

Spatio-temporal modelling for PM₁₀ in Piemonte ^(*)

Modellizzazione spazio-temporale per il PM₁₀ in Piemonte

Stefano Bande^{1,2}, Rosaria Ignaccolo¹, Orietta Nicolis³

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 (PM₁₀) 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 PM₁₀, 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 PM₁₀ concentrations (in $\mu\text{g}/\text{m}^3$) 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.

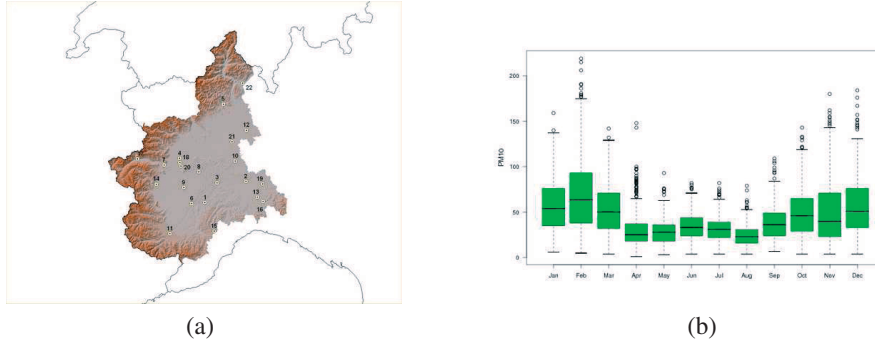
The authors wish to kindly thank Giorgio Arduino, Massimo Muraro, Maria Elena Picollo and Marco Rossino for providing useful information and access to data.

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 $4\text{ km} \times 4\text{ km}$ 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).



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 , $i = 1, \dots, n$, 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 $i = 1, \dots, n = 22$ and $t = 1, \dots, T = 366$. Let $1_S(s_i)$, $1_U(s_i)$, $1_T(s_i)$ 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 d th weekday (with Sunday = 1), 0 otherwise, and the monthly seasonal indicators $u(t, m) = 1$ if the time t is in the m th month, 0 otherwise. We assume for the data the model

$$Z(s_i, t) = \mu(s_i, t) + w(s_i, t), \quad i = 1, \dots, n, \quad t = 1, \dots, T, \quad (1)$$

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

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

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 $2\gamma(h, u) = \text{Var}(w(s + h, t + u) - w(s, t))$.

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 $t_0 \in \{1, \dots, T\}$ 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 PM₁₀ 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 PM₁₀ 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 PM₁₀ 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.