioni.

Visti i risultati di questa ricerca propongo, qui di seguito, ALTRE POSSIBILI

LINEE DI AZIONI che per essere intraprese necessitano di un maggiore impiego di risorse. Finora gli unici strumenti utilizzati sono stati un computer collegato ad internet (attraverso linea adsl) ed il tempo necessario ad eseguire la ricerca. Nel caso in cui si ritenga utile l'utilizzo di ulteriori strumenti di indagine bisogna operare anche una valutazione costi/benefici. Nel campo della statistica, infatti, si ha la possibilità e quando si lavora su potenziali grosse moli di dati diviene d'obbligo valutare se i risultati ottenibili giustificano lo spreco di risorse impiegate nell'indagine. Nel mio caso le risorse usate: internet e tempo hanno un costo pari a zero, poiché ho potuto usufruire dell'accesso ad internet fornito a tutti gli studenti iscritti a scienze statistiche ed il tempo opportunità impiegato nella ricerca non può essere valutato poiché è d'obbligo la stesura di un elaborato finale per conseguire la laurea triennale.

Immagino che ogni associazione culturale abbia una sede; a tale sede ritengo possa essere collegato un numero telefonico. Ritenendo l'elenco telefonico una rappresentazione esaustiva dell'universo di utenze, si potrebbe attraverso esso ricercare ogni associazione che nella denominazione espliciti appunto la propria natura di associazione culturale. Visto che si ritiene impossibile scorrere con attenzione e rapidità un elenco telefonico cartaceo della provincia di Padova, bisognerebbe disporre di una sua versione informatizzata che permetta di eseguire una ricerca attraverso il computer inserendo delle parole chiave. Questo percorso d'azione presenta però la seguente serie di problemi:

- Al giorno d'oggi molti utenti richiedono esplicitamente di non comparire sull'elenco telefonico. Ritenendo, però, che le associazioni culturali preferiscano essere pubblicizzate, questa opzione può essere scartata.
- Le uniche copie informatizzate dell'elenco telefonico che conosco sono i siti di pagine bianche e pagine gialle. Attraverso di esso, però, risulterebbe complicato eseguire una ricerca per denominazione come quella proposta.
- Poiché le denominazioni delle associazioni sono composte a volte da

parole non auto-esplicative delle finalità, le parole chiave da inserire come condizioni di ricerca sono molto varie e non si potrebbe avere mai la certezza di aver rilevato tutte le associazioni veramente presenti.

- Non si è in grado di avere un'idea della numerosità complessiva dell'universo delle associazioni e ritengo poco opportuno considerare tale popolazione infinita, poiché l'indagine si applica ad un particolare momento storico e non a tutte le possibili associazioni che potrebbero nascere in futuro (come avviene nella stima dell'efficacia di una farmaco quando la popolazione di malati viene considerata infinita per ammettere anche tutte le possibili unità future).

Bisognerebbe, altrimenti, fare un'indagine esplorativa attraverso intervista presso alcuni rappresentanti legali od amministratori di associazioni profit (visto che sono la parte di universo che non sono riuscita a rappresentare) per cercare di capire meglio l'ambiente di riferimento; si potrebbe, in particolare, chiedere se esiste un organismo che raccolga informazioni su tutti questi tipi di attività, oppure attraverso quale organo si pubblicizzano, oppure quali documenti sono obbligati a presentare sia al momento della costituzione sia durante la vita della società. Tale ricerca presenta, però, la seguente serie di difficoltà:

- E' necessario reperire in qualche modo il recapito del rappresentante legale oppure dell'amministratore, oppure di un associato di una organizzazione interessata alla musica classica; contattare la persona e sperare nella disponibilità di questa e nella sua conoscenza dell'ambiente nel quale opera.
- Non si può stimare a priori il numero di interviste da eseguire, né il tempo impiegato in tale indagine poiché, se fortunati, con una sola visita si possono ricavare le informazioni necessarie che portano direttamente ad un organismo nazionale che raccoglie tutte le imprese di spettacolo, oppure possono essere necessarie più interviste ad una serie di associazioni sempre più articolate in sottosezioni come delle scatole cinesi.

Si potrebbe, infine, contattare gli organizzatori di rassegne musicali o concorsi

oppure i rappresentanti degli uffici comunali, provinciali o regionali che si occupano di cultura. Attraverso essi, infatti, si potrebbe avere una descrizione dell'ambiente di riferimento, reperire i recapiti dei rappresentanti delle associazioni (come scritto in precedenza) oppure accedere alle liste delle associazioni che si iscrivono per partecipare a tali eventi. I problemi da affrontare se si intraprende una tale linea di azione sono simili a quelli affrontati in precedenza; ma in questo caso è più semplice reperire il recapito degli uffici comunali visto il loro ruolo istituzionale.

3. MEMORIZZAZIONE DEI DATI

Una volta trovato l'elenco è necessario, in seguito, trasferire su supporti idonei tali informazioni, per poterli analizzare, anche attraverso l'uso di alcuni strumenti software. Molto spesso per usare un software statistico di analisi dei dati oppure un GIS i file devono essere importati. Tale operazione risulta molto delicata perché solo un corretto formato dei dati originali permette di importare correttamente le informazioni e facilita tale operazione. In generale i dati originali dovrebbero essere memorizzati in un file excel o formati compatibili, oppure come file txt o simili, in questo ultimo caso lo spaziatore principe usato per dividere le colonne è il tabulatore, se sono presenti altri tipi di spaziatura è necessario prestare attenzione nella procedura di importazione.
In particolare io volevo memorizzare i dati su una tabella excel, per poter poi gestirli a mio piacere, ma ho trovato alcune difficoltà anche in questa operazione.

Il sito della regione Veneto nel quale è presente il registro delle associazioni di volontariato è organizzato in modo tale che risulta molto lunga e complessa una memorizzazione dell'elenco in un tabella. Per prima cosa è possibile eseguire una ricerca dell'associazione solo per denominazione o per provincia; ogni pagina, poi, presenta solo venticinque risultati; in aggiunta il motore di ricerca non fornisce il numero complessivo di risultati trovati e si è quindi

costretti ad operare un calcolo personale per ottenere una stima complessiva delle associazioni cercate. Il risultato della ricerca è di solito disposto in ordine alfabetico, ma è possibile ordinare le organizzazioni per denominazione o comune. Cliccando, poi, su ognuna di esse, nella pagina corrispondente, si trovano le informazioni relative alla sede (indirizzo, località e telefono) e agli obiettivi (obiettivo, età, materia). Risulta quindi evidente che per riportare tali informazioni in un file excel è necessario un lungo ed attento lavoro di copia ed incolla. Per la sola provincia di Padova, sono censite 401 organizzazioni; se si volesse alleggerire l'opera di trascrizione e rilevare solo quelle impegnate nell'ambito culturale-musicale sarebbe comunque necessario visualizzare ogni singolo link per leggere tra gli obiettivi e la materia quali sono l'ambito di competenza e le finalità dell'organizzazione.

Sul sito della città di Padova, invece, il registro delle associazioni non profit è organizzato in modo tale che è possibile eseguire la ricerca attraverso la denominazione, oppure il quartiere nel quale si trova la sede (fuori città; centro; nord; est; sud-est: sud-ovest; ovest), oppure per area tematica (interventi, servizi sociali e socio sanitari; tutela e impegno civile, solidarietà internazionale e pace; economia e lavoro; educazione e formazione; sport; cultura; territorio, urbanistica, tutela ambiente e protezione civile; tempo libero; combattentistica e d'arma). Anche in questo caso non viene presentato il numero totale del risultato della ricerca, ma quantomeno tutte le denominazioni delle organizzazioni cercate vengono presentate in un'unica pagina. Ad ogni associazione è legato, poi, un link nel quale si possono conoscere le generalità (denominazione, sezione tematica e numero di registro), l'indirizzo della sede (indirizzo, quartiere, telefono, fax, e-mail) ed una breve descrizione delle finalità.
Come per il caso precedente anche qui, per poter predisporre i dati per essere elaborati con un software di statistica o GIS è necessario un lungo lavoro di copia ed incolla.

Prima di iniziare un lavoro lungo e noioso ho valutato la possibilità di

continuare nella ricerca di una forma appropriata del registro regionale. Fortunatamente, attraverso i link presenti sul sito www.padovanet.it prima e centro servizi volontariato (www.csvpadova.org) poi, ho trovato la pagina del ministero del welfare nel quale sono elencati i siti delle Regioni italiane nei quali è possibile reperire il registro delle organizzazioni non profit. Visitando, in particolare, il link della regione veneto sono tornata sul sito www.venetosociale.it in una pagina non visitata prima perché contenuta nella sottosezione documenti. In essa ho trovato il file compresso contenente l'elenco delle organizzazioni regolarmente iscritte aggiornato al luglio 2004. Scaricato il file e decompresso ho visto che si componeva di sette cartelle excel corrispondenti ognuna ad una provincia e composte, al loro interno delle colonne: numero, codice, denominazione, indirizzo, cap, comune, obiettivo, età, materia e stato.

Sempre attraverso i link presenti sul sito del centro servizi volontariato ho trovato il collegamento al registro provinciale di Padova. Nel file pdf scaricabile è presente il registro provinciale delle libere forme associative, cooperative, sociali e fondazioni. Tale elenco è suddiviso per sezioni tematiche (interventi socio-sanitari; tutela ed impegno civile; solidarietà internazionale e pace; economia e lavoro; cultura, beni culturali, educazione e formazione; sport; territorio – urbanistica – tutela ambientale e Protezione Civile) e riporta per ogni associazione denominazione, indirizzo della sede, telefono, eventuale e-mail e la data di scadenza dell'iscrizione. Fortunatamente il file non è protetto ed è quindi possibile eseguire il copia ed incolla, con tutti gli inconvenienti già evidenziati nei casi precedenti.

Sul sito del centro sevizi volontariato è presente anche il collegamento al registro comunale che è quello presente su www.padovanet.it.

4. ANALISI DEI DATI

Se usiamo la rappresentazione insiemistica per descrivere la disponibilità di dati finora trovati, possiamo affermare quanto segue: dell'intero universo delle

associazioni od organizzazioni interessate alla musica classica operante nell'ambito della provincia di Padova possiamo rappresentare solo quelle di volontariato attraverso tre diverse tabelle la cui intersezione è non nulla.
Mi preme di sottolineare però che i registri trovati sono aggiornati a periodi diversi: quello regionale a luglio 2004, quello provinciale a novembre 2005 e in quello comunale non viene indicata alcuna data. Questo, purtroppo, crea dei problemi per quanto riguarda i confronti delle tabelle in quanto l'iscrizione ai registri ha una scadenza in genere triennale ed esiste la possibilità di ritirare la propria adesione.
Importando queste tabelle in un dbms come può essere Access è possibile attraverso interrogazioni (query) fare alcune considerazioni preliminari. La flessibilità di tali strumenti software, infatti, permette di fare un numero molto elevato di analisi il cui unico vincolo è posto solo dalla coerenza dei criteri di ricerca e dall'interesse del ricercatore.

Visto che l'argomento dell'elaborato sono i GIS ipotizzo nel seguito una serie di analisi che coinvolgano la spazialità del dato.
In ogni elenco di associazioni finora trovato era sempre presente l'indirizzo della sede dell'organizzazione a conferma di ciò che si affermava nei capitoli precedenti su come il collegamento alla posizione geografica sia sempre presente ma non approfonditamente sfruttato. Grazie alla presenza di questo particolare campo e alle capacità elaborative dei software GIS è possibile rappresentare su una cartina la posizione delle sedi dei gruppi individuati. A questo punto sono numerose le analisi che si possono eseguire e quelle che seguono sono solo alcune proposte.
Come prima cosa si potrebbe partire da un'analisi visiva del risultato della mappatura per iniziare ad avere una visione complessiva della distribuzione territoriale delle associazioni non profit; si potrebbe, in seguito, individuare con colori diversi le differenti aree tematiche a cui afferiscono i gruppi di volontariato. Attraverso queste semplici rappresentazioni visive si possono trarre le prime considerazioni riguardanti la dispersione o concentrazione delle

unità individuate in determinate aree. Si possono evidenziare, in questo modo, situazioni anomale sia per un estremo che per l'altro sia riguardo alle organizzazioni di volontariato in generale che a particolari finalità. Da questi risultati l'analisi si può approfondire decidendo di mappare anche altre entità che possono risultare di interesse per individuare le cause delle situazioni evidenziate in precedenza. Con riferimento particolare alle associazioni dedite alla musica classica si potrebbe aggiungere alla mappa la posizione dei conservatori, delle scuole di musica, dei teatri, degli auditorium, e di altri luoghi nei quali è possibile organizzare uno spettacolo di musica classica. Posso immaginare, infatti, che una concentrazione elevata di associazioni nei pressi di tali strutture sia dovuta alla vicinanza ad una risorsa fisica importante per diffondere la cultura musicale. Una situazione opposta può, invece, essere sintomo o di una egemonia da parte di un unico gruppo o di una carenza di strutture adeguate allo sviluppo dell'organizzazione. In questa analisi si incontrano, però, una serie di problemi:

- Bisogna stabilire a che livello di dettaglio analizzare il territorio, perché rappresentazioni a scale diverse danno effetti visivi diversi di concentrazione o dispersione.
- Bisogna identificare un opportuno grado per identificare le situazioni di concentrazione o dispersione
- Bisogna calcolare a che distanza una risorsa quale un teatro può essere considerato accessibile o meno.

Si potrebbe in seguito incrociare questi dati con altri di tipo socio-demografico-economico. In questo caso le combinazioni penso possano essere considerate infinite e solo lo scopo dell'indagine o la curiosità del ricercatore possono guidare nelle scelte più adeguate. In riferimento all'argomento dell'elaborato si potrebbero ricercare i dati sul numero di abitanti e relativizzarli con la superficie selezionata (operazione possibile in molti GIS) per ottenere la densità di popolazione. In questo modo si potrebbe collegare quest'ultimo risultato con la concentrazione o meno di associazioni. Collegamenti altrettanto interessanti si potrebbero fare con dati di tipo sociale quali ad esempio quelli

rilevati dall'Istat sull'impiego del tempo libero. In questo caso si evidenzia, però, il problema di individuare i dati idonei, di averli a disposizione a livello di dettaglio adeguato (anche se esistono delle tecniche, evidenziate, nel capitolo 2 che permettono di superare questo inconveniente) e di saper interpretare i risultati ottenuti. Infine, un uso di dati di tipo economico, a mio parere, risulta meno idoneo.

Tutte queste analisi potrebbero essere condotte con dati pubblici, la cui diffusione è abbastanza ampia e che, per esempio, potrebbero essere reperiti (a volte con difficoltà) su internet.

Se si disponesse, però, di dati più dettagliati riguardanti l'associazione e il luogo nel quale ha sede si potrebbero costruire tutta una serie di altre indagini molto più specifiche. Sarebbe interessante, infatti, conoscere la residenza degli associati per fare un'analisi delle distanze. Incrociando poi, questi risultati con un'analisi di accessibilità alla sede (anche questa disponibile ormai in molti GIS) si potrebbe individuare un'altra delle cause di concentrazione di alcuni gruppi in luoghi particolari. Si ritiene, infatti, che zone maggiormente coperte dai trasporti pubblici oppure con una viabilità più scorrevole e con adeguati parcheggi siano preferibili di zone più disagiate ed incoraggino anche persone residenti non nelle immediate vicinanze della sede ad iscriversi all'associazione. Con quest'ultimo caso si vuole evidenziare come i GIS e la disponibilità di informazioni adeguate siano di importanza molto rilevante per gli amministratori comunali e le aziende di pubblica utilità.

5. CONCLUSIONI

Potrei fare un lungo elenco sulle possibili applicazioni dei GIS sia in aziende pubbliche che private essendo sicura che non sarebbe mai esaustivo.

In questi mesi di ricerca mi sono imbattuta in alcuni atti di convegni (per esempio quelli dell'esri o di grass o del gis-day), in molte aziende private di consulenza, in alcune dispense di corsi universitari che portano esempi di

problemi di ogni natura ed appartenenti ad ogni campo, che possono essere trattati con l'ausilio dei GIS.

Nel primo capitolo volutamente si è parlato in generale dello strumento GIS, presentando le idee chiave alla base dello sviluppo delle possibili operazioni che si possono eseguire. Nel secondo capitolo, poi, si sono introdotte le problematiche legate al geomarketing. In questo modo ho potuto evidenziare da una parte la flessibilità dei GIS e dall'altra la possibilità di collegamenti con una interessante branca di studi che porta a molte applicazioni pratiche in campo aziendale.

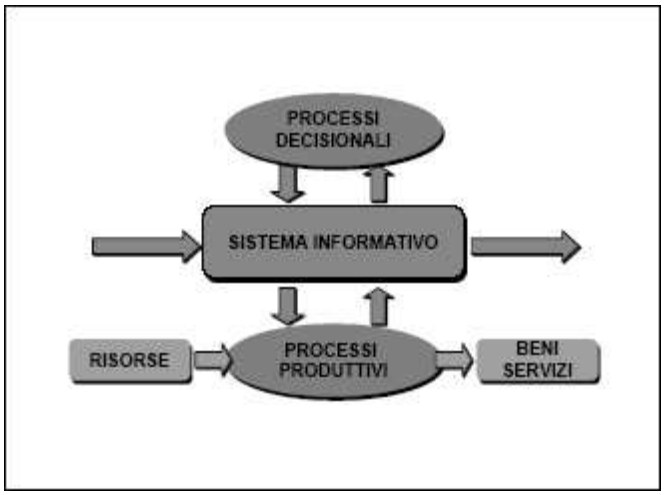
Nella parte precedente si è visto, infatti, come molte aziende di pubblica utilità potrebbero attuare politiche più favorevoli per il cittadino se i GIS fossero usati in modo più profondo e quotidiano. In aggiunta si stanno sempre più implementando anche nelle aziende private i sitemi informativi territoriali. Tale scelta è stata dettata da un mercato competitivo che impone alle aziende di tendere costantemente verso l'efficienza e l'efficacia e che in tal modo permette di apprezzare le potenzialità dei GIS che ne fanno un valido strumento di supporto alle decisioni.

Ammetto che quando ho iniziato questo studio pensavo che i GIS servissero solo a rappresentare “meglio” i dati, cioè visualizzare su mappa informazioni contenute in database che avevano anche un campo dedicato alla localizzazione geografica; in realtà mi sono resa conto che permette di studiare un aspetto trascurato nella maggior parte delle statistiche descrittive ma comunque sempre presente: la posizione geografica. Attraverso questo strumento possiamo pensare territorialmente e renderci realmente conto di come lo spazio influisca sulla conoscenza approfondita di un fenomeno.

In un'epoca nella quale non esiste ancora il teletrasporto ma si proclama la scomparsa di ogni barriera è forse necessario, anzi obbligatorio studiare su dati georiferiti.

APPENDICE
IMMAGINI RELATIVE A CAP. 1

FIGURA 1

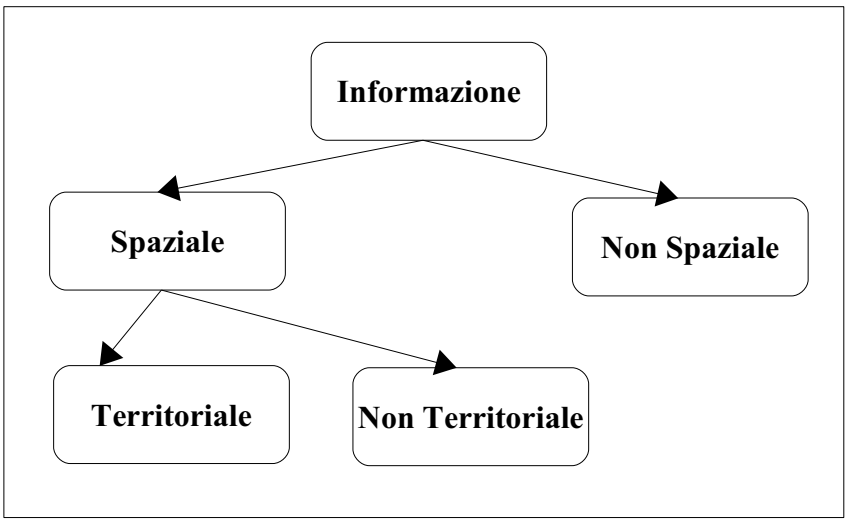


Il ruolo centrale del Sistema Informativo all'interno di un'organizzazione

FIGURA 2

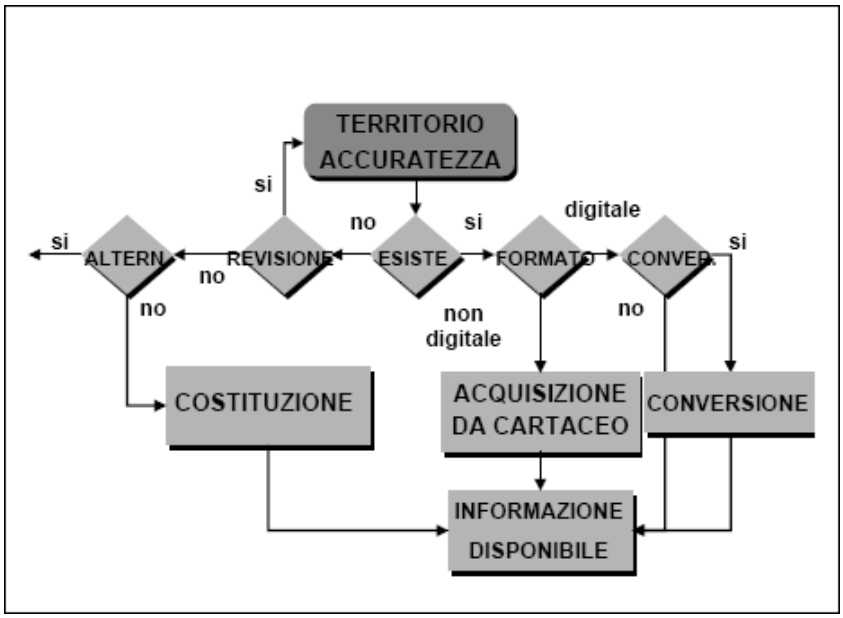
La rappresentazione della struttura di un'organizzazione secondo Anthony

FIGURA 3



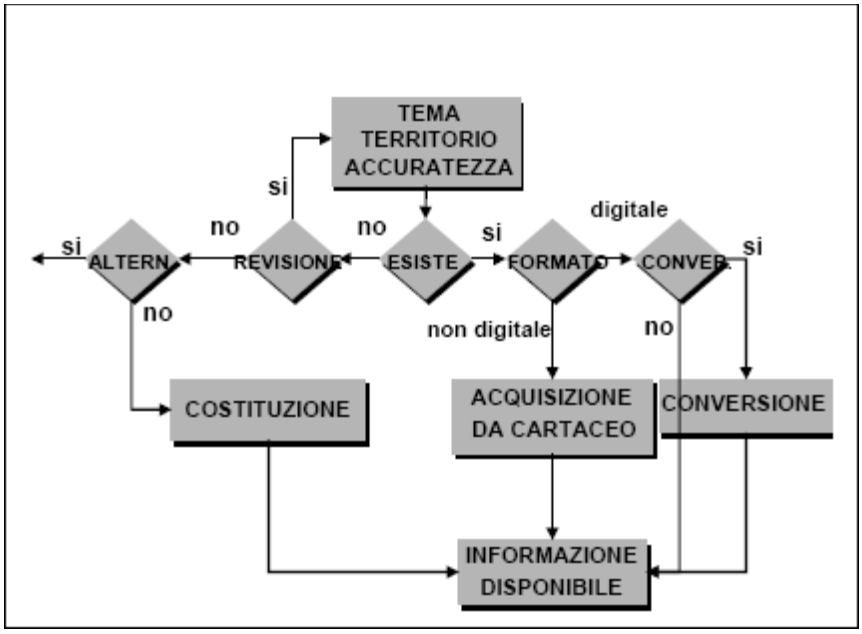
Tassonomia dell'informazione territoriale

FIGURA 4



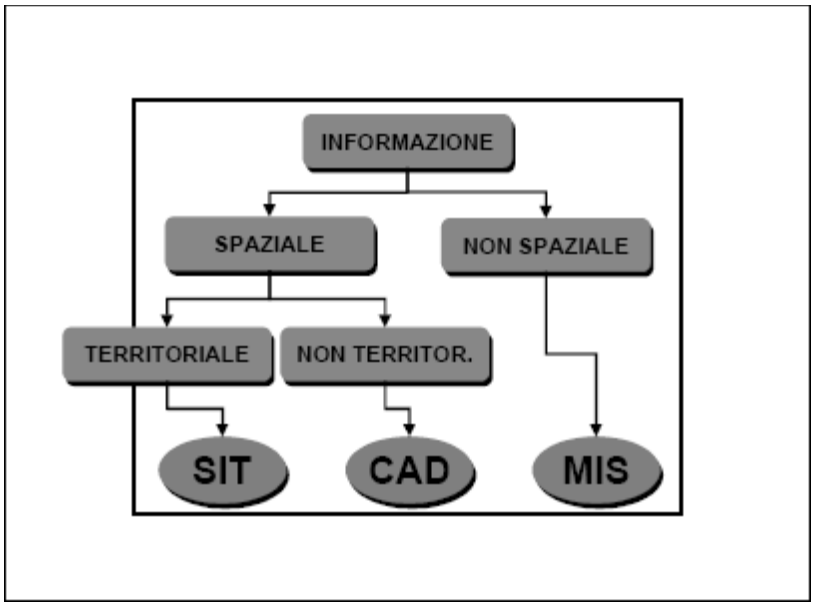
Schema a blocchi esemplificativo del processo di reperimento di informazione territoriale di base

FIGURA 5



Schema a blocchi esemplificativo del processo di reperimento di informazione territoriale tematica

FIGURA 6



Classificazione dei diversi Sistemi Informativi in relazione all'informazione

FIGURA 7

7	66	70	67	26	24	30	30	32	28
7	8	77	77	27	26	25	28	27	27
7	9	75	80	89	24	24	28	28	23
7	10	11	83	90	24	27	27	29	26
8	9	90	82	22	24	30	32	26	28
8	9	86	87	24	26	31	32	29	30
8	8	80	83	75	26	27	29	30	31
8	8	10	77	67	25	27	28	31	29
9	10	11	11	22	24	25	27	29	28
7	7	11	10	10	27	25	25	24	21

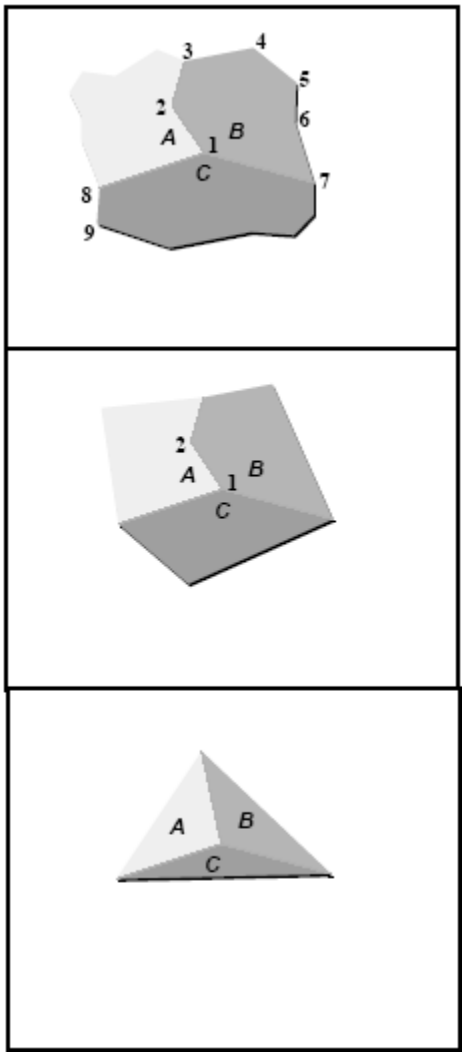
In questa immagine ogni pixel indica un valore di radianza

FIGURA 8

1	2	2	2	3	3	3	3	3	3
1	1	2	2	3	3	3	3	3	3
1	1	2	2	2	3	3	3	3	3
1	1	1	2	2	3	3	3	3	3
1	1	2	2	3	3	3	3	3	3
1	1	2	2	3	3	3	3	3	3
1	1	2	2	2	3	3	3	3	3
1	1	1	2	2	3	3	3	3	3
1	1	1	1	3	3	3	3	3	3
1	1	1	1	1	3	3	3	3	3

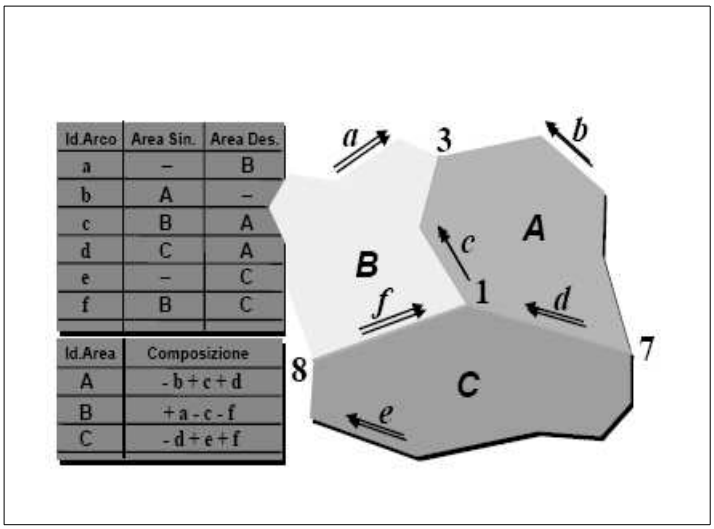
Dopo aver classificato i pixel quelli adiacenti con lo stesso valore sono stati accorpati.

FIGURA 9



Semplificando l'immagine attraverso l'eliminazione progressiva di vertici ritenuti “secondari” si elimina l’informazione geometrica e si lascia quella strutturale, ovvero la relazione di confine tra le aree.

FIGURA 10

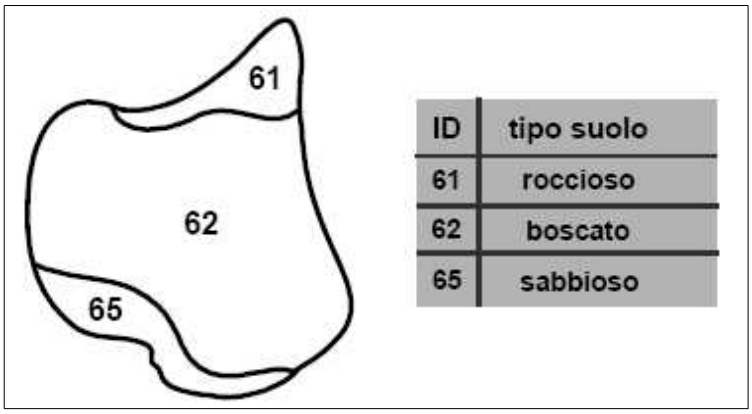


Id.Arco	Area Sin.	Area Des.
a	–	B
b	A	–
c	B	A
d	C	A
e	–	C
f	B	C

Id.Area	Composizione
A	- b + c + d
B	+ a - c - f
C	- d + e + f

Attraverso l'applicazione delle regole per la codifica degli archi si sono costruite la tabella degli archi e quella delle aree.

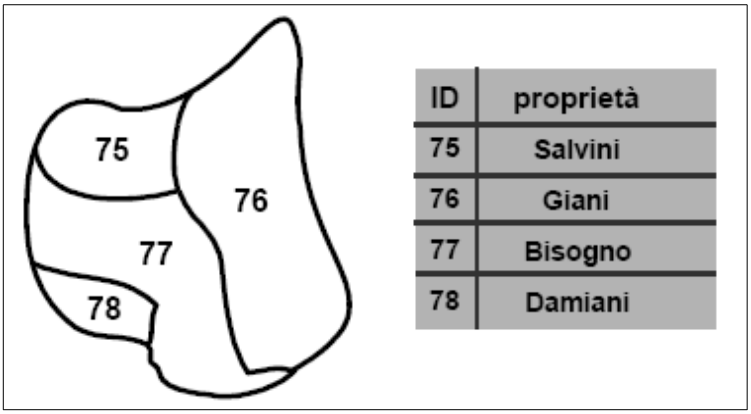
FIGURA 11



ID	tipo suolo
61	roccioso
62	boscato
65	sabbioso

Il primo dei due strati originari: copertura del suolo.

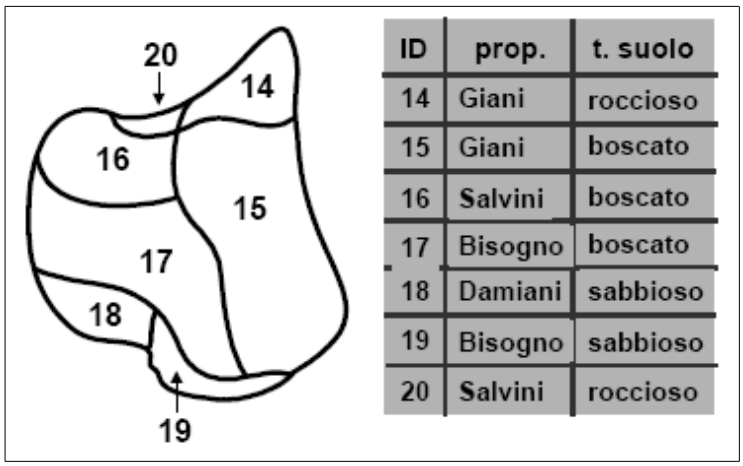
FIGURA 12



ID	proprietà
75	Salvini
76	Giani
77	Bisogno
78	Damiani

Il secondo dei due strati originari: carta delle proprietà.

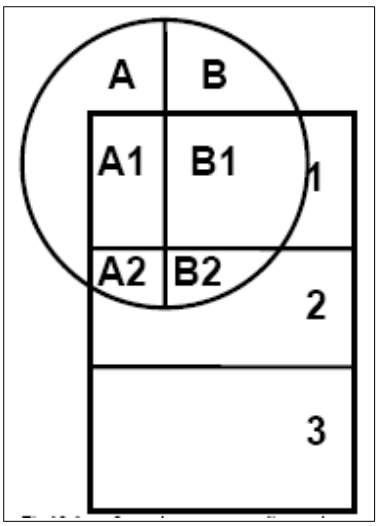
FIGURA 13



ID	prop.	t. suolo
14	Giani	roccioso
15	Giani	boscato
16	Salvini	boscato
17	Bisogno	boscato
18	Damiani	sabbioso
19	Bisogno	sabbioso
20	Salvini	roccioso

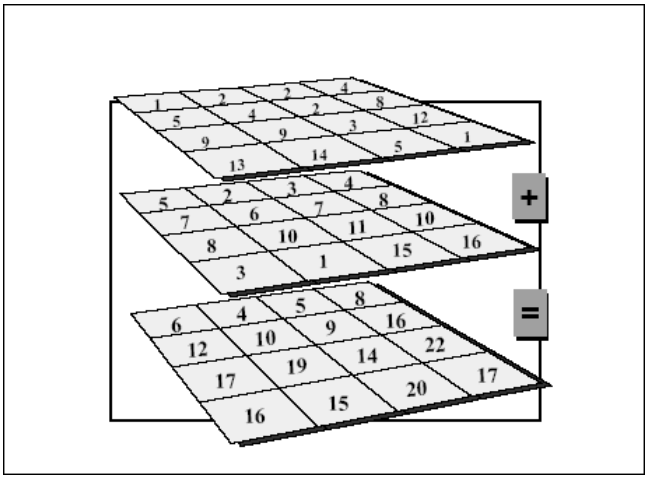
Lo strato finale, risultante dall'operazione di incrocio nel quale si evidenzia che ogni area ha un solo proprietario ed è composta da un unico tipo di suolo.

FIGURA 14



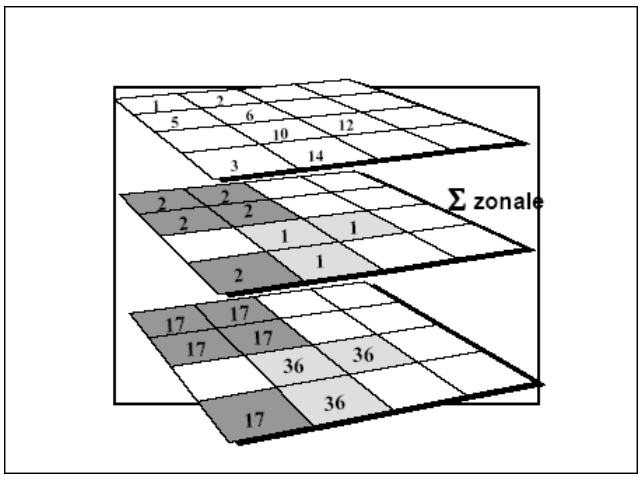
Incrocio tra aree di strati con copertura non coincidente: gli strati sono stati sovrapposti sulla base del loro reale posizionamento sul terreno.

FIGURA 15



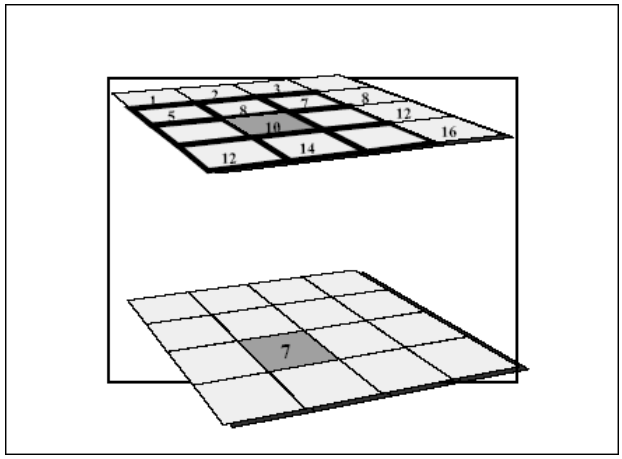
Situazione di perfetta corrispondenza in cui devono trovarsi i diversi strati informativi per potervi eseguire operazioni di map algebra. Questo implica che gli strati informativi devono avere le stesse dimensioni, e daranno uno strato-risultato con le stesse dimensioni.

FIGURA 16



Applicazione dell'operatore somma zonale. Nel primo strato sono presenti gli elementi da sommare, nel secondo strato, invece, vengono identificate le zone nelle quali eseguire l'operazione. Se si calcola, per esempio, la somma delle celle del primo layer appartenenti alla zona 2, così come definita dal secondo layer, si ottiene il valore 17 attribuito, nel layer risultato, a tutte le celle della zona 2.

FIGURA 17



Applicazione della media focale. Nel layer di input viene scelto un elemento e definito un suo intorno, evidenziato dalle caselle con il bordo più spesso. Nel layer di output alla cella prescelta viene attribuita la media aritmetica dei valori delle celle dell'intorno definito in precedenza

BIBLIOGRAFIA

Arbia G.(1990), "Impiego dei sitemi geografici informativi nelle rilevazioni ufficiali", ISTAT *Avanzamenti metodologici e statistice ufficili. Atti delle prime giornate di studio*, Roma, pp.113-128

Arbia G.(1992), "Campionamento areale in presenza di correlazione spaziale", SIS *Atti della XXXVI Riunione scientifica*, Pescara pp.311-321

Arbia G., Espa G. (1995), "Verso un approccio statistico al geomarketing: alcune linee guida ed un caso di studio", *Quaderni di Statistica e matematica applicata alle Scienze Economico-Sociali*, 1, pp.55-81

Arbia G.(1996), *Analisi econometrica di dati spaziali*, Università degli Studi "G. D'Annunzio", Pescara

Benedetti R., Espa G, (1998) "Stime potenziali di mercato per piccole aree", SIS *La statistica per le imprese: atti del Convegno*, Torino, pp.335-349

Bonfatti F. a cura di (1988) Elaborazione automatica dei dati geografici, Masson, Milano

Brovelli M.A., *Dispensa relativa al corso Sistemi Informativi Territoriali* (GIS), disponibile all'url:http://geomatica.como.polimi.it/corsi/sist_informativi/

Carfagna E., (2000) "Note sull'uso dell'autocorrelazione spaziale nel disegno campionario areale", SIS *Atti della XL riunione scientifica*, Firenze, pp. 169-182

Crescenzi F., Lipizzi F. (2000) "Un modello per l'accuratezza posizionale nell'ambito dei GIS", SIS *IL Riunione scientifica*, Firenze, pp.751-754

Crescenzi F. (2002) "I sistemi informativi geografici e la qualità del dato statistico", ISTAT *Atti della VI conferenza nazionale di statistica*, Roma

Cognolat M., *GIS: sistemi informativi per la gestione di dati cartografici*, Tesi di diploma, A.A. 1998-1999

Di Dio Roccazzella M, (2002) " Il geomarketing a supporto delle attività di vendita e comunicazione", *Disegno Digitale e Design* n.718, pp. 132-142

Giupponi C., Rigatti Luchini S., (1992) "Inquadramento generale del software per l'elaborazione dei dati telerilevati e le analisi statistiche", SIS *Atti della XXXVI riunione scientifica*, Pescara, pp.101-107

Giupponi C. , Rigatti Luchini S. (1996) *Metodologie informatiche per l'interpolazione spaziale di dati agrometereologici*, Regione del Veneto

Grassi Conti Puorger A. (1992) "Considerazioni sulle misure di

autocorrelazione spaziale", SIS *Atti della XXXVI riunione scientifica*, Pescara, pp. 77-83

Mogorovich P., Rumor M., *Dispensa relativa la corso Sitemi Informativi Territoriali*, disponibile all'url: http://sit.iuav.it

Napolitano P., Zani S., (1990) "Problemi di rilevazione e di presentazione dei dati spaziali",ISTAT *Avanzamenti metodologici e statistice ufficili. Atti delle prime giornate di studio*, Roma, pp.87-108

Pilotti L., Sedita S.R., (2002) "Competizione spaziale e fattori di localizzazione: un'applicazione strategica di geomarketing", *I&D* n.3, pp. 15-28

Vaccari C. (2000) "Telerilevamento e sistemi GIS", ISTAT *Atti della IV Conferenza nazionale di statistica*, Roma, pp.33-35

Zatelli P., *Dispensa relativa al corso Cartografia numerica e GIS*, disponibile all'url: www.ing.unitn.it/~zatelli/cartografia_numerica.html