Società Italiana di Statistica

Atti della XLIII Riunione Scientifica

Proceedings of the XLIII Scientific Meeting

UNIVERSITÀ DI TORINO

14 - 16 Giugno 2006

Sessioni spontanee

Contributed sessions

2006

STATISTICA

cleup

2006

Società Italiana di Statistica

Atti della XLIII Riunione Scientifica

Proceedings of the XLIII Scientific Meeting

UNIVERSITÀ DI TORINO

14-16 giugno 2006

14-16 June, 2006

SESSIONI SPONTANEE

Contributed sessions

Con il patrocinio dell'Università degli Studi di Torino, del Politecnico di Torino e del CNR (Consiglio Nazionale delle Ricerche)

Supported by Università di Torino, Politecnico di Torino and CNR (Consiglio Nazionale delle Ricerche)

cleup

Si ringrazia per il contributo all'organizzazione della 43ª Riunione Scientifica:
Thanks for support are owed to:

Fondazione CRT
Compagnia di San Paolo
Camera di Commercio, Industria Artigianato e Agricoltura di Torino
Istat - Istituto Nazionale di Statistica
Università degli studi di Torino
Facoltà di Economia
Facoltà di Medicina e Chirurgia
Facoltà di Scienze Politiche
Consiglio di Corso di Laurea in Scienze Statistiche
Dipartimento di Statistica e Matematica applicata "Diego de Castro"
Fondazione Franca e Diego de Castro
Unione Industriale di Torino

Tutti i lavori di questo volume sono stati sottoposti a revisione scientifica da componenti del Comitato Programma

All papers in this volume have been refereed by members of the Program Committee

Prima edizione: giugno 2006

ISBN 88-7178-791-9

© Copyright 2006 by CLEUP sc
"Coop. Libraria Editrice Università di Padova"
Via G. Belzoni, 118/3 – Padova (Tel. 049 650261)
www.cleup.it

Tutti i diritti di traduzione, riproduzione e adattamento, totale o parziale, con qualsiasi mezzo (comprese le copie fotostatiche e i microfilm) sono riservati.

SOCIETÀ ITALIANA DI STATISTICA

Sede: Salita de' Crescenzi 26 - 00186 Roma
Tel +39-06-6869845 - Fax +39-06-68806742
sis@caspur.it Web: http://w3.uniroma1.it/sis/

La Società Italiana di Statistica (SIS), fondata nel 1939, è una società scientifica eretta a Ente morale ed inclusa fra gli Enti di particolare rilevanza scientifica. La SIS promuove lo sviluppo delle scienze statistiche e la loro applicazione in campo economico, sociale, sanitario, demografico, tecnologico, produttivo e in molti altri settori di ricerca.

Organi della società

Presidente: Daniela Cocchi (Università di Bologna)

Segretario Generale: Alessandra De Rose (Università di Roma "La Sapienza")

Tesoriere: Maria Felice Arezzo (Università di Roma "La Sapienza")

Consiglieri: Enrica Aureli (Università di Roma "La Sapienza"); Giuseppe Bove (Università di Roma "La Sapienza"); Marcello Chiodi (Università di Palermo); Piero Demetrio Falorsi (ISTAT); Antonio Giusti (Università di Firenze); Enrico Gori (Università di Udine); Silio Rigatti Luchini (Università di Padova); Giorgio Tassinari (Università di Bologna)

XLIII RIUNIONE SCIENTIFICA

COMITATO PROGRAMMA

Presidente: Silio Rigatti Luchini (Università di Padova)

Membri: Simona Balbi (Università di Napoli); Francesco Battaglia (Università di Roma "La Sapienza"); Salvatore Bologna (Università di Palermo); Giampiero Dalla Zuanna (Università di Padova); Ignazio Drudi (Università di Bologna); Stefano Falorsi (ISTAT); Giovanni Maria Giorgi (Università di Roma "La Sapienza"); Salvatore Ingrassia (Università di Catania); Anna Clara Monti (Università del Sannio); Mario Montinaro (Università di Torino); Tommaso Proietti (Università di Roma Tor Vergata); Federico Mattia Stefanini (Università di Firenze); Susanna Zaccarin (Università di Trieste); Alessandra De Rose (Segretario Generale SIS)

COMITATO ORGANIZZATORE LOCALE

Presidente: Mario Montinaro (Università di Torino)

Membri: Cinzia Carota (Università di Torino); Paolo Chirico (Università di Torino); Dalit Contini (Università di Torino); Roberto Corradetti (Università di Torino); Mauro Gasparini (Politecnico di Torino); Dario Gregori (Università di Torino); Anna Lo Presti (Università di Torino); Carlo Maccheroni (Università di Torino); Massimo Marinacci (Università di Torino); Andrea Scagni (Università di Torino); Grazia Vicario (Politecnico di Torino); Alessandra De Rose (Segretario Generale SIS); Maria Felice Arezzo (Tesoriere SIS)

SIS2006 – SEGRETERIA DEL COMITATO ORGANIZZATORE LOCALE

Sede: Dipartimento di Statistica e Matematica applicata "D. De Castro" –
Sezione: Scipol - Via Maria Vittoria n. 38, 10123 Torino
Tel: 011. 6703130/3080 – *Fax:* 011.6703139
E-mail: sis2006@unito.it *Sito web:* www.sis2006.unito.it

Pubblicazioni della Società Italiana di Statistica
SIS Informazioni (notiziario mensile)
Statistical Methods & Applications (quadrimestrale)
Atti delle Riunioni Scientifiche (biennale)
Atti dei Convegni Intermedi (biennale)

ITALIAN STATISTICAL SOCIETY
Head office: Salita de' Crescenzi 26 - 00186 Roma
Tel +39-06-6869845 - Fax +39-06-68806742
E-mail: sis@caspur.it Web: http://w3.uniroma1.it/sis/

The Italian Statistical Society, founded in 1939, is a no profit scientific society with the aim of promoting the development of the Statistical Sciences and their applications in economic, social, medical, demographic, technological and productive fields and in many other working areas.

Executive and Council

Chairman: Daniela Cocchi (Università di Bologna)
Secretary: Alessandra De Rose (Università di Roma "La Sapienza")
Treasurer: Maria Felice Arezzo (Università di Roma "La Sapienza")
Council members: Enrica Aureli (Università di Roma "La Sapienza"); Giuseppe Bove (Università di Roma "La Sapienza"); Marcello Chiodi (Università di Palermo); Piero Demetrio Falorsi (ISTAT); Antonio Giusti (Università di Firenze); Enrico Gori (Università di Udine); Silio Rigatti Luchini (Università di Padova); Giorgio Tassinari (Università di Bologna)

XLII SCIENTIFIC MEETING OF THE ITALIAN STATISTICAL SOCIETY

PROGRAMME COMMITTEE

Chairman: Silio Rigatti Luchini (Università di Padova)

Members: Simona Balbi (Università di Napoli); Francesco Battaglia (Università di Roma "La Sapienza"); Salvatore Bologna (Università di Palermo); Giampiero Dalla Zuanna (Università di Padova); Ignazio Drudi (Università di Bologna); Stefano Falorsi (ISTAT); Giovanni Maria Giorgi (Università di Roma "La Sapienza"); Salvatore Ingrassia (Università di Catania); Anna Clara Monti (Università del Sannio); Mario Montinaro (Università di Torino); Tommaso Proietti (Università di Roma Tor Vergata); Federico Mattia Stefanini (Università di Firenze); Susanna Zaccarin (Università di Trieste); Alessandra De Rose (Segretario Generale SIS)

LOCAL ORGANISING COMMITTEE

Chairman: Mario Montinaro (Università di Torino)

Members: Cinzia Carota (Università di Torino); Paolo Chirico (Università di Torino); Dalit Contini (Università di Torino); Roberto Corradetti (Università di Torino); Mauro Gasparini (Politecnico di Torino); Dario Gregori (Università di Torino); Anna Lo Presti (Università di Torino); Carlo Maccheroni (Università di Torino); Massimo Marinacci (Università di Torino); Andrea Scagni (Università di Torino); Grazia Vicario (Politecnico di Torino); Alessandra De Rose (Segretario Generale SIS); Maria Felice Arezzo (Tesoriere SIS)

SIS 2006 LOCAL HEAD OFFICE

Sede: Dipartimento di Statistica e Matematica applicata "D. De Castro"
Sezione: Scipol - Via Maria Vittoria n. 38, 10123 Torino
Tel: + 39-011-6703130/3080 – *Fax:* +39-011-6703139
E-mail: sis2006@unito.it *Web page:* www.sis2006.unito.it

Publications of the Italian Statistical Society
SIS Informazioni (monthly)
Statistical Methods & Applications (quarterly)
Proceedings of the Scientific Meeting (biyearly)
Proceedings of the Intermediate Conference (biyearly)

INDEX

SESSIONE SPONTANEA 1 - CONTRIBUTED SESSION 1
Bayesian Methods
(Metodi bayesiani) 21
Chairman and Discussant: FULVIO DE SANTIS (Università di Roma "La Sapienza")

F. BERTOLINO (Università di Cagliari), W. RACUGNO (Università di Cagliari)
Bayesian parametric testing for independent distributions with nuisance parameters....23

S. BOZZA (Università "Cà Foscari" di Venezia), F. TARONI (University of Lausanne), M. MARQUIS (University of Lausanne), M. SCHMITTBUHL ("Louis Pasteur" University, Strasbourg)
Evaluation of evidence in forensic science: a multivariate bayesian approach.....27

S. CABRAS (Università di Cagliari), M. E. CASTELLANOS (Universidad "Rey Juan Carlos", Madrid)
Comparison of Skewness in two populations.....31

C. CONIGLIANI (Università di Roma Tre), A. TANCREDI (Università di Roma "La Sapienza")
A Bayesian semi-parametric approach for cost-effectiveness analysis in health economics 35

A. FARCOMENI (Università di Roma "La Sapienza"), L. TARDELLA (Università di Roma "La Sapienza")
Bayesian Methods for Semiparametric Recapture Models 39

B. LISEO (Università di Roma "La Sapienza"), J. ROUSSEAU (CEREMADE, Université Paris Dauphine)
Sequential importance sampling algorithm for Bayesian nonparametric long range inference 43

C. SCRICCIOLO (Università "Bocconi" di Milano)
On Bayesian Adaptive Regression Function Estimation.....47

Earthquakes clustering based on maximum likeliho estimation of point process conditional intensity funct

Raggruppamento di eventi sismici basato sulla stima di massima verosimiglianza della funzione di intensità di un processo di punto

Giada Adelfio[1], Marcello Chiodi[1], Luciana De Luca[2], Dario Luzio[2]

[1] Dipartimento di Scienze Statistiche e Matematiche "S. Vianelli", Palermo

[2] Dipartimento di Chimica e Fisica della Terra, Palermo

e-mail: adelfio@dssm.unipa.it; chiodi@unipa.it

Riassunto: In questo lavoro viene proposta una tecnica di clustering progettata per separare le due principali componenti dell'attività sismica di un'area sismogenetica: la sismicità di fondo e quella relativa a sequenze sismiche. Il metodo qui proposto assegna ciascun terremoto a una sequenza sismica o all'insieme degli eventi che costituiscono la sismicità di fondo, sulla base dei valori assunti dalla funzione di intensità del processo di punto generatore degli eventi osservati. I parametri che caratterizzano tale funzione sono stimati tramite il metodo della massima verosimiglianza. La partizione dell'insieme di eventi è modificata iterativamente in modo da incrementare la verosimiglianza globale. Questa tecnica impiega metodi di stima non parametrica in una procedura complessa di massimizzazione della funzione di verosimiglianza di un processo di punto.

Keywords: Earthquakes clustering, point process, intensity function, MLE.

1. Introduction

Because of the different seismogenic features controlling the kind of seismic release of background and clustered seismicity (Adelfio *et* al, 2005), to describe the seismicity of an area in space, time and magnitude domains, it is useful to study separately the features of *independent* events and strongly correlated ones. The two different kinds of sets of events give different information on the seismicity of an area. For the estimation of parameters of phenomenological laws useful for the description of seismicity, we need a good definition of the earthquake clusters. Furthermore the prediction of the occurrence of large earthquakes (related to the assessment of seismic risk in space and time) is complicated by the presence of clusters of aftershocks, that are superimposed to the background seismicity and shade its principal characteristics.

For this purposes the preliminary subdivision of a seismic catalog in background seismicity (represented by isolated events, that do not trigger any further event, and the mainshock of each seismic sequence) and clustered events is required. At this regard, a seismic sequences detection technique is presented; it is based on MLE of parameters that identify the conditional intensity function of a model describing seismic activity as a clustering-process and representing a slight modification of the ETAS model (Epidemic Type Aftershocks-Sequences model; Ogata, 1988, Ogata, 1998).

2. Conditional intensity function in point processes

The conditional intensity function of a space-time point process can be defined as:

$$\lambda(t,x,y|H_t) = \lim_{\Delta t,\Delta x,\Delta y \to 0} \frac{Pr_{\Delta t \Delta x \Delta y}(t,x,y|H_t)}{\Delta t \Delta x \Delta y} \tag{1}$$

where H_t is the space-time occurrence history of the process up to time t; $\Delta t, \Delta x, \Delta y$ are time and space increments; $Pr_{\Delta t \Delta x \Delta y}(t,x,y|H_t)$ is the history-dependent probability that an event occurs in the volume $\{[t, t+\Delta t) \times [x, x+\Delta x) \times [y, y+\Delta y)\}$. The conditional intensity function completely identifies the features of the associated point process (i.e. if it is independent of the history but dependent only on the current time and the spatial locations (1) supplies a nonhomogeneous Poisson process; a constant conditional intensity provides a stationary Poisson process).

Wider classes of point processes are defined by different conditional intensity functions (Schoenberg and Bolt, 2000).

ETAS model is a self-exciting point process describing earthquakes catalogs as a realization of a branching or epidemic-type point process. In particular it could be considered as an extension of the Hawkes model (Hawkes, 1971), which is a generalized Poisson cluster process associating to cluster centers a branching process of descendants.

The conditional intensity function provides a quantitative evaluation of future seismic activity. It is proportional to the probability that an event with magnitude m will take place at time t, in the epicenter of coordinates (x, y).

3. The proposed clustering method

The clustering technique that we propose leads to an intensive computational procedure, implemented by software R (R Development Core Team, 2005). It identifies a partition of a catalog of seismic events, $\mathcal{P}_{k+1}$, formed by $k+1$ sets: the background seismicity and the clustered events (k disjoint sets). It iteratively changes the partition assigning events either to the background seismicity or to the $j-th$ cluster ($j = 1, \ldots, k$), on the basis of the likelihood function variation due to their moving from a set to another one.

Let $[T_0, T_{max}]$ and Ω_{xy} time and space domains of observation respectively; according to the theory of point process, the likelihood function to be maximized is:

$$\log L(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log \lambda(x_i, y_i, t_i; \boldsymbol{\theta}) - \int_{T_0}^{T_{max}} \int_{\Omega_{xy}} \lambda(x,y,t;\boldsymbol{\theta}) dx dy dt \tag{2}$$

where:

$$\lambda(x,y,t;\boldsymbol{\theta}) = \lambda_t \mu(x,y) + K_0 \sum_{\substack{j=1 \\ (t_j < t)}}^{k} g_j(x,y) \frac{\exp[\alpha(m_j - m_0)]}{(t - t_j + c_j)^{p_j}} \tag{3}$$

and $\boldsymbol{\theta} = (\lambda_t, K_0, c_j, p_j, \alpha)$. In (3) t_j and m_j are time of the first event and magnitude of the mainshock of the cluster j, $g_j(x,y)$ is the space density of the cluster j and $\mu(x,y)$ is the background one; K_0 and λ_t are the weights of the clustered seismicity and of the background one, respectively. Background seismicity is assumed stationary in time, while

time aftershock activity is represented by a non stationary Poisson process according to the modified Omori formula (Utsu, 1961), of parameters c_j, p_j, relating the occurrence rate of aftershocks to the mainshock magnitude (with α measuring the influence on the relative weight of each sequence and m_0 the completeness threshold of magnitude, that is the lower bound for which earthquakes with higher values of magnitude are surely recorded in the catalog).

In our approach space density both of background seismicity and of each cluster, differently by ETAS model, is estimated by a bivariate kernel estimator: it is computed either using only the independent events (isolated and mainshocks) or points belonging to the cluster, including the mainshock, respectively. In both cases the smoothing constant is evaluated with Silverman's formula (Silverman, 1986).

In the evaluation of (3), different kinds of parametrization, that allow to take into account for different assumptions on the seismicity of an area, are considered (e.g. Omori's law parameters of the k clusters can be assumed equal or distinct in each cluster). The choices can be compared at the end of the procedure.

3.1. Main steps

To start, the proposed method needs an initial partition, found by a like single-linkage method procedure. In this first step, choosing a threshold parameter, an initial number of clusters is determined, but then the classification of the events can change, because each event could change its position moving to the set for which the likelihood function is maximized.

The space densities and the ML estimates of the parameters of the intensity function (3) are evaluated. Moving on, for each event U_i, $(i = 1, \ldots, n)$, isolated or belonging to the cluster s, we first find the cluster which maximizes the conditional intensity (3): r is the index corresponding to the maximum of the k contributions in the sum:

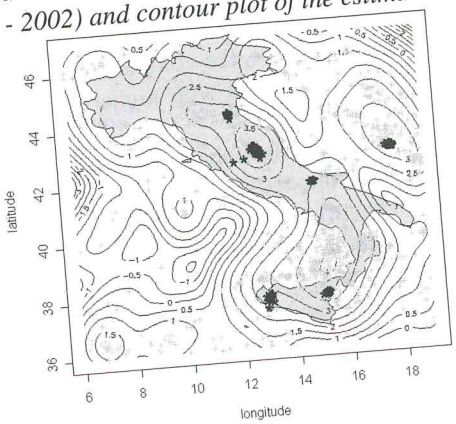
$$\sum_{\substack{j=1 \\ (t_j < t_i)}}^{k} g_j(x_i, y_i) \frac{\exp[\alpha(m_j - m_0)]}{(t_i - t_j + c)^p} \tag{4}$$

Then we try to move all events (one by one) from their current position (either from the cluster s to the cluster r or from the background to the cluster r or from the cluster r to the background) and we compute the change in (2) due to these movements, using the current value of the estimated parameters. If the current value of the likelihood increases the movements become effective. If at least one event changes its position the partition $\mathcal{P}_{k+1}$ is updated and the algorithm restarts from the intensity estimation step; the number of clusters could decrease during the iterative optimization. The iterative procedure stops when the current classification does not change after a whole iteration.

4. Conclusion

Although the proposed method has some critical aspects that have to be solved (as the computational burden and a certain dependence of the convergence of the iterative algorithm on some initial choices) it could be the basis to carry out an analysis of the complexity of the seismogenetic processes relative to each sequence and to the

Figure 1: *Clusters (in black $*$ symbol) and isolated events (in grey $+$ symbol) found in the Italian catalog (1964 - 2002) and contour plot of the estimated space density*



background seismicity, separately. It returns a plausible separation of the different components of seismicity and clusters that have a good interpretability, estimating the space pattern of the induced seismicity through non parametric methods, using only the events of each cluster. It also returns the value of the likelihood function and the estimate of the intensity function for each point (see for example Figure 1).

References

Adelfio, G., Chiodi, M., De Luca, L., Luzio, D., and Vitale, M. (2005) Southerntyrrhenian seismicity in space-time-magnitude domain. *Submitted for publication*.

Gardner, J. and Knopoff, L. (1974) Is the sequence of earthquakes in outhern california, with aftershock removed, poissonian? *Bull. Seismol. Soc. Am.*, 64, pages 1363-1367.

Hawkes, A. (1971) Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58, 1, pages 83–90.

Ogata, Y. (1988) Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the Am. Stat. Ass.* , 83, 401, pages 9–27.

Ogata, Y. (1998) Space-time point-process models for earthquake occurrences. *Ann. Inst. Statist. Math*, vol. 50, no. 2, pages 379–402.

R Development Core Team (2005) R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07- 0*.

Resenberg, P. (1985) Second-order moment of central california seismicity, 1969-1982. *Journal of Geophysical Research*, vol. 90, no. B7, pages 5479–5495.

Schoenberg, F. and Bolt, B. (2000) Short-term exciting, long-term correcting models for earthquake catalogs . *Bull. Seismol. Soc. Am*, 90(4), 849–858.

Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.

Utsu, T. (1961) A statistical Study on the Occurrence of aftershocks. *Geophys. Mag.*, 30, 521–605.

Spatio-temporal modelling for PM_{10} in Piemonte (*)

Modellizzazione spazio-temporale per il PM_{10} in Piemonte

Stefano Bande[1,2], Rosaria Ignaccolo[1], Orietta Nicolis[3]

[1] Dipartimento di Statistica e Matematica Applicata, Università di Torino
e-mail: bande@econ.unito.it

[2] Area Previsione e Monitoraggio Ambientale, ARPA Piemonte

[3] Dipartimento IGI, Università di Bergamo

Riassunto: Il materiale particolato (PM) rappresenta uno degli inquinanti dell'aria a maggiore criticità e per esso la normativa in vigore stabilisce valori limite per la protezione della salute umana. Considerando le misurazioni giornaliere del particolato atmosferico con diametro aerodinamico inferiore a 10 μm (PM_{10}) registrate nel 2004 in Piemonte, in questo lavoro viene proposto un modello spazio-temporale che considera le principali grandezze meteorologiche, le coordinate geografiche e il livello di urbanizzazione del sito di rilevamento dei dati.

Keywords: particulate matter, spatio-temporal modelling, meteorological variables.

1. Introduction

Since several epidemiological studies on particulate matter have shown its adverse effects on human health, recent papers deal with its spatio-temporal modeling based on data recorded at different monitoring networks and using different statistical approaches (see Sahu et al. (2005) and the references therein). Using the data recorded in the year 2004 in Piemonte, in this paper we propose a spatio-temporal model for PM_{10}, that takes into account some meteorological variables and other exogenous information, like the geographical coordinates and the urbanisation of the sites. The temporal seasonality is also included, by dummy variables for the day of the week and for the month, while the spatial variability is embedded in a space-time process component.
The paper is organised as follows. In Section 2, we provide a description of the data set with some details on the source of meteorological data. In Section 3 we define the proposed model and finally Section 4 offers results, conclusions and open discussion.

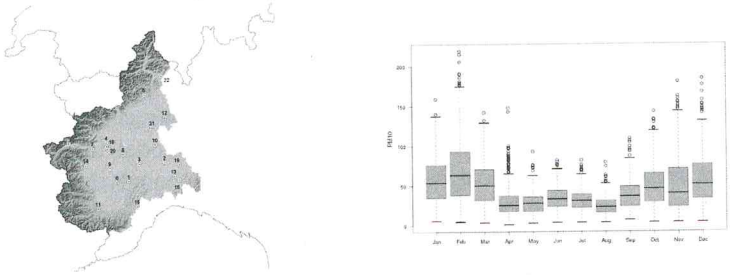
2. The data set

We analyze daily PM_{10} concentrations (in $\mu g/m^3$) during 2004 at 22 irregularly spaced sites in Piemonte, showed in the map in Figure 1(a). The data were collected through the information system *AriaWeb Regione Piemonte.* The 22 records used in this study were selected from Low Volume Gravimetric (LVG) monitors, such that the amount of missing data did not exceed 10%. The missing values were imputed using kernel regression

(*) Work partially supported by MIUR PRIN-Cofin-2004.
The authors wish to kindly thank Giorgio Arduino, Massimo Muraro, Maria Elena Picollo and Marco Rossino for providing useful information and access to data.

smoothing with adaptive plug-in bandwidth (Gasser et al. 1991).
Each of the 22 sites is classified by *area-type* which has three categories (rural, suburban and urban) depending on the edification of the area, and *site-type* which has two categories (traffic and background). The distribution of the PM_{10} concentrations by area-type supports the hypothesis that the pollution levels are lower in rural areas. Fig. 1(b) reveals the seasonal pattern of PM_{10} across months, strongly linked at the different meteorological conditions during the year, and after we consider also the seasonality relative to the day of the week.
Meteorological variables for each site are obtained from the MINERVE (Aria Technologies, 2001) and SURFPRO (Arianet, 2004) models, implemented by the *Area Previsione e Monitoraggio Ambientale* of ARPA Piemonte, and used for the air quality assessment evaluation in Piemonte for the year 2004 (Muraro and De Maria, 2005). MINERVE is a 3D diagnostic meteorological model which produces a mass-consistent wind field using data (both near-ground and upper-air data) from a dispersed meteorological network. SURFPRO provides the gridded fields of main turbulence PBL (Planetary Boundary Layer) scale parameters. Such models provide the estimates on a regular grid with a resolution of *4 km×4 km* for some meteorological variables; in particular we have chosen daily mean solar radiation (RAD), daily temperature range (DT), daily maximum value of wind speed (WS). The solar radiation is the most important variable in the surface energy budget that affects the turbulent properties of the PBL.

Figure 1: *Localization of the sites (a) and distributions of PM_{10} by month (b).*



(a) (b)

PBL turbulent properties will actually rule whether pollutants are dispersed and diluted or they build up and lead to pollution episodes. Moreover local temperature and temperature range seem to be relevant in determining different pollutants behaviors and probable accumulation features. Finally the daily maximum values of wind speed have been used in order to take into account the pollutants removal due to strong wind episodes, often joint to foehn conditions, that frequently occur in Piemonte during the winter-time.

3. The spatio-temporal model

The model developed here is applicable for spatio-temporal data recorded at n sites s_i, $i = 1, \ldots, n$, over a period of T equally spaced time points. Let $z(s_i, t)$ denote the square-root of the observed PM_{10} level at site s_i and at time t where $i = 1, \ldots, n = 22$ and $t = 1, \ldots, T = 366$. Let $\mathbb{1}_S(s_i), \mathbb{1}_U(s_i), \mathbb{1}_T(s_i)$ denote the dummy

variables for the area and site type, respectively if the site s_i is classified Suburban, Urban or Traffic. For every site s_i we consider also the geographical coordinates $x(s_i), y(s_i), z(s_i)$ for the longitude, the latitude and the altitude, and the meteorological variables $RAD(s_i,t), DT(s_i,t), WS(s_i,t)$ for every day. Moreover we define seasonal indicators for the week of the day as $v(t,d) = 1$ if the time t is in the dth weekday (with Sunday = 1), 0 otherwise, and the monthly seasonal indicators $u(t,m) = 1$ if the time t is in the mth month, 0 otherwise. We assume for the data the model

$$Z(s_i,t) = \mu(s_i,t) + w(s_i,t), \qquad i = 1,\ldots,n, \qquad t = 1,\ldots,T, \tag{1}$$

where $\mu(s,t)$ is a mean function depending on the exogenous variables, given by

$$\mu(s,t) = \beta X(s,t) + \sum_{d=2}^{7} \delta_d v(t,d) + \sum_{m=2}^{12} \gamma_m u(t,m), \tag{2}$$

that takes into account all the covariate information in

$$X(s,t) = (1, \mathbb{1}_S(s), \mathbb{1}_U(s), \mathbb{1}_T(s), x(s), y(s), z(s), RAD(s,t), DT(s,t), WS(s,t))'$$

and $w(s,t)$ is a spatio-temporal process with variogram $2\gamma(h,u) = Var(w(s+h,t+u) - w(s,t))$.

The coefficients β, δ_d and γ_m in (2) can be fitted using OLS method and the trend part can be estimated by $\hat{\mu}(s,t) = \hat{\beta}X(s,t) + \sum_{d=2}^{7} \hat{\delta}_d v(t,d) + \sum_{m=2}^{12} \hat{\gamma}_m u(t,m)$. We can consider $\hat{w}(s,t) = Z(s,t) - \hat{\mu}(s,t)$ as a proxy for the spatio-temporal process $w(s,t)$ and estimate it at the spatial location s_0 (even not sampled) for the time-point $t_0 \in \{1,\ldots,T\}$ by the Ordinary Kriging predictor $K_w(s_0,t_0)$ (Stein, 1999), fitting a spatial variogram for the choosen time-point t_0. Hence, the prediction in (s_0,t_0) for the original process (1) is obtained by $\hat{Z}(s_0,t_0) = \hat{\mu}(s_0,t_0) + K_w(s_0,t_0)$.

4. Discussion and results

This paper is about some preliminary results and developments regarding the ongoing research. The estimates of the parameters in the trend part (2) obtained by OLS method are summarized in Table 1. The parameter estimates of monthly seasonal indicator and daily seasonal indicator values show a strong seasonality in PM10 concentrations both across months and days. The seasonality across months is linked to the different dispersion properties of the atmosphere through the year. The seasonality across days is due to the weekly emission cycle. The area-type has the expected effect and indicates the increase in PM_{10} levels due to urbanisation. The meteorological covariate effect seems not to be relevant, except for the wind. A negative relationship has been estimated between the wind speed and the PM_{10} concentrations, according to the pollutant removal caused by strong winds.

On the "residuals" $\hat{w}(s,t)$ a Matern model for the variogram is fitted for every day t and it is used to obtain the prediction $K_w(s_0,t)$ at s_0. Table 2 shows the results of a cross-validation study over the 22 sites in term of RMSE, BIAS and NMSE.

The final aim of our work is to make a forecast map of PM_{10} concentrations over a regular grid. In order to accomplish this we need a spatial model for the type-area and type-site,

Table 1: *The estimates of the parameters with standard deviations (in brackets).*

Interc.	18.963 *(2.22)*	RAD	-0.001 *(0.00)*	Fri	0.836 *(0.06)*	Jul	-2.002 *(0.11)*
Sub.	0.202 *(0.06)*	DT	0.076 *(0.01)*	Sat	0.534 *(0.06)*	Aug	-2.574 *(0.10)*
Urb.	0.719 *(0.07)*	WS	-0.380 *(0.02)*	Feb	0.659 *(0.09)*	Sep	-1.703 *(0.09)*
Traf.	0.194 *(0.04)*	Mon	0.621 *(0.06)*	Mar	-0.300 *(0.09)*	Oct	-0.631 *(0.08)*
x	-0.003 *(0.00)*	Tue	0.847 *(0.06)*	Apr	-1.946 *(0.09)*	Nov	-0.753 *(0.08)*
y	-0.002 *(0.00)*	Wed	0.989 *(0.06)*	May	-1.974 *(0.10)*	Dec	-0.293 *(0.08)*
z	-0.004 *(0.00)*	Thu	1.010 *(0.06)*	Jun	-1.625 *(0.11)*		

which is actually ongoing.

We let the variogram parameters change through time to consider a spatial structure eventually different for every day. The application of the Fuentes separability test (Fuentes, 2006) on our dataset leads to reject the null hypothesis so that a nonseparable space-time covariance will be considered in order to improve the proposed model.

Table 2: *RMSE, BIAS and NMSE for the following validation set: 1 =AL - Nuova Orti; 2 =Alba; 3 = AT - Scuola D'Acquisto ; 4 = Borgaro; 5 = Borgosesia; 6 = Bra; 7 = Buttigliera Alta; 8 = Buttigliera d'Asti; 9 = Carmagnola; 10 = Casale Monferrato; 11 = CN - Piazza II Reggimento Alpini; 12 = NO - Viale Roma; 13 = Novi Ligure; 14 = Pinerolo; 15 = Saliceto; 16 = Serravalle Scrivia; 17 = Susa; 18 = TO - I.T.I.S. GRASSI; 19 = TO - Via Consolata; 20 = Tortona; 21 = VC - Corso Gastaldi; 22 = Verbania.*

i	RMSE	BIAS	NMSE	s	RMSE	BIAS	NMSE	i	RMSE	BIAS	NMSE
1	0.608	10.541	0.1374	9	0.605	8.419	0.123	17	0.778	5.717	0.250
2	0.462	-7.688	0.117	10	0.843	13.491	0.185	18	0.791	-24.534	0.163
3	0.469	4.589	0.118	11	0.787	4.392	0.258	19	0.337	-0.574	0.073
4	0.423	1.378	0.105	12	0.441	-6.334	0.101	20	0.394	0.133	0.073
5	0.637	-4.861	0.263	13	0.459	-3.699	0.122	21	0.464	-4.593	0.102
6	0.622	-12.160	0.183	14	0.603	6.559	0.158	22	0.784	8.962	0.262
7	0.438	0.162	0.114	15	0.685	-5.575	0.336				
8	0.430	-3.082	0.124	16	0.757	3.156	0.243				

References

Fuentes M. (2006) Testing for separability of spatialtemporal covariance functions *Journal of Statistical Planning and Inference*, 136, 447–466.

Gasser T., Kneip A., Koehler W. (1991) A flexible and fast method for automatic smoothing, *Journal of the American Statistical Association*, 86, 643-652.

Muraro M., De Maria R. (2005) Modelling applications and developments for air quality assessment in Piemonte *Workshop "Air quality assessment and management in the Piemonte region according to European Legislation"* Torino, 28 Ottobre 2005.

Sahu S. K., Gelfand A. E. and Holland D. M. (2005) Spatio-temporal modeling of fine particulate matter. *Available from www.maths.soton.ac.uk/staff/Sahu*

Stein M. (1999) *Interpolation of Spatial Data: Some Theory of Kriging.*, Springer, New York.

A hidden Markov model for daily rainfall occurence over southern Calabria Ionica (Italy) (⋆)

HMMs per le occorrenze di pioggia sulla Calabria Ionica meridionale

B. Betrò[1], A. Bodini[1], Q. A. Cossu[2], G. Gullà[3], O. Terranova[3]

[1] CNR-IMATI, Milano
e-mail: antonella.bodini@mi.imati.cnr.it
[2] SAR, Sassari
[3] CNR-IRPI, Cosenza

Riassunto: Si presenta un'analisi statistica della serie temporali giornaliere delle piogge, schematizzate come fenomeno binario, relativamente ad un'area della Calabria Ionica Meridionale. Fra le 45 stazioni disponibili per quest'area sono state selezionate quelle che possiedono un periodo di osservazione comune sufficientemente lungo e completo. In particolare si è considerato il trentennio 1961-1990. Il fenomeno è stato quindi modellato tramite un cosidetto *hidden Markov process*.

Keywords: hidden Markov model, rainfall occurrence.

1. Introduction

The main goal of this work is to examine the probability distribution of local daily rainfall occurrences in part of southern Calabria Ionica to sketch a framework for further studies aiming to define possible scenarios of landslide-triggering due to rainfall. Despite a large literature, the high frequency of landslide events makes necessary deep researches to determine a complete background where the problem of soil slip could be analyzed in a systematic way.

2. Observed rainfall data

We use daily rainfall data collected at 5 stations from the southern Calabria Ionica, listed in Table 1. Among the 45 available, these stations are close to the area of interest and have complete records for the standard 30 years period 1961-1990 (SP), according to the *World Meteorological Organization* suggesting a length of about 30 years for the analysis of weather features, (Thom, 1966). In fact, the current standard period is 1971-2000; our choice is due to the outstanding lack of data for the period 1991-2000. We select the 182-days period beginning October 1 (October-March, OM) corresponding to about 80% of the annual precipitation and to 70% of the raining days. A wet day is defined as having a rainfall of over 0.1mm. In the case of leap years, the rainfall amount at the extra day (i.e., February 29) is added to March 1. The geographical coordinates of the stations together with some rainfall features are shown in Table 1. For comparison, Table 1 contains both

(⋆) The authors are very grateful to Ing. G. Iiritano (Arpacal, Cosenza, Italy) for useful explanations.