

Trilobio Malisano

**Uso dei metodi di Statistica Spaziale in ambienti GIS:
alcune applicazioni epidemiologiche**

Marco Geraci, Matteo Bottai

Working Paper

CNUCE – B4 – 2002 – 002

Gennaio 2002

Sommario

1. INTRODUZIONE	3
2. INTEGRAZIONE IN AMBIENTE GIS	6
3. ALCUNE APPLICAZIONI DI STATISTICA SPAZIALE IN EPIDEMIOLOGIA	9
4. STIMA PARAMETRICA E NON PARAMETRICA DI UN INDICE LICHENICO E VISUALIZZAZIONE SU GIS	12
4.1. DATI	12
4.2. OBIETTIVI DELL'ANALISI	13
4.3. METODI STATISTICI UTILIZZATI	15
4.4. INTEGRAZIONE S-PLUS 4.5 E ARCVIEW 3.1 IN MODALITÀ LOOSE COUPLING	18
5. CONSIDERAZIONI FINALI	20
APPENDICE	26
RIFERIMENTI BIBLIOGRAFICI.....	27

1. Introduzione

“... [I] Sistemi Geografici Informativi...sono tanto rilevanti per l’analisi spaziale quanto le invenzioni del microscopio e del telescopio lo furono per la scienza, il computer per l’economia, e la stampa per la diffusione dell’informazione.” (Department of the Enviroment 1987, p. 8).

Negli ultimi dieci anni circa, la crescente rilevanza della Statistica Spaziale come metodo si è mostrata in un continuo fiorire di applicazioni in numerosi settori scientifici, caratterizzati dall’esigenza comune di dare un senso ed una utilità alle implicazioni, di natura stocastica, scaturenti dalla georeferenziazione dei fenomeni. La stessa disponibilità dei dati variamente raccolti con dettaglio sempre maggiore, i sistemi informatici di gestione e di descrizione e la potenza dei calcolatori hanno costituito un catalizzatore per l’affermarsi di un approccio spaziale.

Come scienza, tuttavia, la statistica spaziale ha posto dei seri problemi di riduzione a sistema codificato ed organizzato di tecniche e teorie che sono apparse in letteratura nelle ultime cinque decadi circa e che contano a tutt’oggi pochi anche se validi testi di riferimento (Cressie, 1993; Ripley, 1991; Griffith, 1988).

Parallelamente, la nascita dei Sistemi Geografici Informativi, inizialmente concepiti come supporto delle Scienze Geografiche, ha ben presto catturato l’attenzione di chi ha intravisto in esso le potenzialità dei suoi principali punti di forza: immissione, editing, gestione e visualizzazione su mappe dei dati geocodificati.

Sin dai primissimi anni novanta, fine ottanta, ci si è posti nella prospettiva di un’integrazione, alquanto difficoltosa, tra l’analisi spaziale (nota come *Spatial Data Analysis* – SDA, termine usato in senso lato e comprendente sia gli aspetti esplorativi, o *data-driven*, sia quelli confermativi, o *model-driven*) e i sistemi geografici territoriali.

Goodchild *et al.* (1992) già riportavano i risultati di un convegno tenutosi a Sheffield tra il 18 e il 20 Marzo 1991, circa i vantaggi scaturenti da un collegamento più stretto tra GIS e analisi spaziale (si veda anche Bailey, 1994), le specifiche aree dell’analisi che sarebbero dovute essere considerate e come tale collegamento, da un punto di vista informatico, si sarebbe potuto fare nel concreto. In particolare, i partecipanti al congresso suggerivano un uso di tecniche descrittive implementate direttamente in GIS per i “non-specialisti” (*full integration*), e di un collegamento basato su files di scambio tra GIS e pacchetti statistici per chi, dell’analisi in sé, ne facesse un obiettivo (*loose coupling*). E sempre sull’integrazione tra modellazione, anche se specificatamente ambientale, e GIS, l’*International Journal of Geographical Information Systems* ne ribadiva l’importanza con un uscita speciale nel numero 4 del volume 9 (1995).

Sul piano applicativo e conoscitivo, purtroppo, l’emergere di problemi propri dell’analisi spaziale, come l’autocorrelazione spaziale, i problemi di confine, il problema dell’unità areale modificabile, ha sicuramente giocato un ruolo negativo

nella diffusione di tecniche di questo tipo di analisi in ambito GIS che Fotheringham e Rogerson (1993) avevano già dettagliatamente illustrato. Nello stesso articolo, gli autori indicavano il potenziale contributo di un sistema GIS allo sviluppo di soluzioni per tali problemi e la strategia preferibile per avvicinare gli utenti GIS all'uso dell'analisi spaziale.

Ad oggi gli sviluppi più significativi si sono avuti indubbiamente sul piano descrittivo dell'analisi spaziale, mentre su quello predittivo sono state privilegiate tecniche e modelli parametrici.

I contributi più recenti sul tema (si veda il volume 2, parte 1, 2000 del *Journal of Geographical Systems*) tentano di far luce sullo stato attuale di tale integrazione e, soprattutto, di cosa quest'ultima necessiterebbe per uno sviluppo ulteriore. Fra gli altri, Boots (2000) spiega qual'è stato il ruolo del GIS e quale potrebbe essere in futuro come mezzo per la promozione dell'analisi spaziale. In particolare, si sostiene la necessità di sviluppare nuovi testi di supporto, maggiori interazioni interdisciplinari e software integrati. Jacquez (2000) sposta un particolare riflettore sull'epidemiologia spaziale (ambientale) e il ruolo del GIS sulla sua diffusione e affermazione. L'autore addita come maggiori limiti che tuttora ne impediscono una crescita, la mancanza, in un sistema territoriale, di una visione dinamica dei fenomeni spaziali e di modelli basati su processi che traducano in risultati epidemiologici i dati spaziali.

La natura interdisciplinare del metodo scientifico, del campo applicativo e del supporto tecnologico spingono ovviamente verso un'integrazione che sviluppa sinergie positive ed insieme difficoltà di inserimento.

Se da un lato, però, le sfide che si prospettano agli studiosi della comunità statistica e di quella della Scienza Informativa Territoriale sono affascinanti e ricche di notevoli linee di approfondimento, dall'altro, l'integrazione tra GIS e Statistica Spaziale da un punto di vista informatico rappresenta uno campo comune dove, se non altro, sperimentare e migliorare modelli già esistenti. A favore si registrano notevoli sforzi verso:

- una maggiore sperimentazione di progetti volti alla diffusione di dati GIS, grezzi ed elaborati, su web e su wap (GIS data server);
- avanzate ricerche di paradigmi e tecnologie, come nel caso della realtà virtuale, per la visualizzazione di dati spaziali, soprattutto in tre dimensioni e/o in contesti dinamici (vedi il numero speciale dell'*International Journal of Geographical Information Systems*, volume 13 numero 4);
- la creazione di sinergie sistemiche tra la statistica ufficiale, nazionale ed internazionale (soprattutto europea), e la componente GIS, in termini di raccolta, elaborazione e condivisione dei dati (a titolo indicativo, lo speciale apparso sulla rivista *Mondo GIS* settembre 98 e all'indirizzo <http://www.istat.it/Primpag/quintaconf/workshop/esp.html>);
- lo sviluppo di programmi e librerie, sia GIS che statistici, disponibili al pubblico.

In questo lavoro verranno presentate le attuali possibilità di integrazione informatica tra GIS ed analisi spaziale (capitolo 2) ed alcune applicazioni più significative nell'ambito epidemiologico, di cui una appositamente sperimentata, al fine di mostrare l'indubbio vantaggio che un'analista statistico trae da tale integrazione in termini di potenzialità di visualizzazione GIS (capitolo 3), e l'importanza di condurre l'analisi con l'ausilio della moderna metodologia statistica.

2. Integrazione in ambiente GIS

Visualizzazione, come momento di percezione visiva di un insieme di dati e come supporto all'analisi pre e post esplorativa, gestione efficiente di dati geo codificati su base *vector* o *raster*, esecuzione d'importanti operazioni spaziali di selezione e manipolazione, rappresentano caratteristiche proprie di un sistema informativo territoriale.

Ora, data la superiorità dei pacchetti statistici riguardo gli algoritmi d'analisi e del GIS per le proprietà anzidette, sembrerebbe naturale sostenere un'integrazione in funzione di un bacino d'utenza di tipo accademico, avvalendosi delle caratteristiche per le quali è nato un software e della sua flessibilità d'impiego. Non dimentichiamo, infatti, che la continua produzione letteraria teorica deve contare sulla possibilità di un'immediata sperimentazione applicativa e computazionale di risultati e tecniche volte alla risoluzione dei problemi tuttora aperti. Se, tuttavia, l'integrazione venisse effettuata a favore di utenti meno esperti, la semplificazione delle procedure d'analisi e d'inferenza porterebbe probabilmente a risultati assai scadenti. Tale trade-off, ovviamente, non può risolvere obiettivi fra loro incompatibili.

Lo stato attuale dei software legati al trattamento dei dati geografici è decisamente variegato e il collegamento può avvenire in diversi modi (si veda in Appendice A per una rassegna indicativa dei programmi adatti all'analisi statistica dei dati spaziali):

- programmi specifici: software progettati per eseguire una o più parti dell'analisi spaziale ma che sfruttano esclusivamente risorse interne, sia per quanto riguarda gli algoritmi d'analisi, sia per ciò che concerne l'eventuale struttura data-relazionale e visualizzativa di tipo GIS. Non è esclusa la possibilità di importare ed esportare file in formato standard o comune al GIS;
- collegamento libero (*loose coupling*): il collegamento tra GIS e pacchetto statistico, risorse ben distinte, si basa su uno scambio di file di formato standard per esportare l'output d'analisi di un programma, da importare come input d'analisi per l'altro. Può avvenire per mezzo di un'estensione che modifica l'interfaccia grafica con l'inserimento di un menù che facilita le operazioni di scambio, minimizzando l'intervento dell'utente;
- collegamento stretto (*close coupling*): in questo caso il programma statistico viene lanciato da GIS o viceversa. Tutte le funzionalità sono utilizzabili;
- *system enhancement*: macros e librerie di funzioni GIS compilate e caricate dinamicamente nell'ambiente eseguibile statistico (e viceversa). Sono richiamabili dall'utente e adattabili alle esigenze dell'utilizzatore. Funzionalità limitate a quelle definite;
- programmi integrati: moduli di analisi realizzati direttamente dal produttore di GIS, secondo modalità e facilitazioni proprie della casa.

Nella tabella 1, sono state riportate le possibili modalità di collegamento ed esempi ad esse associate, e alcune valutazioni in merito alle caratteristiche d'uso che comportano tali soluzioni, posto che la scala ha nel nostro caso un valore sostanzialmente ordinale. Il punto principale è, infatti, sapere in che rapporto stiano fra loro i programmi così categorizzati. Ad esempio è naturale considerare che il collegamento libero o stretto portano con sé, solitamente, il vantaggio di flessibilità elevata e la necessità di conoscenze maggiori a causa della compresenza di uno strumento GIS e di uno statistico, diversamente da quanto accade nel caso di una integrazione completa, dove l'utente è maggiormente guidato nell'analisi per la minore incidenza della componente statistica.

L'esistenza di una convergenza tra GIS e analisi spaziale, richiesta soprattutto da potenziali utenti GIS di settori scientifici diversi, si è manifestata, oltre che nella produzione informatica di soluzioni diverse, anche in quella letteraria, di cui, a titolo indicativo, verranno di seguito segnalati alcuni lavori.

Modalità	Livello di integrazione	Competenze necessarie	Flessibilità	Esempi
<i>Programmi specifici</i>	nullo	basse	bassa	GAM Stat! GS+
<i>Collegamento libero</i>	basso	alte	alta	SpaceStat for ArcView SAGE
<i>Collegamento stretto</i>	medio	alte	alta	S+ for ArcView S+Grassland Link
<i>System enhancement</i>	medio	medie	media	-
<i>Programmi integrati</i>	alto	basse	bassa	Geostatistical Analyst for ArcGis

Tabella 1. Modalità di collegamento.

Il *Journal of Geographical Systems* presenta diversi articoli relativi alle possibili soluzioni informatiche nella parte 3 del volume 2, 2000.

Aspetti tecnici e concettuali dell'implementazione di S+ in due diversi ambienti GIS (Arcview, *vector-oriented*, e Grassland, *raster-based*) li ritroviamo in Bao *et al.* (2000).

Anselin (2000) descrive funzionalità e architettura di SpaceStat, della sua estensione per ArcView e del modulo DynESDA, il quale facilita la visualizzazione dinamica dell'analisi esplorativa su ArcView (si veda anche Anselin, Bao, 1997).

Su SAGE, Haining, Wise e Ma (2000) per gli aspetti progettuali e la filosofia d'analisi statistica sottostante al programma; Haining, Wise e Signoretta (2000) per gli aspetti grafici.

Una segnalazione particolare va fatta per i lavori di Bivand, i quali costituiscono un ottimo riferimento per programmi e linguaggi disponibili al pubblico.

L'implementazione di funzioni di analisi spaziale utilizzando il linguaggio R si trova in Bivand e Gebhardt (2000) , mentre in Bivand (1999, 2000) e Bivand e Neteler (2000) l'integrazione del linguaggio R e del sistema GIS GRASS.

Quest'ultimo è disponibile su <http://www.baylor.edu/grass/>, dove tra l'altro è presente una sezione dedicata all'analisi geostatistica in ambiente GIS (<http://www.baylor.edu/grass/statsgrass/index.html>); il linguaggio R si trova su <http://www.r-project.org/>.

3. Alcune applicazioni di statistica spaziale in epidemiologia

La crescente sensibilità rivolta alla tutela della salute umana, e lo sviluppo notevole dei sistemi tecnologici, costituiscono fattori sicuramente positivi per la maturazione di tecniche, metodi e approcci interdisciplinari per la valutazione dei rischi ambientali in rapporto agli aspetti epidemiologici che questi comportano. Indagini epidemiologiche mirate o studi basati su dati censuari coinvolgono un numero elevato di soggetti e professionalità che vanno da enti pubblici, amministrativi e di ricerca, ad unità sanitarie locali e nazionali, da statistici, matematici a epidemiologi ed informatici. Appare non trascurabile, per questo motivo, la forte esigenza di sviluppare procedure standard in ambienti software *user friendly*, laddove possano costituire uno strumento utile d'intervento focalizzato da parte di autorità locali e nazionali in tema di sanità e salvaguardia dell'ambiente.

Il monitoraggio, la valutazione dei fattori di rischio, la mappatura della malattia e di agenti inquinanti e i risultati dell'analisi, al fine di un controllo e di una prevenzione più efficaci, dovrebbero contare su un supporto GIS - per una discussione sul ruolo del GIS nell'ambito dell'analisi spaziale e dell'epidemiologia (sinteticamente, Epidemiologia Spaziale), si veda Jacquez (2000); Croner *et al.* (1996) per una introduzione ai punti principali di un GIS e per il suo uso in progetti su scala nazionale per la tutela della salute; Briggs ed Elliott (1995) sull'uso del GIS negli studi sull'ambiente e la salute ed esempi applicativi.

Emergono in questo senso il rilievo spaziale dell'approccio e l'importanza dell'informazione geografica. La Statistica Spaziale, da una parte, con le sue molteplici applicazioni, costituisce un valido metodo scientifico per trattare i problemi che sorgono ogni qual volta lo studio implica la georeferenziazione di un fenomeno, a prescindere dai limiti che tuttora ne impediscono uno sviluppo ed una diffusione su larga scala. I GIS, dall'altra, rappresentano un naturale "sbocco informatico" per quel che concerne la gestione e la visualizzazione di dati geografici e delle loro elaborazioni.

E' bene sottolineare che le problematiche e gli approcci legati alla gestione del territorio per i fini anzidetti sono numerosi e altrettanto numerose sono le analogie riscontrabili in settori applicativi diversi. Un esempio indicativo è l'analisi del crimine e l'individuazione di "hot spots" (concentrazioni) nelle aree urbane o la comparazione di tassi di delinquenza fra paesi diversi, di cui un congruo numero di applicazioni è presente nella letteratura di Econometria Spaziale (si veda, Baller *et al.*, 2000; Ratcliffe, McCullagh, 1999; Anselin, Florax, 1995; Griffith, 1988). Il passaggio successivo è quello di modellare la presenza del crimine individuando fattori, in questo caso, di natura economica e sociale o di esplicitare le cause di tassi differenti fra vari paesi, considerando ovviamente la componente informativa di natura spaziale.

Uno dei punti principali nella valutazione del rischio per l'esposizione a fattori inquinanti, è sicuramente l'individuazione di aggregazioni di malattie e la

modellazione di un possibile legame della distribuzione spaziale della malattia con la sorgente inquinante ritenuta responsabile.

Un'applicazione significativa in un momento in cui l'integrazione GIS - analisi spaziale era solo auspicata è quella di Diggle, Gatrell e Lovett (1990). Essi esplorano con un approccio parametrico la possibile relazione tra la distribuzione del cancro alla laringe e la posizione di un inceneritore in un caso empirico. Oltre al riferimento d'approccio seguito da Openshaw in GAM, viene introdotto il crescente interesse da parte di statistici, epidemiologi e geografi che si sviluppa di pari passo con la disponibilità di dati epidemiologici e l'avvento dei sistemi informatici territoriali. L'accento si sposta dalle tecniche di individuazione delle aggregazioni di malattie, alla possibilità di verificare, attraverso modelli, ipotesi sull'esistenza di "associazioni ambientali" tra le aggregazioni e fattori esterni - per una rassegna di metodi d'analisi dei *point pattern* per la modellazione spaziale di eventi nell'intorno di sorgenti d'inquinamento, si veda Lawson e Waller (1996); per l'analisi spaziale condotta utilizzando metodi di *clustering* in ambito GIS, Murray, Estivill-Castro (1998) e Murray (1999), e la mappatura dell'inquinamento urbano tramite GIS, Briggs et al (1997); per un'introduzione recente al tema della statistica applicata all'epidemiologia ambientale, Thomas (2000).

Hjalmars *et al.* (1996) ricorrono al GIS e ad una tecnica statistica di *clustering* per individuare concentrazioni di casi di leucemia infantile in Svezia.

In anni recenti, trasmettitori radio e televisivi sono stati oggetto di numerose discussioni circa la loro dannosità per la salute umana a causa dell'insorgere di vari casi di tumore nelle vicinanze degli stessi. Dolk *et al.* (1997a, 1997b) in due studi paralleli, applicano test statistici per verificare la associazione tra la presenza di un trasmettitore radio e casi di varie tipologie di cancro localizzati negli intorni. Considerano in un caso, un singolo trasmettitore (Sutton Coldfield), e nell'altro, tutti i trasmettitori ad alta potenza sul suolo britannico.

Un uso particolare del GIS è quello fatto da Childs *et al.* (1998) che applicano una regressione logistica per stimare il rischio di infestazione da ratti nella città di New York, sottolineando come un'analisi combinata di tipo geografico ed epidemiologico possa aiutare a focalizzare attività di controllo.

Wall e Devine (2000), esaminando la mortalità per cancro ai polmoni nello stato dell'Ohio, mostrano la capacità di un GIS per lo *smoothing* dei tassi d'incidenza, per l'individuazione di aggregazioni inusuali di casi, e per la valutazione statistica delle inferenze condotte su trend spaziali.

Altrettanto importante è la possibilità di predire il rischio di malattie infettive a partire dalla conoscenza di fattori ambientali e climatici legati a tale rischio. Kleinschmidt *et al.* (2000) per il controllo della diffusione della malaria, tramite la mappatura del rischio ottenuta con un modello di regressione logistica ed un affinamento tramite kriging. Sempre sulla mappatura della malaria, tramite *Generalized Linear Mixed Model*, Kleinschmidt *et al.* (2001).

Zhu *et al.* (2000), utilizzano Arc/Info come GIS per mappare le stime del rischio di asma associato alla densità del traffico, nel caso speciale di dati spazio-temporali non allineati. L'approccio è di tipo gerarchico. Lo è anche quello utilizzato da

Mugglin *et al.* (2000) per l'inferenza su dati multivariati non allineati. Il metodo, dettagliatamente illustrato, viene sperimentato su un dataset complesso relativo alle misurazioni di esposizione inalatoria al radon (Ohio).

4. Stima parametrica e non parametrica di un indice lichenico e visualizzazione su GIS

Fra le sorgenti lineari di inquinamento, si possono naturalmente ricomprendere le strade di grande comunicazione ad alto traffico. Di sicuro interesse è la conoscenza del livello d'inquinamento in corrispondenza di punti non osservati e la visualizzazione su mappa dei relativi valori.

I metodi per la stima di una superficie interpolante sono diversi e li possiamo classificare in parametrici e non parametrici. La scelta dipende da numerosi fattori, quali la disponibilità a priori di informazioni sul fenomeno e sulla zona di osservazione per indagini precedenti, il tipo di obiettivo conoscitivo che ci si propone, il grado di adattamento dei dati osservati alle ipotesi sottostanti l'approccio parametrico e, in un certo senso, la snellezza del metodo.

Come caso applicativo abbiamo considerato un campione di 69 osservazioni dell'indice di sviluppo della flora lichenica rilevate da una precedente indagine ambientale del Dipartimento delle Scienze Ambientali dell'Università di Pisa nel Comune di Cascina tra il 1991 e il 1992, una successiva sperimentazione dell'uso del GIS su questi dati e quelli ottenuti da un'indagine epidemiologica condotta nel medesimo periodo (Della Maggiore *et al.*, 2000) e, infine, un lavoro sugli stessi dati per la modellazione non parametrica della superficie interpolante e dell'errore di predizione spaziale (Bottai, 2001).

4.1. Dati

L'indice lichenico è ormai considerato un ottimo bioindicatore della qualità dell'aria e possiede il vantaggio di una facile modalità di rilevazione. Il luogo sotto indagine ambientale è un tratto di strada ad alto traffico veicolare, di circa 10 km, passante per diversi comuni, tra cui Cascina, Navacchio, San benedetto e Titignano. Le stazioni di rilevamento sono state considerate distribuite in modo sufficientemente uniforme e riportate su GIS con simbologia visibile in legenda (figura 1). Si potrà osservare la presenza di punti a distanze non eccessive dalla strada principale (via Tosco Romagnola).

Mappa delle stazioni di rilevamento

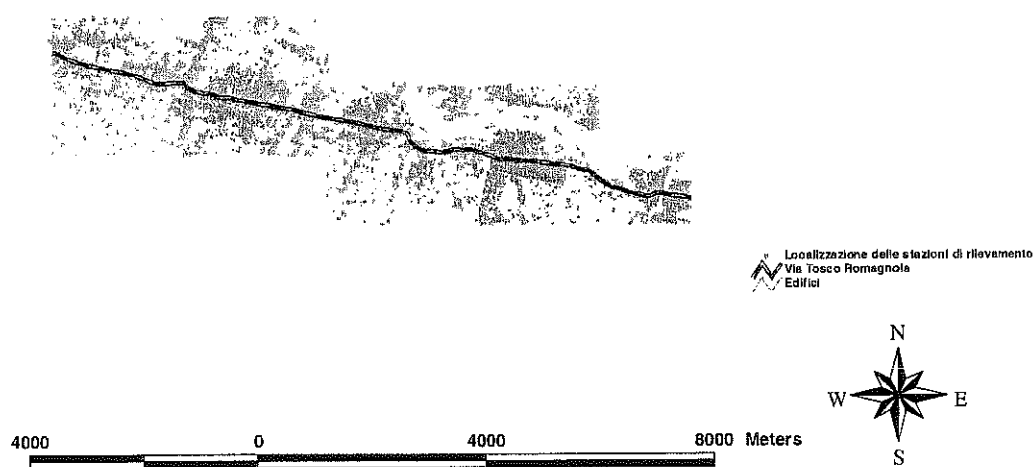


Figura 1. Via Tosco Romagnola nel tratto passante per Cascina, Navacchio, San Benedetto e Titignano.

4.2. Obiettivi dell'analisi

In della Maggiore *et al.* (2000), i dati lichenici sono stati sottoposti ad una analisi geostatistica (*variogram fitting*, *ordinary kriging*) condotta su S-Plus tramite il modulo S+SpatialStats (figura 2 e figura 3). Tramite Spatial Analyst sono stati interpolati, con *splines*, i valori lichenici. Successivamente, sono stati considerati i risultati al test Ab-BP-DNA (quest'ultimo valutato come *biomarker* dell'esposizione a sostanze ambientali cancerogene, Petruzzelli *et al.*, 1998) per soggetti non fumatori e stimate le mappe di rischio in corrispondenza dei centri abitati sempre con *splines*, sovrapponendo queste ultime alla mappa degli edifici e a quella della superficie lichenica.

Un possibile miglioramento ai fini di una maggiore chiarezza interpretativa è quello di usufruire delle analisi statistiche di S-Plus e visualizzare i risultati su ArcView per georeferenziare le predizioni attraverso la costruzione di temi. La mancanza di un controllo totale dell'analisi statistica condotta con Spatial Analyst rischia di portare a scelte arbitrarie riguardo i parametri dell'interpolazione e all'impossibilità di ricorrere a tecniche alternative che possono essere meglio

implementate per le caratteristiche dei dati osservati. Se da una parte, infatti, l'analisi esplorativa visiva è importante, dall'altra, molto dipende da come la superficie viene stimata. A questo proposito abbiamo considerato il kriging come tecnica parametrica e la regressione localmente ponderata (*Locally-weighted Regression*, nota come "loess") come controparte non parametrica, confrontando i risultati al variare del parametro di *span*, che può essere scelto a piacere e misura il grado di lisciamento della superficie interpolante (si veda il paragrafo 4.3).

L'utilità della componente statistica ed il ricorso a tecniche e modelli particolari come il *bootstrap* e lo *smoothing* non parametrico vengono dimostrati nel lavoro di Bottai (2001), dove mancando una visualizzazione GIS, si ripropone l'opportunità di un'integrazione.

Variogramma

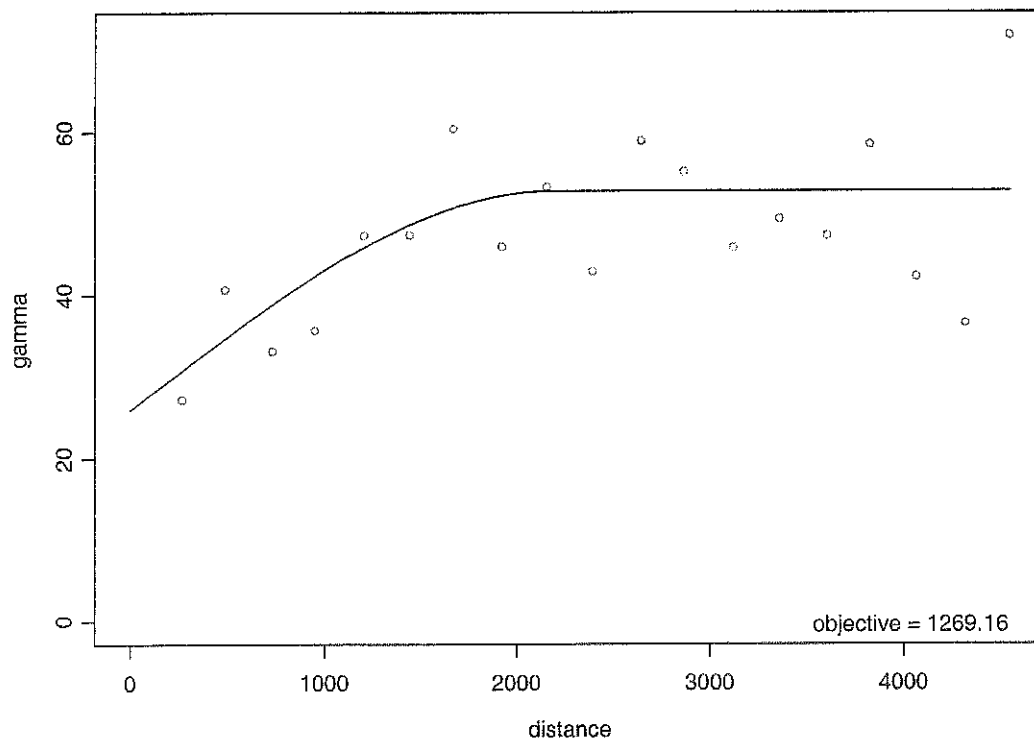


Figura 2. Variogramma empirico (cerchi) delle 69 osservazioni dell'indice lichenico e sovrapposizione della stima robusta del variogramma sferico (linea continua).

Interpolazione kriging

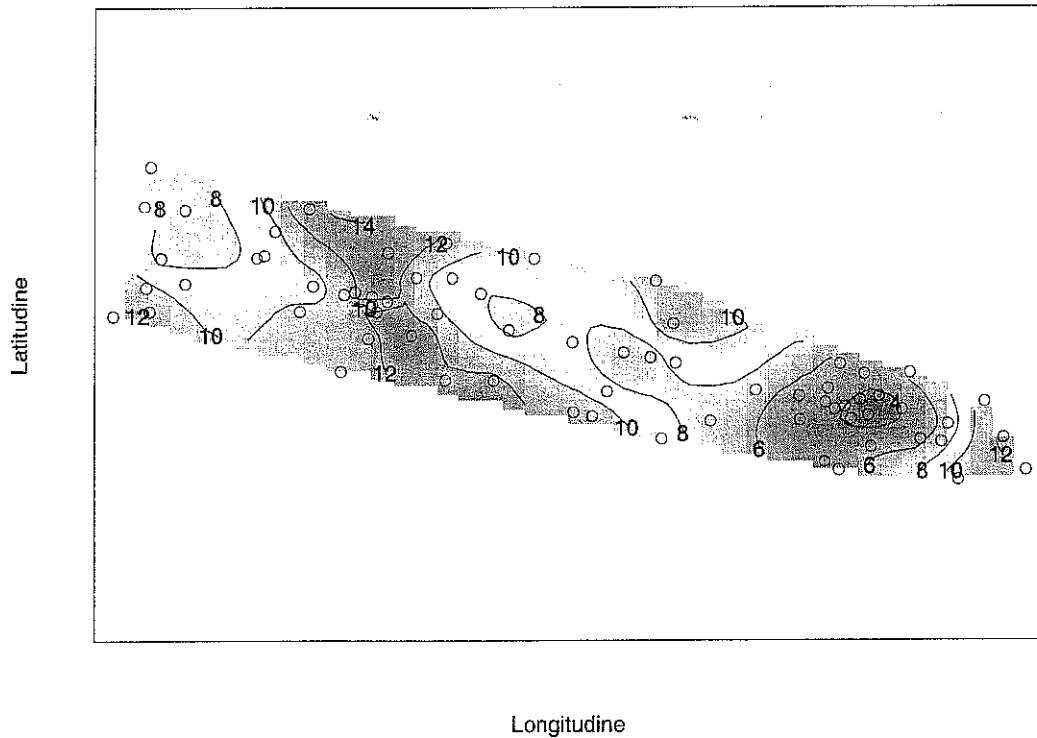


Figura 3. Kriging ordinario delle 69 osservazioni dell'indice lichenico. In grigio le aree inquinate.

4.3. Metodi statistici utilizzati

In questo paragrafo saranno definiti con notazione più precisa i metodi statistici utilizzati nell'analisi spaziale dei valori lichenici.

Kriging ordinario.

Sia $s \in \mathcal{R}^d$ una generica locazione in uno spazio euclideo d -dimensionale (solitamente bidimensionale) e supponiamo che Z sia una variabile casuale misurata in corrispondenza di n punti s_i , di coordinate (x_i, y_i) , con $i = \overline{1, n}$ (Cressie, 1993). Si lasci variare s nell'insieme indicizzante $D \subset \mathcal{R}^d$, così che generi il processo casuale:

$$\{Z(s); s \in D\}.$$

Introduciamo la funzione variogramma come

$$2\gamma(s_1 - s_2) \equiv \text{var}(Z(s_1) - Z(s_2)),$$

per tutti gli s_1 e s_2 in D . Se consideriamo solo le coppie di punti che distano di una grandezza h (lag) e supponiamo che il processo Z sia stazionario del secondo

ordine, uno stimatore naturale di $\gamma(\mathbf{h})$, dovuto a Matheron (1962), basato sul metodo dei momenti è

$$2\bar{\gamma}(\mathbf{h}) \equiv \frac{1}{|N(\mathbf{h})|} \sum_{N(\mathbf{h})} (Z(s_i) - Z(s_j))^2,$$

con $\mathbf{h} \in \mathfrak{R}^d$, dove $N(\mathbf{h}) \equiv \{(s_i, s_j) : s_i - s_j = \mathbf{h}; i, j = 1, K, n\}$ ed $|N(\mathbf{h})|$ è il numero di coppie distinte in $N(\mathbf{h})$.

Uno stimatore robusto (rispetto a valori estremi), proposto da Cressie ed Hawkins (1980), è dato da

$$2\bar{\gamma}(\mathbf{h}) \equiv \left\{ \frac{1}{|N(\mathbf{h})|} \sum_{N(\mathbf{h})} (|Z(s_i) - Z(s_j)|)^{1/2} \right\}^4 / (0.457 + 0.494/|N(\mathbf{h})|).$$

L'obiettivo è di predire valori in corrispondenza di punti non osservati, a partire dai dati osservati

$$\mathbf{Z} \equiv (Z(s_1), K, Z(s_n))'$$

Assunzioni:

i. di modello:

$$Z(s) = \mu + \delta(s),$$

dove μ , con $\mu \in \mathfrak{R}$, è una costante (*trend*) incognita, e $\delta(s)$, con $s \in D$, modella l'errore (correlato spazialmente) del processo casuale;

ii. di stimatore:

$$p(\mathbf{Z}; B) = \sum_{i=1}^n \lambda_i \cdot Z(s_i),$$

con $\sum_{i=1}^n \lambda_i = 1$ e $B \subset D$;

iii. di variogramma:

$$2\gamma(\mathbf{h}) = \text{var}(Z(s + \mathbf{h}) - Z(s))$$

con $\mathbf{h} \in \mathfrak{R}^d$.

Lo stimatore ottimale $p(\bullet; B)$ è quello che minimizzerà l'errore quadratico medio di predizione

$$\sigma_e^2 \equiv E(Z(B) - p(\mathbf{Z}; B))^2$$

nella classe degli stimatori lineari $\sum_{i=1}^n \lambda_i \cdot Z(s_i)$ che soddisfano $\sum_{i=1}^n \lambda_i = 1$.

Supponendo che $B = \{s_0\}$, per il kriging ordinario è necessario minimizzare la quantità

$$E \left(Z(s_0) - \sum_{i=1}^n \lambda_i \cdot Z(s_i) \right)^2 - 2 \cdot m \cdot \left(\sum_{i=1}^n \lambda_i - 1 \right)$$

rispetto a $\lambda_1, \dots, \lambda_n$, ed m (il parametro m è il moltiplicatore di Lagrange che assicura $\sum_{i=1}^n \lambda_i = 1$).

Così, i parametri ottimali $\lambda_1, \dots, \lambda_n$ possono essere ottenuti da

$$\lambda_0 = \Gamma_0^{-1} \cdot \gamma_0$$

dove

$$\begin{aligned} \lambda_0 &\equiv (\lambda_1, \dots, \lambda_n, m)' \\ \gamma_0 &\equiv (\gamma(s_0 - s_1), \dots, \gamma(s_0 - s_n), 1)', \\ \Gamma_0 &\equiv \begin{cases} \gamma(s_i - s_j), & i = 1, \dots, n, j = 1, \dots, n, \\ 1, & i = n+1, j = 1, \dots, n, \\ 0, & i = n+1, j = n+1, \end{cases} \end{aligned}$$

e Γ_0 è una matrice simmetrica $(n+1) \times (n+1)$.

Locally-weighted Regression (loess).

Consideriamo il caso di polinomi di primo grado, nonostante qualunque polinomio possa essere adattato localmente (Hastie, Tibshirani, 1990).

Sia $s \in \mathcal{R}^d$ una generica locazione in uno spazio euclideo d -dimensionale (solitamente bidimensionale) e supponiamo che Z sia una variabile casuale misurata in corrispondenza di n punti s_i , di coordinate (x_i, y_i) , con $i = 1, \dots, n$. Si lasci variare s nell'insieme indicizzante $D \subset \mathcal{R}^d$, così che generi il processo casuale:

$$\{Z(s) : s \in D\}.$$

Con alcuni passaggi, viene calcolato il lisciamento di $z(s_0)$, con s_0 generico punto, attraverso una retta localmente ponderata utilizzando i k punti più vicini:

- i. viene individuato l'insieme dei k punti più vicini di s_0 , indicato con $N(s_0)$ e definito intorno di s_0 ;
- ii. viene calcolata la distanza tra s_0 e il punto $s_i \in N(s_0)$ più lontano, $\Delta(s_0) = \max_{s_i \in N(s_0)} |s_0 - s_i|$;
- iii. sono assegnati a ciascun punto in $N(s_0)$ i pesi w_i , questi ultimi generati utilizzando la funzione *tri-cubica*

$$W\left(\frac{|s_0 - s_i|}{\Delta(s_0)}\right),$$

dove

$$W(u) = \begin{cases} (1 - u^3)^3, & \text{per } 0 \leq u < 1; \\ 0, & \text{altrimenti;} \end{cases}$$

- iv. relativamente all'intorno di s_0 , $N(s_0)$, si ottiene il valore stimato di Z in s_0 , $z(s_0)$, con il metodo dei minimi quadrati ponderati, con pesi calcolati in (iii).

Il termine *span* è riferito alla proporzione di punti più vicini ad s_0 , nel caso in cui il numero dei punti k venga espresso come percentuale del totale.

4.4. Integrazione S-Plus 4.5 e ArcView 3.1 in modalità loose coupling

Per realizzare l'obiettivo del nostro lavoro, abbiamo considerato S-Plus come supporto all'analisi statistica e ArcView 3.1 come supporto GIS. I moduli S+SpatialStats, S+ for ArcView e Spatial Analyst 1.1, i primi due per S+ ed il terzo per ArcView GIS, sono intervenuti nella fase di analisi statistica propriamente spaziale e di interpolazione dei valori generati. I passi sono stati così individuati.

1. Innanzitutto, sono stati preparati i dati. La disponibilità su GIS delle osservazioni dell'indice lichenico con le relative coordinate spaziali ha permesso l'esportazione di un file in formato DBASE successivamente importato su S-Plus. Le colonne principali riguardano una serie di coordinate longitudinali (colonna x), latitudinali (colonna y) e di misure corrispondenti dell'indice lichenico (colonna z), per un totale di 69 *records* nel dataset Licheni.

Stima Non Parametrica

2. Nel passo successivo è stato applicato un modello di regressione locale (loess) con *span* (proporzione dei valori utilizzati per lo *smoothing*) pari a 0.75, 0.50 e 0.30 alla relazione funzionale data da $z = f(x, y) + \varepsilon_{x,y}$, con $\varepsilon_{x,y}$ variabile casuale distribuita simmetricamente con media nulla. La funzione ha così generato 69 predizioni (*fit*) in corrispondenza dei punti di osservazione e i relativi errori standard (*se.fit*). L'interpolazione triangolare dei valori predetti ha generato un oggetto *list* contenente una griglia rettangolare di 1.600 (40x40) valori z , con ascisse e ordinate massime e minime corrispondenti agli estremi del campo di variazione delle x e delle y del dataset Licheni, e due vettori, x e y , di dimensione 40x1. Per *default* l'interpolazione è stata considerata nel convesso, dando luogo così a valori non definiti nei punti al di fuori. Si è evitato in tal modo di eseguire estrapolazioni difficilmente

valutabili. L'interpolazione è stata fatta sia per il valore di *fit* sia per l'errore standard, generando dunque oggetti di tipo *list* differenti. Con alcuni passaggi, dall'oggetto *list* è stata creata una matrice di tipo *data.frame* esportata in formato DBASE ed importata in ArcView come oggetto *table*. I punti nel convesso sono risultati 664. A questo punto l'estensione Spatial Analyst ha permesso l'interpolazione dei valori con il metodo *Inverse Distance Weighted*. Questa doppia interpolazione è stata necessaria per l'impossibilità di generare direttamente la superficie da S+ in ArcView.

Stima Parametrica

3. Al variogramma empirico del dataset Licheni, abbiamo adattato un modello sferico stimando i parametri con il metodo dei minimi quadrati per funzioni non lineari. E' stato ottenuto un *range* pari a 2189.67, *sill* 26.78 e *nugget* 25.99. L'applicazione del kriging ordinario e della funzione di predizione sui punti del convesso hanno generato direttamente un *data.frame* di 664 *records* per longitudine, latitudine, valore predetto e errore standard della predizione. I punti sono stati appositamente scelti per effettuare un confronto tra valori predetti con loess e valori predetti con kriging. L'esportazione e l'interpolazione su ArcView sono avvenute come in 2. Si nota, comunque, che per comodità di confronto dei valori, l'analisi geostatistica è stata anch'essa condotta su S-Plus, rimanendo la possibilità di condurla su Spatial Analyst.

Errori Standard e Variabili Ulteriori

4. Ripetendo i passaggi su elencati, come variabile Z da interpolare, sono stati considerati i valori dell'errore standard di predizione per entrambi i modelli, con il fine di dare una visualizzazione del giudizio di credibilità delle stime. Tale procedimento potrebbe essere ovviamente ripetuto per variabili epidemiologiche, preventivamente georeferenziate, avendo anche in questo caso la possibilità di scegliere tra un numero maggiore di tecniche statistiche.

5. Considerazioni finali

E' necessario innanzitutto fare alcune considerazioni di ordine metodologico relative all'applicazione di tecniche e modelli statistici.

Consideriamo le stime prodotte da entrambi i metodi, i relativi errori di predizione, le differenze tra le stime prodotte con loess e quelle prodotte con kriging e le statistiche principali.

I grafici di figura 4 mostrano i risultati di stima ed errore di predizione ottenuti con il kriging ordinario e con diversi valori di *span* per il loess.

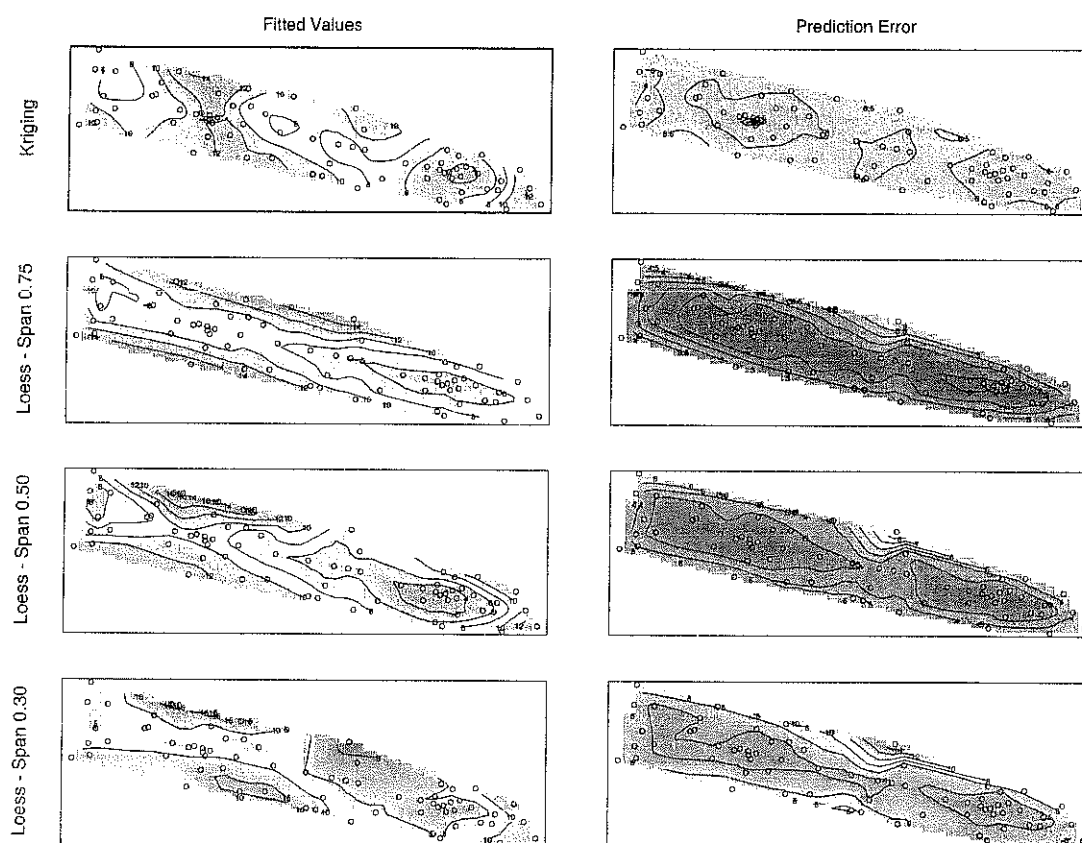


Figura 4. Nella colonna di sinistra, superfici stimate con loess e kriging. Nella colonna di destra, la stima corrispondente dell'errore di predizione. Grigio scuro per valori bassi, verde per valori alti.

Come ci si poteva aspettare, gli estremi del campo di variazione e la deviazione standard tendono a diminuire con l'aumentare dello *span* (tabella 2). Lo *smoothing* produce dunque una superficie sempre più piatta attorno ad un valor medio crescente

ed è naturale attendersi un aumento della distorsione indotta da un appiattimento eccessivo.

Alcune osservazioni sui valori numerici in senso stretto.

La prima riguarda il minimo per il loess 0.30 (tabella 2). Esso non ha alcun significato, potendo l'indice assumere solo valori positivi.

La seconda, il kriging (tabella 2). Ricordando che il kriging è un interpolatore esatto, non deve stupire il fatto che gli estremi del campo di variazione dei valori stimati non coincidono con quelli del dataset. Infatti, la predizione non ha riguardato necessariamente le coordinate dei punti osservati, ma un griglia rettangolare di punti equispaziati.

	Dataset	Loess 0.30	Loess 0.50	Loess 0.75	Kriging
<i>Min</i>	0.000	-1.431	2.408	3.665	2.984
<i>Mean</i>	8.725	7.813	8.375	8.927	9.037
<i>Max</i>	28.000	17.462	16.480	14.229	14.290
<i>Std Dev</i>	7.338	3.841	2.812	2.207	2.383

Tabella 2. Statistiche dei valori stimati.

	Loess 0.30	Loess 0.50	Loess 0.75	Kriging
<i>Min</i>	2.746	2.008	1.679	4.123
<i>Mean</i>	5.213	3.488	2.830	6.043
<i>Max</i>	15.270	8.755	6.302	6.727
<i>Std Dev</i>	1.884	1.097	0.852	0.219

Tabella 3. Statistiche degli errori di predizione.

In figura 5, i diagrammi a dispersione di stime kriging e loess, unitamente al valore della correlazione, confermano in modo chiaro quanto affermato precedentemente.

Il kriging produce delle stime più precise, ma condizionate all'esistenza del variogramma (sull'argomento si veda, Diblasi, Bowman, 2001) e alla sua bontà di adattamento.

Se noi poniamo ad un semplice confronto i due metodi, si noterà, all'aumentare dello *span*, un progressivo avvicinamento dei valori delle statistiche a quelli del kriging. Nella tabella 3, il confronto in termini di errore di predizione, ci dice che il loess, all'aumentare dello *span*, produce, in media, una variabilità decrescente (fino ad una riduzione di oltre il 50 % in media per uno *span* di 0.75) e minore di quella prodotta dal kriging, ma comunque soggetta ad un'incertezza maggiore.

Sul piano grafico, l'indubbio vantaggio di una sovrapposizione di temi efficientemente gestiti da un sistema territoriale comporta delle riflessioni in merito a quali informazioni rappresentare e come rappresentarle. Anche in questo senso, il supporto di una conoscenza statistica è resa necessaria, ad esempio, per valutare la qualità dei dati in ordine al disegno campionario progettato ed applicato, o per individuare quali fattori indagare sulla possibile esistenza di correlazioni tra variabili

ambientali, epidemiologiche e socio-demografiche. La scelta del numero di temi o *layer* da visualizzare contestualmente, può creare una “confusione visiva” che non è possibile apprezzare compiutamente, a meno di una riduzione delle informazioni a quelle necessarie o successive combinazioni semplici di temi che permettano all’occhio umano un lavoro non troppo difficoltoso.

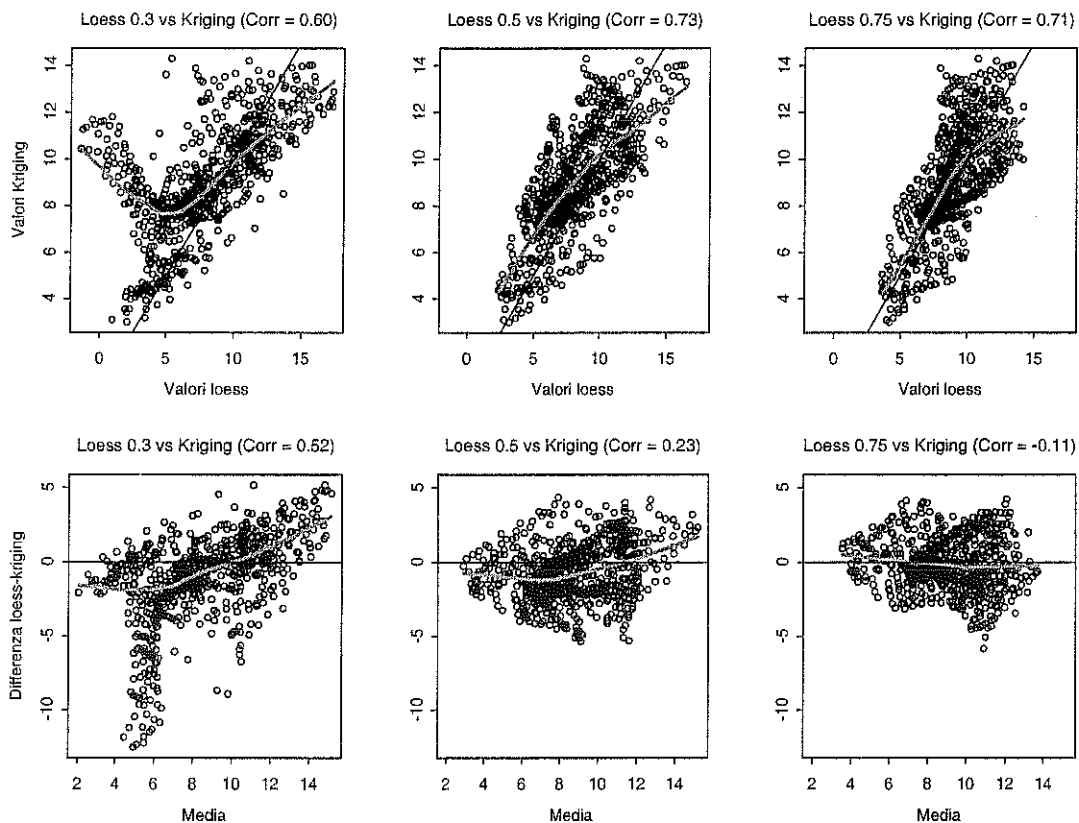


Figura 5. Scatter plot dei confronti loess-kriging.

Per dare compimento al nostro lavoro, in figura 6 è riportata la visualizzazione della superficie stimata con loess (*span* 0.75) in ambiente GIS, mentre in figura 7 la controparte del kriging. Sono stati considerati i principali punti di riferimento come la strada e gli edifici senza compromettere, a causa della risoluzione di stampa, la chiarezza visiva.

In particolare, in figura 7, si è realizzata l’integrazione tra la figura 1 e la figura 3. Considerate isolatamente, l’utilità della figura 3 è limitata alla capacità di identificare mentalmente le zone d’interesse o, nel caso la scala di entrambe lo permetta, di spostare lo sguardo da una figura all’altra. Naturalmente con la sovrapposizione tramite *layer* questa *impasse* è superata.

Da un punto di vista informatico, lo scambio grezzo dei dati attraverso un formato comune comporta ovviamente diversi passaggi ed una laboriosità maggiore.

L'esportazione da ArcView di *tables* in oggetti di S+ e viceversa per mezzo del modulo S+ for ArcView GIS, è più veloce ma limitata alla gestione di quel tipo di oggetti, appunto *data.frame*. Risulta, dunque, necessario un intervento manuale nel creare appositamente una serie di *records* a partire da un oggetto di tipo *list*.

I metodi interpolativi presenti su Spatial Analyst 1.1 sono: l'*Inverse Distance Weighted*, il Kriging, lo *Spline* e un interpolatore lineare globale. Questi metodi, teoricamente sufficienti per un'analisi parziale, sono fortemente condizionati dalla scelta di parametri non chiaramente definiti ed esplicitati per utenti meno esperti, e, per analisi statistiche più accurate, praticamente insufficienti.

Superficie dell'indice lichenico ottenuta con loess

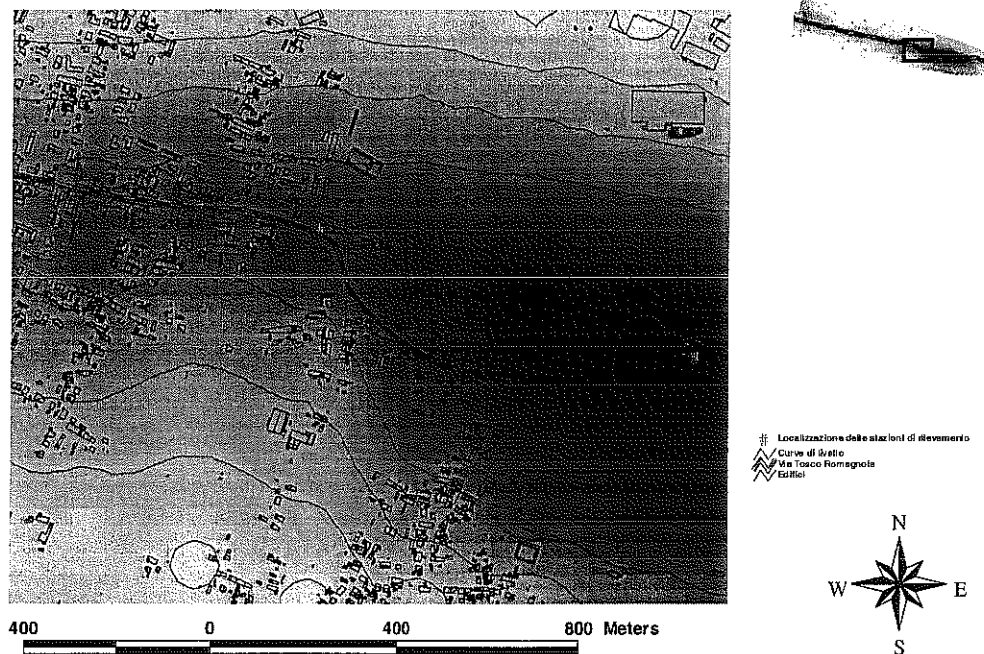


Figura 6. Visualizzazione su ArcView della superficie lichenica.

Superficie dell'indice lichenico ottenuta con kriging

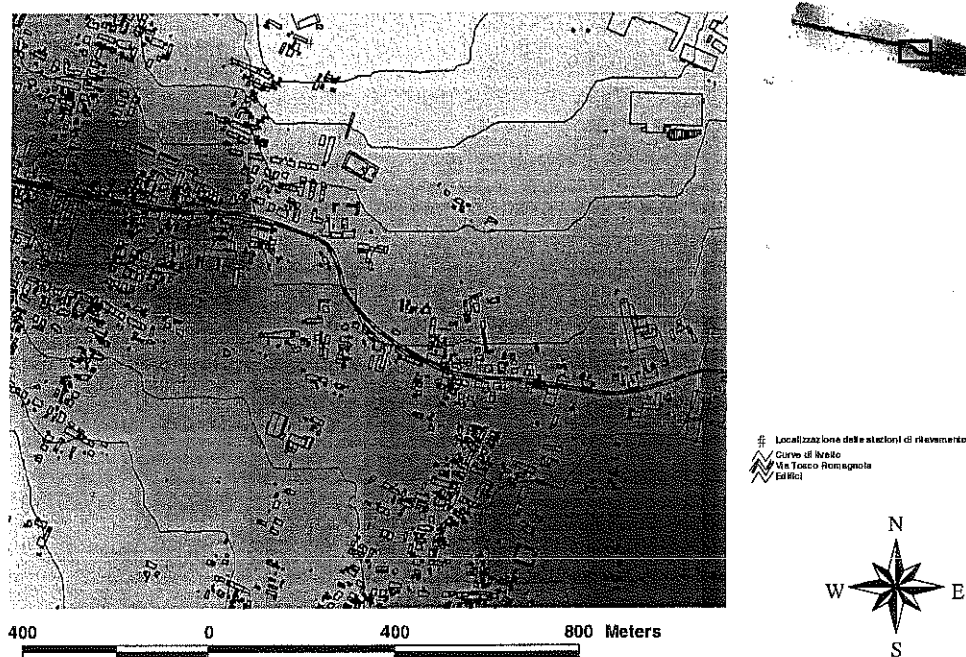


Figura 7. Visualizzazione su ArcView della superficie lichenica.

Innanzitutto, la possibilità di generare un grafico S-Plus, come la semplice interpolazione di un modello loess, direttamente su ArcView faciliterebbe notevolmente i passaggi. L'opzione del menù S-Plus per l'importazione di grafici è, infatti, riferita all'uso di grafici statistici come immagini su *layout* di ArcView, ma non come *layer* sovrapponibili cui è possibile applicare le funzioni di un GIS.

Un ulteriore aspetto, poi, di notevole importanza, attiene alla stima dell'errore di predizione. L'informazione relativa alla fiducia o meno sulla validità di una stima, è imprescindibile da qualsivoglia analisi di natura inferenziale. E' insufficiente, a nostro parere, la mera interpolazione della superficie ed il ricorso a pacchetti di tipo statistico produce risultati sicuramente più confortanti.

Appendice

Lo sviluppo di linguaggi di programmazione di alto livello e di sistemi operativi flessibili, ha favorito la creazione *ad hoc* di programmi di analisi spaziale talvolta specifici per alcune operazioni, coerenti con l'impostazione teorica del programmatore e limitati ovviamente alle risorse a lui disponibili. Abbiamo considerato soprattutto programmi realizzati sulla base di risorse private, in quanto sul mercato svolgono un ruolo di *benchmark*, non dimenticando, però, che ogni programma, gratuito o meno, dovrebbe essere comunque valutato in ordine alle *performance* che garantisce e alle risorse richieste. Non è scopo del presente lavoro realizzare un simile confronto, quanto quello di indicare alcuni prodotti significativi e le loro funzionalità principali - per un lavoro di questo genere, si veda Anselin, 2000.

A tal fine abbiamo considerato: S+SpatialStats, distribuito dalla Insightful, SpaceStat e Stat! dalla BioMedware Inc. e il Geographical Analysis Machine (GAM).

S+SpatialStats è un modulo estensivo delle funzionalità del pacchetto S+, finalizzato all'analisi sia delle mappe di punti sia dei dati geostatistici e *lattice*. Le potenzialità e la flessibilità sono quelle del linguaggio S.

SpaceStat è un software nato e sviluppato dall'originaria idea di Anselin per l'ottimizzazione non lineare di procedure di stima di massima verosimiglianza per modelli autoregressivi spaziali. L'attuale versione, la 1.90 (Anselin, 2000), è caratterizzata da un'ampia gamma di tecniche per l'analisi di dati *lattice*: costruzione di matrici spaziali, *Exploratory Spatial Data Analysis tools*, modelli regressivi particolarmente adatti per un approccio d'analisi econometrico. Dati geostatistici e *point pattern* non sono considerati.

Il GAM è un esempio di una completa automazione nella ricerca di *cluster* spaziali. La versione preliminare, variamente criticata per la mancanza di una base statistica, (Openshaw *et al.*, 1987; Openshaw, 1990) è stata progressivamente sviluppata e potenziata per raffinare l'unica funzione cui è deputato il programma: l'esplorazione automatica di concentrazioni di punti o piccole aree.

Stat!, realizzato da Jacquez (1996), è finalizzato, analogamente al GAM, alla ricerca tramite statistiche di vario genere di *cluster* spaziali e, diversamente dal GAM, temporali, e di interazioni spatio-temporali.

Riferimenti bibliografici

- ANSELIN, L. (2000). Computing environments for spatial data analysis. *Journal of Geographical Systems* **2**, 201-220.
- ANSELIN, L., E BAO, S. (1997). Exploratory spatial data analysis linking SpaceStat and ArcView. In M. Fischer and A. Getis (eds.), *Recent developments in spatial analysis*, Berlin, Springer-Verlag.
- ANSELIN, L., E FLORAX, R. J. G. M. (1995). *New Directions in Spatial Econometrics*, Berlin, Springer-Verlag.
- BAILEY, T. (1994). A review of spatial statistical analysis for geographical information systems. In *Spatial analysis and GIS* (ed. A. S. Fotheringham and P. Rogerson), 13-44, London, Taylor & Francis.
- BALLER, R. D., ANSELIN, L., MESSNER, S. F., DEANE, G., E HAWKINS, D. F. (2000). Structural covariates of U.S. county homicide rates: incorporating spatial effects. Forthcoming, *Criminology*, disponibile su <http://geog55.gis.uiuc.edu/~luc/papers.html>.
- BAO, S., ANSELIN, L., MARTIN, D., E STRALBERG, D. (2000). Seamless integration of spatial statistics and GIS: The S-PLUS for ArcView and the S+Grassland Links. *Journal of Geographical Systems* **2**, 287-306.
- BIVAND, R. S. (1999). Integrating GRASS 5.0 and R: GIS an modern statistics for data analysis. *Proceedings 7th Scandinavian Research Conference on Geographical Information Science*, Aalborg, Denmark, 111-127.
- BIVAND, R. S. (2000). Using the R statistical data analysis language on GRASS 5.0 GIS database files. *Computers and Geosciences* **26**, 1043-1052.
- BIVAND, R. S., E GEBHARDT, A. (2000). Implementing functions for spatial statistical analysis using the R language. *Journal of Geographical Systems* **2**, 307-317.
- BIVAND, R. S., E NETELER, M. (2000). Open source geocomputation: using the R data analysis language integrated with GRASS GIS and PostgreSQL database systems. *Proceedings 5th conference on GeoComputation*, 23-25 August 2000, University of Greenwich, U.K.
- BOOTS, B. (2000). Using GIS to promote spatial analysis. *Journal of Geographical Systems* **2**, 17-21.
- BOTTAI, M. (2001). Bootstrap Confidence Bands for Spatial Smoothing. *Atti del IV congresso della International Biometric Society*, regione Italiana, Rimini, 141-144.
- BRIGGS, D. J., COLLINS, S., ELLIOTT, P., FISCHER, P., KINGHAM, S., LEBRET, E., PRYL, K., VAN REEUWIJK, H, SMALLBONE, K, E VAN DER VEEN, A. (1997).

- Mapping urban air pollution using GIS: a regression-based approach. *International Journal of Geographical Information Systems* **11**, 699-718.
- BRIGGS, D. J., E ELLIOTT, P. (1995). The use of geographical information systems in studies on environment and health. *World Health Statistics Quarterly* **48**, 85-94.
- CHILDS, J. E., MCLAFFERTY, S. L., SADEK, R., MILLER, G. L., KHAN, A. S., DUPREE, E. R., ADVANI, R., MILLS, J. N., E GLASS, G. E. (1998). Epidemiology of rodent bites and prediction of rat infestation in New York City. *American Journal of Epidemiology* **148**, 78-87.
- CRESSIE, N. A. C. (1993). *Statistics for Spatial Data*, (Revised Edition), New York, Wiley.
- CRESSIE, N. A. C., E HAWKINS, D., M. (1980). Robust estimation of the variogram, I. *Journal of the International Association for Mathematical Geology* **12**, 115-125.
- CRONER, C. M., SPERLING, J., E BROOME, F. R. (1996). Geographic information systems (GIS): new perspectives in understanding human health and environmental relationships. *Statistics in Medicine* **15**, 1961-1977.
- DELLA MAGGIORE, R., BOTTAI, M., MAMMINI, U., MURA, E., E FRESCO, R. (2000). L'uso di tecnologie informatiche per l'analisi spaziale applicata al monitoraggio dell'inquinamento atmosferico, *Nota Tecnica CNUCE-B4-2000-025*, Pisa, CNUCE.
- DEPARTMENT OF THE ENVIRONMENT, 1987. Handling Geographic Information: Report to the Secretary of State for the Environment of the Committee of Enquiry into the Handling of Geographic Information, London, HMSO.
- DIBLASI, A., E BOWMAN, A. W. (2001). On the Use of the Variogram in Checking for Independence in Spatial Data. *Biometrics* **57**, 211-218.
- DIGGLE, P. J., GATRELL, A. C., E LOVETT, A. A. (1990). Modelling the Prevalence of Cancer of the Larynx in Part of Lancashire: A New Methodology for Spatial Epidemiology. In R. W. Thomas (ed.), *Spatial Epidemiology*, London, Pion Limited.
- DOLK, H., SHADDICK, G., WALLS, P., GRUNDY, C., THAKRAR, B., KLEINSCHMIDT, I., E ELLIOT, P. (1997a). Cancer Incidence near Radio and Television Transmitters in Great Britain. I. Sutton Coldfield Transmitter. *American Journal of Epidemiology* **145**, 1-9.
- DOLK, H., SHADDICK, G., WALLS, P., GRUNDY, C., THAKRAR, B., KLEINSCHMIDT, I., E ELLIOT, P. (1997b). Cancer Incidence near Radio and Television Transmitters in Great Britain. II. All High Power Transmitters. *American Journal of Epidemiology* **145**, 10-17.
- FOTHERINGHAM, A. S., E ROGERSON, P. A. (1993). GIS and spatial analytical problem. *International Journal of Geographical Information Systems* **7**, 3-19.

- GOODCHILD, M. F., HAINING, R., WISE, S., and 12 others (1992). Integrating GIS and spatial data analysis: problems and possibilities. *International Journal of Geographical Information Systems* **6**, 407-423.
- GRIFFITH, D. A. (1988). *Advanced Spatial Statistics*, Dordrecht, Kluwer Academic Publishers.
- HAINING, R., WISE, S., E MA, J. (2000). Designing and implementing software for spatial statistical analysis in a GIS environment. *Journal of Geographical Systems* **2**, 257-286.
- HAINING, R., WISE, S., E SIGNORETTA, P. (2000). Providing scientific visualization for spatial data analysis: Criteria and an assessment of SAGE. *Journal of Geographical Systems* **2**, 257-286.
- HASTIE, T., J, E TIBSHIRANI, R., J. (1990). *Generalized Additive Models*, New York, Chapman and Hall.
- HJALMARS, U., KULLDORFF, M., GUSTAFSSON, G., E NAGARWALLA, N. (1996). Childhood leukaemia in Sweden: using GIS and a spatial scan statistic for cluster detection. *Statistics in Medicine* **15**, 707-715.
- JACQUEZ, G. M. (1996). Statistical software for the clustering of health events. *Statistics in Medicine* **15**, 951-952.
- JACQUEZ, G. M. (2000). Spatial analysis in epidemiology: Nascent science or a failure of GIS?. *Journal of Geographical Systems* **2**, 91-97.
- KLEINSCHMIDT, I., BAGAYOKO, M., CLARKE, G. P., CRAIG, M., E LE SUEUR D. (2000). A spatial statistical approach to malaria mapping. *International Journal of Epidemiology* **29**, 355-361.
- KLEINSCHMIDT, I., SHARP, B. L., CLARKE, G. P., CURTIS, B., E FRASER, C. (2001). Use of generalized linear mixed models in the spatial analysis of small-area malaria incidence rates in Kwazulu Natal, South Africa. *American Journal of Epidemiology* **153**, 1213-1221.
- LAWSON, A. B., E WALLER, L. A. (1996). A review of point pattern methods for spatial modelling of events around sources of pollution. *Environmetrics* **7**, 471-487.
- MATHERON, G. (1962). *Traite de Geostatistique Appliquee, Tome I. Memoires du Bureau de Recherches Geologiques et Minières* **14**, Paris, Editions Technip.
- MUGGLIN, A. S., CARLIN, B. P., E GELFAND, A. E. (2000). Fully Model-Based Approaches for Spatially Misaligned Data. *Journal of the American Statistical Association* **95**, 877-887.
- MURRAY, A. T. (1999). Spatial analysis using clustering methods: Evaluating central point and median approaches. *Journal of Geographical Systems* **1**, 367-383.

- MURRAY, A. T., E ESTIVILL-CASTRO, V. (1998). Cluster discovery techniques for exploratory spatial data analysis. *International Journal of Geographical Information Systems* **12**, 431-443.
- OPENSHAW, S. (1990). Automating the Search for Cancer Clusters: A Review of Problems, Progress, and Opportunities. In R. W. Thomas (ed.), *Spatial Epidemiology*, London, Pion Limited.
- OPENSHAW, S., CHARLTON, M., WYMER, C., E CRAFT, A. W. (1987). A Mark I geographical analysis machine for the automated analysis of point data sets. *International Journal of Geographical Information Systems* **1**, 335-358.
- PETRUZZELLI, S., CELI, A., PULERA, N., BALIVA, F., VIEGLI, G., CARROZZI, L., CIACCHINI, G., BOTTAI, M., DI PEDE, F., PAOLETTI, P., E GIUNTINI C. (1998). Serum antibodies to benzo(a)pyrene diol epoxide-DNA adducts in the general population: effects of air pollution, tobacco smoking, and family history of lung diseases. *Cancer Research* **58**, 4122-4126.
- RATCLIFFE, J. H., E MCCULLAGH, M. J. (1999). Hotbeds of crime and the search for spatial accuracy. *Journal of Geographical Systems* **1**, 385-398.
- RIPLEY, B. D. (1991). *Spatial Statistics*, New York, Wiley.
- THOMAS, D. C. (2000). Some Contributions of Statistics to Enviromental Epidemiology. *Journal of the American Statistical Association* **95**, 315-319.
- THOMAS, R. W. (1990). *Spatial Epidemiology*, London, Pion Limited.
- WALL, P. A., E DEVINE, O. J. (2000). Interactive analysis of the spatial distribution of disease using a geographic information systems. *Journal of Geographical Systems* **2**, 243-256.
- ZHU, L., CARLIN, B., ENGLISH, P., E SCALF, R. (2000). Hierarchical modeling of spatio-temporally misaligned data: relating traffic density to pediatric asthma hospitalizations. *Environmetrics* **11**, 43-61.