

## **L'incidenza dei fattori ambientali sul rischio di malattia: un modello anisotropico**

**Eugenia Nissi**<sup>1</sup>

**Annalina Sarra**<sup>1</sup>

*Summary:* In the recent years, there has been growing interest in the development of statistical methods useful in the detection of disease cases related to environmental pollution.

*Our starting point is an inhomogeneous Poisson point process model for the spatial variation in the incidence of cases and controls in a designated geographic region, as proposed by Diggle.*

*In this paper we focus on a particular model with directional effects to analyse the health risk around a point source.*

*In order to evaluate the adequacy of the proposed model we perform a Monte Carlo simulation study. In particular we analyse the maximum likelihood estimators properties of the model parameters.*

**Keywords:** processi spaziali di punto, regressione binaria, epidemiologia ambientale, modello con effetti direzionali.

### **1. Introduzione**

In letteratura nell'ultimo decennio, sono stati pubblicati diversi contributi che hanno mostrato come la teoria dei processi spaziali di punto sia in grado di fornire un'appropriata struttura di riferimento per problemi tipici dell'epidemiologia ambientale.

Come è noto, l'epidemiologia ambientale riguarda lo studio della variazione spaziale (e/o temporale) dell'incidenza delle malattie in relazione a possibili fattori ambientali. Gli eventi oggetto di studio, sono i casi di

---

<sup>1</sup> Dipartimento di Metodi Quantitativi e Teoria Economica – Viale Pindaro, 42 – 65127 Pescara (nissi@dmqte.unich.it – asarra@dmqte.unich.it)

malattia (o di morte) la cui distribuzione sul territorio può essere analizzata sia a fini semplicemente descrittivi, sia per fini eziologici.

Un'inevitabile caratteristica dei problemi di natura epidemiologica riguarda la marcata struttura spaziale della popolazione a rischio a causa di vincoli economici e fisici. Un utile punto di partenza è quello di condurre uno studio caso-controllo e quindi confrontare due strutture spaziali di punto: quella dei casi di malattia, ovvero delle localizzazioni di tutti i casi conosciuti della malattia in questione e quella di un campione casuale di individui (pattern dei controlli) che dovrebbe riflettere la distribuzione spaziale della popolazione a rischio sottostante. Nel presente lavoro, è illustrata la metodologia, basata sui processi spaziali di punto, utilizzata per stimare il rischio di malattia delle persone che vivono vicino ad una fonte di inquinamento ambientale. L'obiettivo è quello di proporre l'utilizzo di un modello flessibile per valutare l'incremento del rischio di malattia, quando questo dipende sia dalla distanza che dall'orientamento rispetto al punto sorgente.

Il lavoro è così articolato: nella sezione 2, dopo aver dato una definizione formale e statistica del problema, sono analizzati alcuni modelli che si basano su studi di tipo caso-controllo; in particolare è stato rivisitato il modello presentato da Diggle e Rowlingson, in cui la modellizzazione del rischio intorno ai cosiddetti punti sorgenti avviene tramite una regressione binaria, assumendo come misura approssimata dell'esposizione ambientale la distanza euclidea. Nella sezione 3, viene focalizzata l'attenzione su quelle situazioni in cui la particolare modalità di dispersione di certi inquinanti e la topografia variabile di determinate aree, impongono l'adozione di modelli più complessi rispetto a quelli basati sulla distanza euclidea. Viene proposto l'uso di un particolare modello che include effetti direzionali ("the directional plume model") per descrivere l'eccesso di rischio di malattia vicino ad una fonte di inquinamento ambientale nel caso in cui venga rimossa l'ipotesi di isotropia. Nella sezione 4 sono presentati i risultati di un esperimento simulativo, condotto per verificare la bontà della stima congiunta dei parametri che figurano nel modello con effetti direzionali. Infine nella sezione 5 sono riportate alcune considerazioni conclusive e possibili sviluppi metodologici.

## **2. Modello teorico di riferimento per la variazione spaziale del rischio**

Si ipotizzi di voler descrivere l'aumento del rischio di malattia in una regione geografica A, all'interno della quale, sia posizionata una fonte di inquinamento atmosferica.

Per le malattie non infettive, per le quali non si ha interazione tra gli eventi, si può assumere che la posizione dei casi di malattia segua un processo di Poisson non omogeneo con funzione di intensità  $\lambda(x)$  data da

$$\lambda(x) = \rho \lambda_0(x) f\{d(x); \theta\}. \quad (1)$$

Premesso che  $\rho$  rappresenti un fattore di scala, la funzione di intensità del processo è dunque il prodotto di due componenti: la prima  $\lambda_0(x)$ , che rappresenta la variabilità spaziale dovuta alla non omogenea distribuzione della popolazione a rischio; la seconda  $f\{d(x); \theta\}$ , che esprime la relazione tra il rischio di malattia e la sorgente di inquinamento ambientale.

Per  $f\{d(x); \theta\}$  si esplicita un opportuno modello parametrico che assume come variabile di esposizione il quadrato della distanza. Il termine esposizione si riferisce a tutte le variabili esplicative che si può ritenere abbiano una potenziale associazione con lo "stato di malattia". Osserviamo, infatti, che nella componente parametrica una misura dell'esposizione è ottenuta tramite la distanza euclidea, che fornisce un'adeguata rappresentazione della modalità di diffusione sul territorio dell'agente inquinante.

La scelta della distanza, come misura approssimata dell'esposizione, è motivata inoltre dalla carenza di dati ambientali. Infatti l'ottenimento di dati affidabili sull'esposizione è problematico ed in molti casi rappresenta il punto debole di molte applicazioni epidemiologiche. Molto raramente, infatti, è possibile disporre di stime dirette della concentrazione degli inquinanti nell'area di studio: il campionamento sull'esposizione individuale ed un monitoraggio biologico sono entrambi costosi.

Nella formalizzazione dell'intensità  $\lambda(x)$ , la funzione  $f(d(x); \theta)$  è esplicitata nel modo seguente:

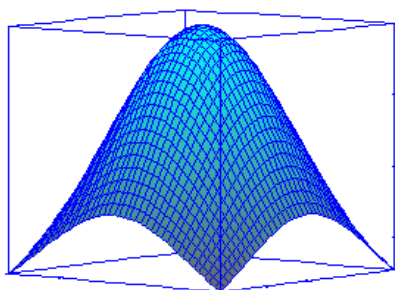
$$f(d(x); \theta) = 1 + \alpha \exp(-\beta d^2). \quad (2)$$

in cui  $\alpha$  rappresenta l'aumento del rischio nel punto in cui è posizionata la fonte di inquinamento, mentre  $\beta$  misura la scala spaziale degli effetti: al crescere del parametro  $\beta$  il rischio tende a concentrarsi in prossimità della sorgente di inquinamento ambientale, (Diggle, 1990).

Può essere utile rappresentare graficamente la funzione che descrive il cambiamento nell'intensità del processo al variare della distanza dalla fonte, per aver un'interpretazione geometrica dei parametri della stessa. (*Figura 1*).

Nel grafico sottoindicato, il parametro  $\alpha$  determina l'altezza del cono, ovvero la maggiore densità dei casi alla fonte: più alti valori di  $\alpha$  rappresentano una più alta concentrazione alla fonte. La funzione che modella il rischio, decresce all'aumentare della distanza dal punto sorgente fino a valere 1 quando l'effetto della fonte di inquinamento è nullo.

Il parametro  $\beta$  controlla quanto velocemente la funzione ritorna ad 1: più alti sono i valori di  $\beta$ , maggiore è la velocità con cui la funzione  $f(d(x); \theta)$  decresce.



**Figura 1** Funzione che descrive la relazione tra rischio di malattia e punto sorgente  $f(d; \theta) = 1 + \alpha \exp(-\beta d^2)$

Vale la pena sottolineare, come la parametrizzazione scelta, permetta di definire il rischio legato all'esposizione ambientale come eccesso di rischio rispetto ad un livello di base e rientri nella categoria dei modelli additivi. I modelli additivi conservano un comportamento appropriato a grandi distanze, quando l'effetto della sorgente è trascurabile, perché preservano il rischio di base. Se invece si adotta un modello di tipo moltiplicativo la funzione  $f(d; \theta)$  dovrà essere definita in modo da agire direttamente sul livello di base, ovvero sul rischio di base (Moolgavkar e Venzon, 1987). Tale parametrizzazione alternativa non garantisce a priori un maggiore rischio in prossimità della sorgente, mentre a grandi distanze può condurre ad un livello di rischio eccessivamente basso. Per questo motivo è consigliabile scegliere una parametrizzazione di tipo additivo, anche se talvolta la formulazione di un modello moltiplicativo fornisce un migliore adattamento (valore più elevato della funzione di verosimiglianza) in tutte le situazioni in cui oltre alle componenti radiali, nella funzione che descrive il rischio di malattia, sono presenti anche quelle direzionali.

## 2.1 Dal modello semiparametrico all'approccio basato sulla regressione binaria.

Con riferimento al modello teorico delineato nella sezione precedente, Diggle (1990), in un lavoro in cui analizza clusters di malattia in relazione ad una sorgente di inquinamento ambientale, suggerisce di utilizzare un campione di controlli per ottenere una stima kernel di  $\lambda_0(x)$ . L'analisi della funzione di intensità del processo permette di esaminare la distribuzione dei casi di malattia sul territorio e di saggiare l'ipotesi nulla di assenza dell'effetto del punto sorgente, cioè consente di verificare che la distribuzione dei casi di malattia sia proporzionale alla densità della popolazione:

$$H_0 : \boldsymbol{\theta} = (\alpha, \beta) = 0$$

Intuitivamente l'idea sottostante è la seguente: se non c'è relazione tra esposizione ed incidenza di malattia, allora la distribuzione dell'esposizione tra i casi dovrebbe essere la stessa della distribuzione dei controlli. La scelta di stimare in modo non parametrico  $\lambda_0(x)$  si riflette sulla stima di massima verosimiglianza dei parametri del modello  $(\alpha, \beta)$ .

La log-verosimiglianza profilo per i parametri  $(\alpha, \beta)$  è, a meno di una costante, pari a

$$l(\alpha, \beta) = \sum_{i=1}^m \log f(d(x); \boldsymbol{\theta}) - m \log \int_A \lambda_0(x) f(d(x); \boldsymbol{\theta}). \quad (3)$$

La massimizzazione dell'equazione precedente dipende dal modo in cui si è proceduto a specificare  $\lambda_0(x)$ . Diggle (1990) ha rilevato che la soluzione di massimo in forma chiusa può essere ottenuta utilizzando uno stimatore kernel gaussiano, invece, impiegando altre funzioni kernel, le stime di massima verosimiglianza possono essere ricavate da approssimazioni numeriche.

Tali considerazioni hanno suggerito la ricerca di modalità alternative di analisi per la valutazione dell'associazione tra rischio di malattia e sorgente di inquinamento. Per eliminare l'effetto attribuibile alla stima non parametrica di  $\lambda_0(x)$ , ossia della componente relativa alla distribuzione della popolazione, Diggle e Rowlingson (1994) hanno proposto un approccio che converte il precedente modello, basato sulla formalizzazione dell'intensità di un processo di Poisson non omogeneo, in una regressione binaria. Secondo tale approccio, le localizzazioni dei casi e dei controlli possono essere viste

come un insieme di etichette, il cui valore binario è determinato dalla probabilità dipendente dalla localizzazione.

Se si assume che casi e controlli siano le realizzazioni di due processi di Poisson, con intensità rispettivamente pari a  $\lambda_1(x)$  e  $\lambda_0(x)$ , e che la loro sovrapposizione sia ancora un processo di Poisson con intensità  $\lambda_0(x) + \lambda_1(x)$ , condizionandosi sulle posizioni dei casi e dei controlli, si ricava che le  $(n + m)$  etichette, dove  $n$  denotano i casi ed  $m$  i controlli, sono un insieme di v.c bernoulliane  $Y_i$ , tali che  $Y_i = 1$  se il punto è un caso di malattia e  $Y_i = 0$  altrimenti.

La probabilità che un individuo che risiede nel punto  $x_i$  sia affetto dalla malattia, è

$$P(Y_i = 1 | x_i) = p(x_i) = \frac{\lambda_1(x_i)}{\lambda_0(x_i) + \lambda_1(x_i)}. \quad (4)$$

Si procede stimando  $\frac{p(x)}{1-p(x)}$ , che esprime l'odds ratio, ovvero la probabilità che una persona che si trovi nella localizzazione  $x$  sia un caso piuttosto che un controllo; la stessa può essere anche espressa dal rapporto tra la funzione di intensità dei casi e quella dei controlli  $\left( \frac{\lambda_1(x)}{\lambda_0(x)} \right)$ .

Tornando alla parametrizzazione del rischio legato al punto sorgente si ricava che

$$p(x) = \frac{\rho f(d(x); \boldsymbol{\theta})}{1 + \rho f(d(x); \boldsymbol{\theta})} \quad (5)$$

e quindi si ha che

$$\log it(p(x)) = \log \left( \frac{\lambda_1(x)}{\lambda_0(x)} \right) = \log \rho + \log (f(d(x); \boldsymbol{\theta})) \quad (6)$$

I termini che figurano nel modello di regressione binaria a cui si è giunti misurano un eccesso di rischio relativo. Questo nuovo approccio evita di dover stimare  $\lambda_0(x)$  e quindi supera i problemi inferenziali sottolineati in precedenza. La stima dei parametri si ricava dalla funzione di verosimiglianza di una regressione binaria. Va sottolineata la maggiore flessibilità di questo nuovo approccio che contempla la possibilità di includere nel modello più fonti di inquinamento in forma moltiplicativa e quella di inserire eventuali covariate che permettano di tener conto dei

cosiddetti fattori di confondimento socio-economici (*Biggeri et al,1996*). Questi ultimi sono variabili esplicative da ritenersi non intrinsecamente spaziali, ma riferibili alle caratteristiche note dei singoli individui della popolazione, che incidono sul rischio di malattia: età, sesso, stato socio-economico. L'influenza di questi fattori non deve essere trascurata perché potrebbe essere talmente elevata da mascherare gli effetti sulla salute collegabili all'inquinamento ambientale e la loro non adeguata considerazione condurrebbe a conclusioni sbagliate.

### **3. Modello con effetti direzionali : “the directional plume model”.**

Il limite del modello visto nella sezione precedente resta senza dubbio il fatto che esso si basa su semplici ed indiretti indicatori di esposizione, quale ad esempio la distanza euclidea, tradizionalmente impiegata per analizzare gli effetti sulla salute di una fonte di inquinamento ambientale.

Una classe di modelli più generale, è quella introdotta da Lawson (1993) e da Lawson e Williams (1994), in cui il rischio dipende sia dalla distanza che dall'orientamento. Poiché fonti diverse di inquinamento possono dar luogo a differenti forme per l'associazione con il rischio di malattia, tali autori sottolineano, che per un'adeguata modellizzazione dell'esposizione, è necessaria un'attenta specificazione delle variabili o di trasformazioni delle stesse, tali da esplicitare il collegamento tra l'incidenza della malattia e la fonte di inquinamento ambientale. La funzione che descrive l'associazione tra rischio di malattia e punto sorgente potrà includere effetti basati solo sulla distanza, o sul logaritmo della stessa per descrivere picchi di intensità ad una certa lontananza dalla fonte, o effetti direzionali o effetti che tengono conto della correlazione distanza–direzione.

In generale, si introducono effetti basati solo sulla distanza quando la fonte di inquinamento è rappresentata da campi elettromagnetici, installazioni nucleari, inceneritori per rifiuti, ecc. che si suppone agiscano senza una preferenza direzionale. In altre situazioni, invece, la suddetta modellizzazione può risultare inadeguata: si pensi ad una fonte di inquinamento atmosferico, in cui la presenza di venti fa sì che la diffusione del presunto agente inquinante nell'area di studio sia anisotropica, cioè non radiale ma caratterizzata da direzioni prevalenti.

L'obiettivo del presente lavoro è proprio quello di analizzare il comportamento di un particolare modello che contempla effetti direzionali: “the directional plume model”. Il modello proposto ha l'apprezzabile peculiarità di fornire una parsimoniosa rappresentazione della modalità di diffusione dell'agente inquinante nella zona circostante la fonte di inquinamento ambientale, cioè quando è realistico ipotizzare che ci sia un

addensamento dei casi di malattia lungo una determinata direzione, individuata in base ai venti prevalenti o alla conformazione orografica del territorio.

Gli effetti angolari sono gestiti includendo le coordinate polari nella funzione che descrive il decadimento esponenziale: ogni punto sarà così individuato dalla distanza  $d_i$  e dall'angolo  $\mathcal{G}_i$  rispetto alla posizione della sorgente. L'effetto direzionale, che si manifesta in un pennacchio, da cui il termine "plume", è introdotto nel modello calcolando il coseno della differenza tra l'angolo di ciascun punto rispetto alla fonte di inquinamento e quello della direzione preferenziale, mentre l'addensamento lungo la direzione preferenziale  $\phi$ , ovvero la forma del pennacchio, è modulata tramite il parametro  $k$ . In particolare si avrà che, nel suddetto modello, la parametrizzazione del rischio di malattia legata al punto sorgente, è espressa dalla seguente funzione:

$$f(d, \mathcal{G}_i) = 1 + \alpha \exp\left(-[d \exp\{k \cos(\mathcal{G} - \phi)\} / \beta]^2\right) \quad (7)$$

Le variabili  $d$  e  $\mathcal{G}$  indicano le coordinate polari, mentre i parametri del modello contenuti nel vettore  $\theta$  sono:  $\alpha$  che esprime, anche per questo modello, l'aumento del rischio alla fonte;  $\beta$  che determina il tasso di decadimento del rischio con la distanza dalla fonte;  $\phi$  che indica la direzione del "pennacchio";  $k$  che descrive la concentrazione lungo la direzione preferenziale.

Anche per questa versione del modello che permette di tener conto dell'anisotropia, la stima dei parametri si ricava dalla funzione di verosimiglianza di una regressione binaria.

In particolare la funzione di log-verosimiglianza per  $\rho$  e per il vettore dei parametri  $\theta$  prende la seguente forma bernoulliana

$$L(d; \theta) = \sum_{i=1}^n \log p(x_i) + \sum_{i=n+1}^{n+m} \log\{1 - p(x_i)\} \quad (8)$$

Ricordando che  $p(x_i) = \Pr(\text{caso ad } x_i \mid \text{caso o controllo a } x_i)$ , per le considerazioni sull'etichettatura binaria si ha che tale probabilità risulta espressa dalla (5).

Quindi la funzione di log-verosimiglianza può essere riformulata come segue:

$$L(\rho, \theta) = n \log \rho + \sum_{i=1}^n \log f(d, \mathcal{G}_i) - \sum_{i=1}^{n+m} \log\{1 + \rho f(d, \mathcal{G}_i)\} \quad (9)$$



Un modello con effetti direzionali per l'analisi del rischio di malattia in prossimità di punti sorgenti

La funzione che esprime l'associazione tra rischio di malattia e punto sorgente, conserva, nel modello con effetti direzionali, una parametrizzazione tale da assumere valore 1 quando è nullo il vettore dei parametri  $\theta$ . Ne deriva che la log-verosimiglianza per  $\theta = 0$  avrà la seguente espressione:

$$L(\rho) = n \log \rho - (n + m) \log(1 + \rho) \quad (10)$$

la cui stima di massima verosimiglianza per  $\rho$  è  $\hat{\rho}_0 = \frac{n}{m}$ . Le stime dei parametri  $\hat{\rho}$  e  $\hat{\theta}$  sono ottenute massimizzando la funzione di verosimiglianza con metodi di ottimizzazione numerica.

#### 4. Esperimento con dati simulati per il modello con effetti direzionali.

Come emerso in precedenza, il modello direzionale si rivela uno strumento più flessibile per descrivere l'eccesso di rischio di malattia nelle vicinanze di un punto sorgente: esso risulta più adatto in tutte quelle situazioni in cui deve essere rimossa l'ipotesi di isotropia. La verifica della bontà delle stime di massima verosimiglianza dei parametri del suddetto modello e quindi della sua adeguatezza è stata effettuata attraverso un esperimento simulativo. Lo schema di simulazione ha richiesto innanzitutto la definizione della regione di studio  $A$  in cui generare le strutture spaziali di punto.

A tale proposito si assuma che  $A$  sia un cerchio di raggio unitario, con centro nell'origine, in cui si ipotizzi sia posizionata anche la ipotetica fonte di inquinamento ambientale. Le  $n$  localizzazioni dei casi di malattia sono state generate da un processo di Poisson non omogeneo, con funzione di intensità data da:

$$\lambda(x) = \rho \lambda_0(x) \left( 1 + \alpha \exp \left( - [d \exp \{k \cos(\vartheta - \phi)\} / \beta]^2 \right) \right) \quad (11)$$

La simulazione di un processo di Poisson non omogeneo è stata fatta ricorrendo all'algoritmo basato sul rifiuto campionario, introdotto da Lewis e Sheldler (1979), in cui attraverso la cancellazione di alcuni eventi, un processo di Poisson omogeneo è trasformato in un processo di Poisson con intensità variabile.

Più in dettaglio, l'algoritmo, per generare un pattern spaziale che rappresenti la realizzazione di un processo di Poisson con intensità variabile  $\lambda(s)$ , può essere schematizzato come segue:

- 1) nello step iniziale si genera un processo di Poisson omogeneo con intensità grande almeno quanto quella del processo non omogeneo: si simula cioè un processo di Poisson omogeneo su  $A$  con intensità  $\lambda \geq \max \{\lambda(s)\}$ , ottenendo una n-pla di coordinate  $(x_1, y_1), \dots, (x_n, y_n)$ ;
  - 2) successivamente si generano  $n$  v.c. uniformi (0,1) per ciascun evento in  $A$ , ottenendo  $U_1, \dots, U_n$ ;
  - 3) il terzo step è chiamato “thinning step” e consiste nell’eliminare tutti quegli eventi la cui intensità risulta superiore ad una certa soglia, in altri termini se  $U_i \leq \lambda(s)/\lambda$  si conserva l’evento, altrimenti lo si scarta.
- Nell’algoritmo di Lewis e Sheldler la proporzione tra l’intensità ad  $s$  e l’intensità  $\lambda$  è proprio la frequenza relativa appropriata.

Nell’esperimento simulativo condotto  $\lambda_0(x)$  è stato posto pari ad una costante e quindi le localizzazioni del campione dei controlli sono state ottenute come realizzazioni di un processo di Poisson omogeneo; questo con l’obiettivo di generare un campione di controlli che rappresenti bene la popolazione a rischio, distribuita casualmente sul territorio.

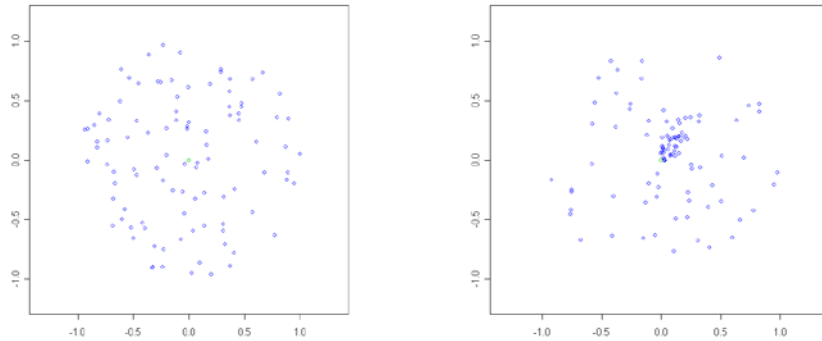
La performance del modello direzionale è stata valutata per diversi valori del parametro  $k$  ( $k= 1, 2, 3,4,5,6$ ); in particolare, in corrispondenza di ciascun valore del suddetto parametro, sono state eseguite 100 iterazioni, per ognuna delle quali il numero dei casi generati è stato posto pari a 100, mentre quello dei controlli è stato fissato a 900 ( $\rho$  di conseguenza, riflettendo la proporzione tra casi e controlli, è uguale a 0.11). I valori dei parametri del modello sono stati fissati rispettivamente pari a  $\alpha=80, \beta=0.005, \varphi=0.7854$ . I risultati, relativi all’accuratezza dei parametri, ottenuti per alcuni valori di  $k$  ritenuti significativi, sono sintetizzati nella tabella 4.1.

Alla scelta dei suddetti valori dei parametri siamo giunti dopo aver sperimentato diverse combinazioni dei valori degli stessi che, pur in presenza di un valore elevato di  $k$ , erano tali da non rendere visibile l’effetto direzionale del pattern simulato. Il parametro  $k$ , come già sottolineato in precedenza, influenza la forma del pennacchio.

In particolare per  $k=0$  il pattern non mostrerà la presenza di nessun effetto direzionale: si torna ad una situazione di isotropia, poiché si annulla nel modello il termine che tiene conto degli effetti angolari. (*Figura 2*)

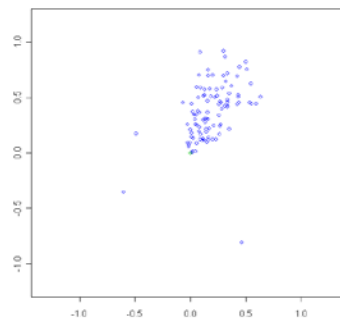
Al crescere di  $k$  la forma del pennacchio diventa sempre più allungata, mostrando nel pattern osservato la presenza di una direzione preferenziale lungo la quale si addensano i casi di malattia (*Figura 3 e Figura 4*). Da notare che la configurazione ottenuta per  $k=0$ , anche se non utilizzata per la simulazione, viene ugualmente riportata a fini di confronto.

Un modello con effetti direzionali per l'analisi del rischio di malattia in prossimità di punti sorgenti



**Figura 2:** casi di malattia generati per  $k=0$

**Figura 3** casi di malattia generati per  $k=3$



**Figura 4:** casi di malattia generati per  $k=5$

Dalla tabella sottoindicata emerge che lo stimatore di massima verosimiglianza fornisce risultati migliori (minore distorsione e maggiore efficienza, rilevata dai valori più bassi del MSE) in corrispondenza di un valore di  $k=5$ , mentre per valori minori di  $k$  (cioè per  $k=1$  e  $k=3$ ) i suddetti stimatori sono molto più distorti, evidenziando inoltre una più elevata deviazione standard.

È un risultato questo che non deve sorprendere e che trova una sua giustificazione nel fatto che il modello direzionale dovrebbe trovare applicazione in tutte quelle situazioni in cui l'addensamento dei casi di malattia lungo una direzione preferenziale è piuttosto evidente; laddove questo non accade non ha senso ricorrere ad un modello più complesso rispetto a quello gaussiano isotropico, che si lascia preferire per il ridotto numero di parametri da stimare.

**Tabella 1** Risultati dell'esperimento simulativo

<b>k=1</b>	<b><math>\alpha</math></b>	<b><math>\beta</math></b>	<b><math>\varphi</math></b>	<b>k</b>
Valori Parametri	80.0000	0.0050	0.7854	1.0000
Media	80.2290	0.0006	2.6893	2.0526
Deviaz. Standard	17.2756	0.0055	1.6854	2.3798
Distorsione	-0.2290	0.0044	-1.9039	-1.0526
E.Q.M	298.4995	0.0000	6.4652	6.7716
<b>k=5</b>	<b><math>\alpha</math></b>	<b><math>\beta</math></b>	<b><math>\varphi</math></b>	<b>k</b>
Valori Parametri	80.0000	0.0050	0.7854	5.0000
Media	75.4892	0.0051	0.7886	5.2738
Deviaz. Standard	4.8737	0.0041	0.0343	0.7666
Distorsione	4.5108	-0.0001	-0.0032	-0.2738
E.Q.M	44.1005	0.0000	0.0012	0.6626

<b>k=3</b>	<b><math>\alpha</math></b>	<b><math>\beta</math></b>	<b><math>\varphi</math></b>	<b>k</b>
Valori Parametri	80.0000	0.0050	0.7854	3.0000
Media	80.0397	0.0036	0.7393	3.4915
Deviaz. Standard	24.5441	0.0079	0.1914	1.4834
Distorsione	-0.0397	0.0014	0.0461	-0.4915
E.Q.M	602.4154	0.0001	0.0388	2.4421
<b>k=6</b>	<b><math>\alpha</math></b>	<b><math>\beta</math></b>	<b><math>\varphi</math></b>	<b>k</b>
Valori Parametri	80.0000	0.0050	0.7854	6.0000
Media	74.4678	0.0062	0.7749	5.9220
Deviaz. Standard	9.2181	0.0084	0.0907	1.2902
Distorsione	5.5322	-0.0012	0.0105	0.0780
E.Q.M	115.5791	0.0001	0.0083	1.6706

A tale proposito può essere interessante valutare cosa accade se, generati i punti dal modello direzionale (“directional plume model”), si adatta invece il modello isotropico.

In altri termini se nei punti osservati è presente un effetto direzionale e si adatta un modello che non contempla tali effetti, è naturale attendersi che lo stimatore di massima verosimiglianza utilizzato non potrà dare buoni risultati.

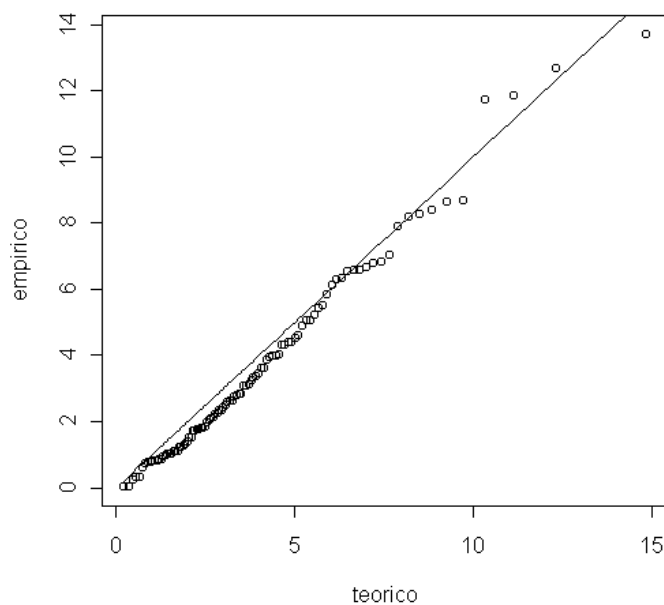
Le simulazioni effettuate a tale scopo confermano una forte distorsione degli stimatori del modello gaussiano isotropico, con un valore della stessa che aumenta al crescere dell'effetto direzionale; più in dettaglio è emerso che entrambi i parametri sono abbondantemente sovrastimati.

Tornando al modello con effetti direzionali la verifica di ipotesi sul vettore dei parametri è stata effettuata utilizzando il test del rapporto di verosimiglianza generalizzato.

L'aspetto cruciale dell'indagine prevede il confronto di 2 modelli spaziali per le localizzazioni dei casi: uno che non ha relazione con il punto sorgente e nel quale, perciò, il vettore dei parametri  $\theta$  è nullo ed uno in cui la struttura dei casi di malattia è influenzata dalla fonte. Formalmente si dirà che l'ipotesi nulla ( $H_0$ ) da sottoporre a verifica è che i casi di malattia ed i controlli hanno la stessa distribuzione spaziale sottostante; mentre nell'ipotesi alternativa  $H_1$  si assume che le localizzazioni dei casi abbiano una diversa struttura spaziale rispetto a quella dei controlli e che la densità delle localizzazioni dei casi sia più alta rispetto a quella dei controlli nelle vicinanze del punto sorgente.

L'ipotesi nulla di non associazione con il punto sorgente ( $\theta=0$ ) richiede un confronto tra la statistica test  $D = 2\{L(\hat{\rho}, \hat{\theta}) - L_0(\hat{\rho}_0)\}$  ed i valori teorici di una distribuzione  $\chi^2$  con un numero di gradi di libertà pari al numero dei parametri che figurano nel "*directional plume model*". Si è mostrato, attraverso un esperimento simulativo, che la statistica test del rapporto di verosimiglianza generalizzato, sotto l'ipotesi nulla di assenza di associazione con il punto sorgente, può essere ben approssimata da una distribuzione  $\chi^2$  con 4 gradi di libertà.

A conferma di ciò sono state eseguite 100 simulazioni in cui il vettore che contiene i parametri del modello  $\theta$  è stato posto uguale a zero e sono stati calcolati i corrispondenti 100 valori della statistica del rapporto di verosimiglianza generalizzato per testare ( $\vartheta_1 = \vartheta_2 = \vartheta_3 = \vartheta_4 = 0$ ). La figura che segue, mostra il Q-Q plot costruito confrontando la distribuzione empirica della statistica del rapporto di verosimiglianza generalizzato ed un  $\chi^2$  con 4 gradi di libertà, tanti quanti sono i parametri del modello. Si può affermare che l'approssimazione della distribuzione campionaria della statistica D sotto l'ipotesi nulla, di assenza di influenza del punto sorgente, alla distribuzione di un  $\chi^2$  con 4 gradi di libertà appare soddisfacente.



**Figura 5** Confronto tra la distribuzione empirica della statistica test del rapporto di verosimiglianza generalizzato e quella di un Chi-quadrato con 4 g.l..

## 5. Conclusioni

In questo lavoro è stato proposto un modello con effetti direzionali che risulta particolarmente flessibile per lo studio della variazione del rischio di malattia in prossimità di punti sorgenti quando deve essere abbandonata l'ipotesi di isotropia.

La performance dello stimatore di massima verosimiglianza, utilizzato per stimare i parametri del modello, è stata verificata attraverso un esperimento simulativo.

I buoni risultati ottenuti incoraggiano a sperimentare ulteriormente il modello; in particolare ci si propone sia di utilizzarlo per lo studio di dati di tipo caso-controllo osservati, sia dal punto di vista metodologico per valutare che si preservino le condizioni per l'applicabilità della teoria della massima verosimiglianza asintotica.

Un modello con effetti direzionali per l'analisi del rischio di malattia in prossimità di punti sorgenti

### **Ringraziamenti**

Si ringraziano i revisori anonimi, il prof. Mauro Coli e il Dott. Fabio Critani i cui preziosi suggerimenti hanno consentito sostanziali miglioramenti alla versione finale di questo lavoro.

Lavoro svolto nell'ambito del progetto di ricerca fondi ex 60% dal titolo "Modelli multivariati per lo studio della diffusione di inquinanti atmosferici." Responsabile Prof. Eugenia Nissi.

### **Riferimenti Bibliografici**

Biggeri, A., Barbone, F., Lagazio, C., Bovenzi, M. and Stanta, G. 1996. Air Pollution and lung cancer in Trieste, Italy: Spatial Analysis of Risk as a Function of Distance from sources, *Environmental Health Perspectives* **104**, pp.750-754.

Diggle, P.J. 1983. *Statistical analysis of Spatial Point Pattern*. Academic Press, New York.

Diggle, P.J. 1990. A point process Modelling Approach to raised Incidence of a Rare Phenomenon in the vicinity of a pre-specified point. *Journal of the Royal Statistical Society. Series A* **153**, pp.349-362.

Diggle, P.J.; Elliot, P., Morris, S.E. and Shaddick, G. 1997. Regression modelling of disease risk in relation to point sources. *Journal of the Royal Statistical Society. Series A* **160**, pp. 491-505.

Diggle, P.J.; Rowlingson, B.S. 1994. A conditional approach to point process modelling of elevated risk. *Journal of the Royal Statistical Society. Series A* **157**, pp.433-440.

Lawson, A.B. 1993. On the analysis of mortality events associated with a pre-specified fixed point. *Journal of the Royal Statistical Society. Series A* **156**, pp.363-377.

Lawson, A.B, Biggeri, A., Böhning, D., Lesaffre, E., Viel, J.F., Bertollini, R. 1999. *Disease mapping and risk assessment for public health*. John Wiley & Sons. LTD.

Lewis, P. A. .W and Shelder, G. S. 1979. Simulation of non-homogenous Poisson processes by thinning. *Naval Research Logistic Quarterly*, **26**, pp. 403-413.

Moolgavkar, H.S. and Venzon, J.D. 1987. General relative risk regression models for epidemiologic studies. *American Journal of Epidemiology*, **126**, pp. 946-961.