

Costruzione di mappe di rischio da dati sanitari

Roberto DELLA MAGGIORE (*), Daniela NUVOLONE (*), Roberto FRESCO (*),
Elia PEROTTO (*)

(*) Istituto di Scienza e Tecnologie dell'Informazione "Alessandro Faedo" Area della Ricerca di Pisa, CNR
Via G.Moruzzi 1 56124 Pisa, Italia tel. 0503152944, e-mail: Roberto.dellaMaggiore@isti.cnr.it

Riassunto

Nell'ambito della linea di ricerca dell'ISTI-CNR "Metodologia GIS per il territorio e l'ambiente" vengono sottoposti ad analisi spaziale i dati ottenuti da indagini epidemiologiche. Nel presente studio i soggetti partecipanti ad una indagine sono stati considerati quali portatori di informazione ambientale al fine di ricavare una mappa indicativa della qualità dell'aria nel territorio da essi abitato. I dati utilizzati riguardano sintomi e malattie respiratori dei soggetti così come il loro stile di vita e le loro abitudini personali.

Abstract

Research conducted by ISTI-CNR about "GIS Technologies for Environmental Studies" focuses on the spatial analysis of data gathered in epidemiological surveys. In this work subjects participating to a survey performed in Pisa area are considered as environmental indicators: we aim to produce a map representing air quality of the area where subjects live. The method applied deals with data about respiratory symptoms/diseases, life style and personal habits of subjects.

Introduzione

I dati epidemiologici sono rappresentativi, per la loro stessa natura, di fenomeni legati al territorio e una loro analisi deve considerare la loro componente geografica intrinseca. È stato accertato dagli studiosi del settore che l'incidenza spaziale delle patologie è un punto chiave della loro origine (Schærström, 1996). Molti studi esaminano le associazioni tra *cluster* di incidenza di patologie con le possibili cause e viene generalmente assunto che l'esposizione di un individuo sia associata alle condizioni ambientali che caratterizzano il luogo attuale di residenza (Bentham, 1988). Vi sono tuttavia numerosi fattori che complicano l'esame dei dati raccolti attraverso le indagini epidemiologiche e che distorcono quella che potrebbe essere una relazione diretta tra causa ed effetto. Le difficoltà maggiori derivano dall'alta mobilità dei soggetti sul territorio, dalle diverse abitudini degli individui e dalle attività svolte. Per questi motivi è opportuno ai fini dell'analisi spaziale di dati epidemiologici effettuare una fase preliminare di elaborazione, che operi una sorta di filtro sui dati stessi, volto a rimuovere l'incertezza costituita dalle sorgenti di esposizione estranee al luogo di residenza dei soggetti.

Scopo

L'obiettivo principale dello studio presentato è quello di estrarre informazione a valenza ambientale dai dati epidemiologici: nello studio viene proposta una elaborazione dei dati relativi alla salute respiratoria degli individui partecipanti ad un'indagine epidemiologica. Ciò per ottenere un'indicazione della qualità dell'aria delle diverse zone su cui l'indagine è stata svolta. L'indicazione che si mira ad ottenere è di tipo qualitativo, concretizzata nella costruzione di una mappa del territorio nella quale sono evidenziate le aree più o meno salubri, cioè quelle in cui la popolazione è meno o più sofferente di malattie respiratorie (o ne manifesta i sintomi).

Lo studio è effettuato a livello di microaree ed è propedeutico per la valutazione delle diverse realtà locali presenti sul territorio.

Background

La sperimentazione descritta in questo articolo utilizza i dati dell'indagine epidemiologica che fu tenuta dal CNR nel corso degli anni 1991 – 1993, nel territorio della zona di Pisa e Cascina; l'indagine era rivolta ad analizzare gli effetti dell'inquinamento atmosferico sulla salute dell'uomo. L'area è caratterizzata da un inquinamento da traffico autoveicolare in particolar modo concentrato lungo la strada di comunicazione tra Pisa e Firenze (statale Tosco-Romagnola) e lo studio si proponeva di valutarne gli effetti provocati sulla salute respiratoria delle persone abitanti in prossimità della strada. Il campione di popolazione sottoposto ad indagine comprende anche un gruppo di persone abitanti nel quartiere periferico pisano di Porta a Mare, caratterizzato dalla presenza di un grande stabilimento industriale per la lavorazione del vetro.

Tutti i soggetti partecipanti risposero ad un questionario standardizzato, la cui compilazione assistita fu proposta alle famiglie che parteciparono all'indagine.

I dati raccolti sono stati oggetto di numerosi studi, svolti con gli strumenti della statistica classica, da parte dell'equipe di epidemiologi facente capo all'Unità di Epidemiologia Polmonare Ambientale dell'Istituto di Fisiologia clinica del CNR di Pisa (Viegi et al., 1994; Viegi et al., 1999). Più recentemente è stato fatto ricorso al GIS come supporto tecnologico di ausilio alla classificazione dei soggetti (della Maggiore et al., 2004; Peretti et al., 2005; Maio et al., 2005; Nitta et al., 2003; Garshick et al., 2003; Livingstone et al., 1996).

Nell'ambito della linea di ricerca dell'ISTI "Metodologia GIS per il territorio e l'ambiente" è stata proposta una metodologia per ricavare la mappa della salute respiratoria nelle aree sottoposte ad indagine (Nuvolone et al., 2005; della Maggiore et al., 2004). La metodologia si fonda su di un "indice di sintomatologia locale" sviluppato *ad hoc*. La georeferenziazione dei soggetti è realizzata secondo uno schema puntuale, localizzando ogni individuo all'indirizzo della propria abitazione (via e numero civico). In uno stesso punto possono essere localizzati più soggetti, dato che l'indagine è stata organizzata su base familiare ed anche perché vi sono casi in cui due o più famiglie partecipanti risiedono nello stesso edificio.

Materiali e metodi

Il campione generale di popolazione sottoposto ad indagine è costituito da individui di entrambi i sessi, di età variabile fra gli 8 e i 97 anni, estratti in maniera casuale fra i cittadini residenti in sezioni di censimento selezionate secondo criterio socio-economico tale da garantire una buona variabilità nella tipologia del campione. Il numero totale dei soggetti partecipanti è 2841, per un totale di 1131 punti distinti sul territorio.

Nel corso di precedenti attività di ricerca (Nuvolone et al., 2005; della Maggiore et al., 2004) è stato definito un "indice di sintomatologia locale" che pare sufficientemente adeguato allo scopo di rappresentare sul territorio lo stato della salute respiratoria della popolazione sottoposta ad indagine. L'indice è calcolato, per ogni punto, in funzione della quantità di sintomi o malattie riportati dai soggetti residenti in quel punto. La figura 1 mostra la mappa di rischio ottenuta con i dati dell'intera popolazione ed è una rappresentazione della qualità della salute respiratoria dell'intera popolazione sottoposta ad indagine (le zone più chiare corrispondono ad una salute migliore). Per i motivi esposti in premessa non si ha alcuna certezza che la mappa rappresenti anche la qualità dell'aria presente nella zona, dato l'elevato grado di aleatorietà dei sintomi e malattie riportati dai soggetti, aleatorietà che è legata a fattori personali. Di tali fattori è stata fatta pertanto una dettagliata analisi al fine di rimuovere il "rumore" insito nei dati e limitarne l'incertezza. L'operazione è stata resa possibile dalle numerose domande contenute nel questionario, a cui tutti partecipanti hanno risposto fornendo preziose informazioni relative alle abitudini personali ed allo stile di vita di ciascuno.

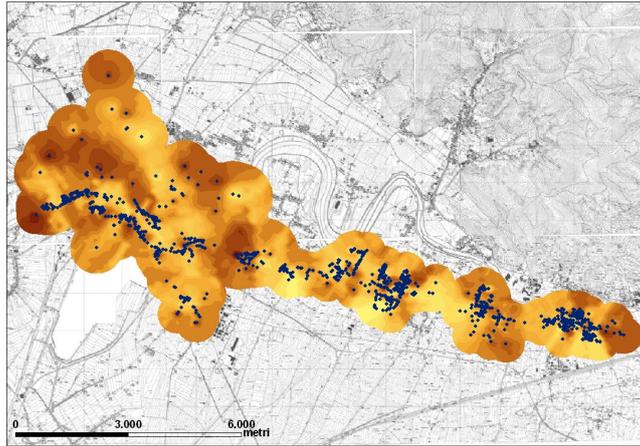


Figura 1 – Mappa di rischio per l'intero campione

Dall'analisi delle domande del questionario sono stati individuati casi in cui la salute respiratoria di una persona non è strettamente dipendente dal luogo in cui è fissata la sua residenza. Infatti alcune situazioni personali dei soggetti sono tali da sconsigliare decisamente di utilizzare i dati riportati dai medesimi come possibili indicatori della qualità dell'aria della zona di abitazione, in quanto alcuni fattori influenzano la salute respiratoria in maniera determinante, sovrastando l'eventuale azione ambientale locale. Per altri fattori si è ritenuto invece che essi possano influenzare la risposta individuale allo stimolo ambientale in modo non così deciso da dover scartare del tutto il dato. Pertanto si è provveduto a recuperare il contenuto informativo di tali dati trattandoli con un approccio basato sulla logica *fuzzy*, particolarmente indicata per la gestione dell'incertezza.

I principali fattori di confondimento individuati sono i seguenti:

- malattie pregresse ovvero patologie clinicamente accertate di tipologia e di rilevanza tale da costituire esse stesse il principale motivo di differenziazione da un normale profilo di funzionalità respiratoria;
- abitudine al fumo: si considera che il fumo sia un potenziale fattore di confondimento a causa della quantità e qualità delle sostanze inalate in modo diretto dai soggetti, indipendentemente dal luogo dove la persona risiede;
- fumo passato: studi epidemiologici stabiliscono che il soggetto che smette di fumare riacquista funzionalità polmonare simile ad un soggetto non fumatore solo dopo un determinato numero di anni;
- durata della residenza nell'indirizzo di indagine: i soggetti che risiedono da un periodo superiore a 5 anni sono considerati quelli che portano un contributo informativo più radicato al territorio, invece per le altre fasce (“fra 2 e 5 anni” e “meno di 2 anni”) si può avere incertezza;
- esposizione nell'ambiente di lavoro: la presenza di fumi, gas e polveri può avere influenza preponderante rispetto al luogo di residenza, più o meno mitigata dalla presenza di un efficiente impianto di ventilazione;
- esposizione ad agenti tossici nel luogo di lavoro;
- esposizione al fumo passivo, sia nel luogo di residenza che nell'eventuale posto di lavoro;
- esposizione a sostanze aeroinquinanti (quali il fumo dei gas di scarico degli autoveicoli) in luoghi diversi dall'abitazione.

Per l'operazione di filtraggio dei dati è stato sviluppato un supporto software a due stadi che fornisce come risultato il sottoinsieme di soggetti “geograficamente affidabili” da utilizzare per la costruzione della mappa di rischio. Per ogni stadio è stata lasciata la possibilità di scegliere le categorie di fattori di confondimento da trattare. Il primo stadio opera una scelta binaria, escludendo i soggetti portatori dei fattori prescelti. Nel secondo stadio invece è possibile tarare il filtro (mediante il programma che opera con logica *fuzzy*) secondo due criteri principali: la scelta della

cosiddetta “percentuale di affidabilità geografica” del soggetto e l’impostazione delle funzioni di appartenenza *fuzzy* che determinano l’esclusione del soggetto dall’insieme finale o la sua inclusione.

Usando il supporto software sono state effettuate più elaborazioni, variando le impostazioni anche in funzione della numerosità del sottoinsieme ottenuto (*trial and error*).

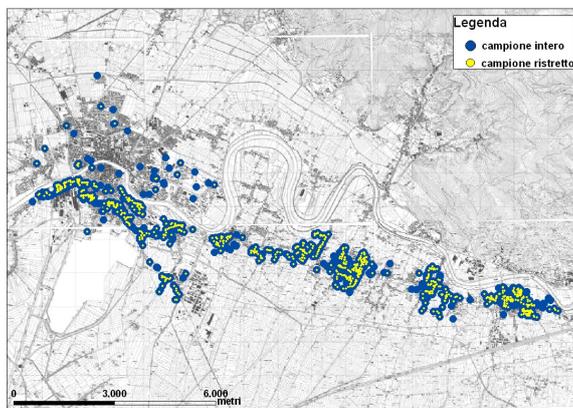


Figura 2 – Distribuzione spaziale dei campioni

Fra i sottoinsiemi selezionati con le numerose prove fatte ne abbiamo scelto uno che abbiamo ritenuto significativo per la discussione dei risultati. Il sottoinsieme è stato ottenuto con le seguenti impostazioni: col primo stadio sono stati esclusi i soggetti portatori di malattie pregresse, i fumatori abituali, gli ex fumatori che hanno smesso di fumare da meno di 7 anni e tutte le persone residenti all’indirizzo di indagine da meno di 5 anni. Nel secondo stadio (elaborazione *fuzzy*) sono stati modellati i soggetti esposti ad agenti tossici nel luogo di lavoro ed i soggetti esposti al fumo passivo; soprattutto in questa fase è stato utilizzato l’approccio *trial and error*. La soglia di affidabilità è stata fissata al 50 %. Sono stati così selezionati 1332 soggetti geograficamente affidabili, per un totale di 782 punti distinti.

La figura 2 mostra come la notevole riduzione del campione originale non modifichi significativamente la distribuzione spaziale dei punti di rilevamento. Quindi anche l’insieme ridotto garantisce una buona copertura dell’intero territorio di indagine.

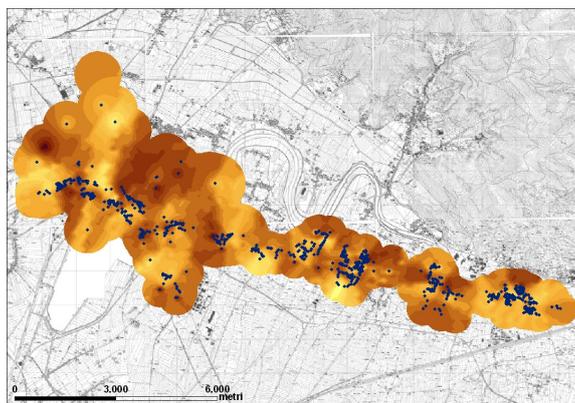


Figura 3 – Mappa di rischio del campione ridotto

Seguendo il procedimento sperimentato in precedenti lavori è stato ricavato l’indice di sintomatologia locale per il sottoinsieme e con questo si è ottenuta, per interpolazione superficiale, la mappa del territorio a valori continui presentata in figura 3.

Utilizzando questa mappa, con la definizione di opportuni valori di soglia dell'indice, il territorio è stato classificato in zone omogenee rispetto alla salute respiratoria della popolazione "geograficamente affidabile" (mappa a quattro valori discreti di figura 4).

Tale classificazione è stata validata ricorrendo alla statistica classica: i soggetti sono stati suddivisi in quattro classi a seconda dell'appartenenza della rispettiva abitazione alle suddette zone omogenee. Su tali classi sono state eseguite analisi bivariate e di regressione logistica. La presenza di numerose associazioni statisticamente significative fra i sintomi e le classi di soggetti indica la buona qualità della classificazione fatta.

Discussione dei Risultati

Basandoci sui risultati ottenuti col processo di validazione possiamo affermare che la mappa di figura 4 costituisce una buona rappresentazione del fenomeno "Salute Respiratoria" nel territorio di indagine ricavata dalle informazioni fornite dai soggetti ivi abitanti.

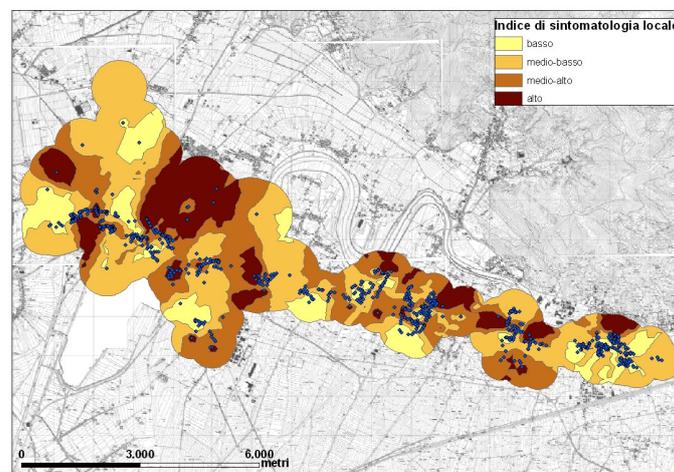


Figura 4 – Classificazione del territorio

Questa conclusione è praticamente la stessa a cui eravamo giunti con il precedente lavoro, basato sull'intero campione di popolazione invece che sul campione ristretto. Da un confronto delle mappe ottenute (fig.1 e fig.3) si osserva una sensibile differenza fra esse: però, mentre per la prima possiamo solo dire che essa è rappresentativa della salute respiratoria della popolazione residente nell'area, per la seconda possiamo azzardare anche un'affermazione più importante, e cioè che essa rappresenti la qualità dell'aria del territorio di indagine. Questo perché il campione ristretto è costituito da persone che hanno un legame più stretto col territorio di residenza e verosimilmente la loro eventuale patologia respiratoria è poco condizionata da fattori estranei ad esso. In definitiva il lavoro presentato costituisce un tentativo di utilizzare l'uomo come indicatore ambientale.

Possiamo fare alcune considerazioni prendendo a riferimento i numerosi studi di stampo prettamente epidemiologico (Viegi et al., 1994; Viegi et al., 1999) che sono stati effettuati con gli stessi dati utilizzati per questo lavoro. Da tali studi infatti sono emerse principalmente due indicazioni: vivere nei pressi della statale Tosco Romagnola non costituisce un rischio elevato per la salute respiratoria, pur se apprezzabile; il rischio risulta invece di misura maggiore per chi abita nel quartiere di Porta a Mare, nel quale è ubicato lo stabilimento di produzione del vetro. L'analisi qualitativa delle mappe prodotte conferma queste tendenze: la classificazione del territorio lungo la statale Tosco Romagnola (v. fig. 4) presenta zone a rischio più elevato alternate con zone a minor rischio, ma manca una indicazione di omogeneità o di stabilità lungo tutto il tratto di strada preso in esame; per il quartiere di Porta a Mare si ha invece un rischio elevato, ben evidenziato nella mappa di fig. 5 nella quale lo stabilimento di produzione del vetro risalta al centro dell'area a più alto rischio.

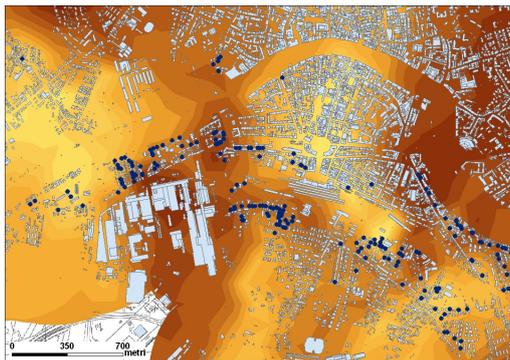


Figura 5 – Zona di Porta a Mare

Bibliografia

- Bentham G. (1988), "Migration and Morbidity: Implications for Geographical Studies of Disease" *Social Science & Medicine* 26: 49-54. Elsevier, ISSN: 0277-9536
- Della Maggiore R., Nuvolone D., Fresco R., Perotto E. (2004), "Metodi geostatistici per la produzione di mappe di rischio per la salute", *Atti della 8th Conferenza Nazionale ASITA: GEOMATICA Standardizzazione, interoperabilità e nuove tecnologie Roma, 14-17 Dicembre 2004, Volume II ISBN 88-900943-6-2 pp.: 997-1002*
- Della Maggiore R., Maio S., Nuvolone D., Fresco R., Baldacci S., Martini F., Borbotti M., Angino A., Viegi G. (2004), "Supporto GIS alla classificazione dei soggetti nelle indagini epidemiologiche", *Atti della 8th Conferenza Nazionale ASITA: GEOMATICA Standardizzazione, interoperabilità e nuove tecnologie Roma, 14-17 Dicembre 2004, Volume II ISBN 88-900943-6-2 pp.: 991-996*
- Garshick E., Laden F., Hart J.E., Caron A. (2003), "Residence near a major road and respiratory symptoms in U.S. veterans", *Epidemiology* 2003;14(6):728-736
- Livingstone A.E., Shaddick G., Grundy C., Elliott P. (1996), "Do people living near inner city main roads have more asthma needing treatment? Case-control study", *British Medical Journal* 1996; 312:676-677
- Maio S., Baldacci S., della Maggiore R., Nuvolone D., Martini F., Borbotti M., Angino A., Viegi G. (2005), "Applicazione della tecnologia GIS ai dati epidemiologici di un campione di popolazione generale residente a Pisa", *Atti del 29° Convegno annuale di epidemiologia AIE, Pisa 7-9 Settembre 2005, Volume unico pp.: 497-498*
- Nitta Y., Shima M., Adachi M. (2003), "Traffic-related air pollution and respiratory symptoms in children living along trunk roads in Chiba Prefecture", *Japan. J Epidemiol.* 2003;13(2):108-19
- Nuvolone D., Maio S., Fresco R., Baldacci S., Angino A., Martini F., Borbotti M., Viegi G., della Maggiore R. (2005), "Application of Geostatistical Methods for Public Health Risk Mapping", *Atti della 8th AGILE Conference on GIScience. Estoril, Portugal May 26-28 2005 Volume unico ISBN 972-893-13-6 pp.: 539-548*
- Peretti I., Baldacci S., della Maggiore R., Martini F., Nuvolone D., Stortini A., Viegi G. (2005), "Effetti dell'aerosol marino sulla salute umana: indagine epidemiologica su un campione georeferenziato della popolazione adulta di Livorno", *Atti del 29° Convegno annuale di epidemiologia AIE, Pisa 7-9 Settembre 2005, Volume unico pp.: 99-100*
- Schærström A. (1996), "Pathogenic Paths? A time geographical approach in medical geography" *Lund University Press, Lund, Sweden*
- Viegi G., Baldacci S., Vellutini M., Carrozzi L., Modena P., Pedreschi M., Maggiorelli F., Di Pede F., Paoletti P., Giuntini C. (1994), "Prevalence rates of diagnosis of asthma in general population samples of Northern and Central Italy", *Monaldi Arch Chest Dis*, pp. 191-196
- Viegi G., Pedreschi M., Baldacci S. (1999), "Prevalence rates of respiratory symptoms and diseases in general population samples of North and Central Italy", *Int J Tuberc Lung Dis*, pp 1034-1042