

# Giornata di Studio

## METODI STATISTICI PER L'ANALISI DI DATI SPAZIALI: CASI DI STUDIO E APPLICAZIONI

**giovedì 13 gennaio 2011**

*Aula Tesi, Dipartimento di Statistica,  
Università degli Studi di Milano-Bicocca  
Edificio U7, Via degli Arcimboldi, 8 Milano*

### Programma

- |                 |   |
|-----------------|---|
| Ore 13.30       | Inizio lavori   |
| Ore 13.30-13.45 | Saluto del Professor <b>Giovanni Corrao</b> , Preside della facoltà di Scienze Statistiche dell'Università degli studi di Milano-Bicocca  |
| Ore 13.45-14.30 | <i>L'uso dei modelli geostatistici multivariati nelle applicazioni ambientali</i> , Professor <b>Marco Minozzo</b> , Dipartimento di Scienze Economiche, Università degli Studi di Verona   |
| Ore 14.30-14.45 | <b>Discussione</b>  |
| Ore 14.45-15.15 | <i>Un approccio geostatistico per la previsione e la valutazione dei fattori geologici ed antropici sulle concentrazioni di radon indoor in Lombardia</i> Dott.ssa <b>Valeria Tritto</b> , Dipartimento di Statistica, Università degli Studi di Milano-Bicocca |
| Ore 15.15-15.25 | <b>Discussione</b>  |
| Ore 15.25-15.55 | <i>L'analisi spaziale nel controllo di processo industriale</i> Dott. <b>Luigi Radaelli</b> , Statistical Methods Eng. PC & ROBUSTNESS GROUP - MICRON ITALIA srl  |
| Ore 15.55-16.05 | <b>Discussione</b>  |
| Ore 16.05-16.15 | <b>Chiusura dei lavori</b>  |

La partecipazione al workshop e' gratuita.  
Per motivi organizzativi è gradita l'iscrizione entro il 10 gennaio 2011

Per informazioni: Riccardo Borgoni Tel. 0264485845 e-mail: [riccardo.borgoni@unimib.it](mailto:riccardo.borgoni@unimib.it)

## RIASSUNTI

### **L'uso dei modelli geostatistici multivariati nelle applicazioni ambientali**

Marco Minozzo

Dipartimento di Scienze Economiche Università' degli Studi di Verona

In ambito geostatistico l'analisi di dati spaziali multivariati viene tradizionalmente condotta facendo ricorso al modello lineare di coregionalizzazione (Wackernagel, 1995; Chiles e Delfiner, 1999) e alla "factorial kriging analysis" (Matheron, 1982). Sebbene queste analisi, e le previsioni che ne conseguono basate sul cokriging fattoriale, tralascino ogni considerazione sulla legge distributiva dei dati e si basino solamente sulla dipendenza spaziale misurata da variogrammi e cross-variogrammi, esse sono perfettamente adeguate, nel senso di fornire previsioni non distorte con errore quadratico medio minimo, nel caso in cui i dati possono assumersi avere distribuzione (multivariata) gaussiana. D'altro canto, nel caso di dati di conteggio, di dati asimmetrici, o comunque di dati non gaussiani, come in molte applicazioni ambientali, queste analisi possono dar luogo a previsioni dei fattori latenti altamente inefficienti. Sull'orma delle più interessanti proposte apparse in letteratura per la modellizzazione di dati spaziali univariati non gaussiani (Diggle, Moyeed e Tawn, 1998), qui si propone l'utilizzo, per dati spaziali multivariati, di un modello fattoriale gerarchico costruito su di una variante del classico modello lineare di coregionalizzazione. Stime di verosimiglianza dei parametri di questo modello possono essere ottenute utilizzando algoritmi EM stocastici (Fort e Moulines, 2003) e verosimiglianze simulate (Geyer, 1994; Christensen, 2004), mentre mappe dei fattori spaziali latenti (non osservabili) possono essere ottenute utilizzando algoritmi "Markov chain Monte Carlo". L'utilizzo di questi modelli geostatistici multivariati verrà illustrato nel caso di alcune basi di dati di origine ecologica e ambientale.

#### **Bibliografia**

Chiles, J.P. and Delfiner, P. (1999). *Geostatistics: Modeling Spatial Uncertainty*, Wiley, New York.

Christensen, O.F. (2004). Monte Carlo maximum likelihood in model-based geostatistics. *Journal of Computational and Graphical Statistics*, 13, 702-718.

Diggle, P.J., Moyeed, R.A. and Tawn, J.A. (1998). Model-based geostatistics. *Applied Statistics*, 47, 299-350.

Fort, G. and Moulines, E. (2003). Convergence of the Monte Carlo expectation maximization for curved exponential families. *The Annals of Statistics*, 31, 1220-1259.

Geyer, C.J. (1994). On the convergence of Monte Carlo maximum likelihood calculations. *Journal of the Royal Statistical Society, Series B*, 56, 261-274.

Matheron, G. (1982). *Pour une analyse krigeante des données régionalisées*. Publication N-732, Centre de Geostatistique, Fontainebleau, France.

Wackernagel, H. (1995). *Multivariate Geostatistics*. Springer, Berlin.

### **Un approccio geostatistico per la previsione e la valutazione dei fattori geologici ed antropici sulle concentrazioni di radon indoor in Lombardia**

Valeria Tritto

Dipartimento di Statistica Università degli Studi di Milano-Bicocca

La problematica del radon indoor è da anni ampiamente studiata e discussa a livello mondiale, essendo proprio il radon la principale causa di tumore al polmone dopo il fumo di tabacco. La Regione Lombardia ha svolto fra il 2003 e il 2004 una campagna di misura della concentrazione di radon indoor nel suo territorio, allo scopo di individuare le aree a maggior probabilità di alte concentrazioni di radon. In particolare, sono state effettuate circa 3600 misure di concentrazione media annua di radon in aria indoor in unità immobiliari site al piano terra. Statistiche preliminari di sintesi hanno mostrato un netto trend sud-nord, indicando maggiori concentrazioni nella parte settentrionale della Regione. Questo aspetto è dovuto principalmente alla morfologia della Lombardia caratterizzata da territori montuosi a nord e dalla Pianura Padana a sud. La litologia del territorio è un aspetto di grande importanza essendo proprio il suolo la principale sorgente di radon. Sulla base di ciò, sono state prodotte delle mappe che permettono di rappresentare in modo continuo l'andamento della concentrazione di radon indoor sul territorio. Nello specifico sono state applicate tecniche di previsione kriging (kriging con trend esterno, considerando come covariata esterna proprio la litologia; Goovaerts, 1997) e di simulazioni (Chiles, 1999). La componente territoriale, legata agli aspetti fisici e geologici, non riesce a spiegare del tutto la variabilità del fenomeno radon indoor, essendo parte di essa legata a fattori antropici specifici degli edifici, come i materiali di costruzione, la qualità degli infissi o il tipo di locale. Per valutare congiuntamente il potenziale effetto di tutte queste componenti sulla concentrazione di radon indoor è stato implementato un modello di regressione multivariato inserendo variabili esplicative della componente territoriale come la litologia e l'altitudine e variabili di natura antropica come il materiale di costruzione delle pareti, la tipologia di edificio e

l'attacco a terra. I parametri del modello sono stati stimati tramite GLS e impiegando un algoritmo proposto in Schamberger O., Gotway C.A. (2005).

#### **Bibliografia**

ARPA Lombardia, Indagine per l'individuazione delle radon prone areas in Lombardia, Rapporto a cura della U.O. Agenti Fisici ed Energia Settore Aria e Agenti Fisici ARPA della Lombardia, Milano  
Chiles JP, Delfiner P (1999), Geostatistics: modeling spatial uncertainty, John Wiley & Sons, New York  
Goovaerts P. (1997), Geostatistics for Natural Resources Evaluation, Oxford University Press, New York  
Schamberger O, Gotway CA (2005), *Statistical methods for spatial data analysis*, Chapman & Hall/CRC  
IReR – Istituto Regionale di ricerca della Lombardia (2010), *Politiche di sicurezza per lo sviluppo sostenibile del territorio: rischio reale e percepito. Allargamento ai rischi emergenti in Regione Lombardia –Dossier Radon*, Milano

#### **L'analisi spaziale nel controllo di processo industriale**

Luigi Radaelli

Statistical Methods Eng. PC & ROBUSTNESS GROUP – MICRON ITALIA srl

L'analisi spaziale dei dati è una metodologia che permette di studiare il comportamento e la correlazione di variabili relativamente alla loro dislocazione su un'area di interesse definita. Tale metodologia trova già un'ampia diffusione nello studio dell'ambiente e nel settore del marketing. In ambito industriale, ed in particolare nel settore della microelettronica, poche sono le applicazioni sviluppate di statistica spaziale. In questa presentazione verranno descritti due lavori che testimoniano come tale approccio può essere d'aiuto nella fase di controllo di processo. Partendo da due problemi diversi, uno l'identificazione di potenziali cluster di particelle di polvere e l'altro l'identificazione di una mappa ridotta per il monitoraggio di un parametro di processo, si mostra come con l'utilizzo di tecniche come il kriging, il simulated annealing e la modellazione di un processo di Poisson sia possibile proporre delle valide soluzioni ai quesiti di partenza. Da ultimo, ma non per questo meno importante, verrà mostrato l'aspetto della ricaduta, in termini di benefici, di questi studi in seguito alla loro implementazione sulla linea produttiva.

#### **Bibliografia**

Borgoni R, Radaelli L, Tritto V, Zappa D. 2010. Optimal reduction of a monitoring grid for SiO<sub>2</sub> deposition surface over a wafer for semiconductor devices. Atti della XLV Riunione Scientifica della Società Italiana di Statistica  
Borgoni R, Radaelli L, Zappa D. 2010. Spatial defect pattern recognition on semiconductor wafers. Future Fab International 33 [www.future-fab.com](http://www.future-fab.com)  
Diggle PJ. 2003, Statistical Analysis of Spatial Point Patterns. 2-nd edition, Arnold London  
Illian J, Penttinen A, Stoyan D. 2008. Statistical Analysis and Modeling of Spatial Point Patterns. Wiley New York  
van Groenigen, JW, Siderius, W, Stein, A. 1999. Constrained optimisation of soil sampling for minimisation of the kriging variance Geoderma, 87, 239-259