Web Working Papers
by
The Italian Group of Environmental Statistics

Gruppo di Ricerca per le Applicazioni della Statistica
ai Problemi Ambientali

www.graspa.org

Spatio-temporal analysis of PM10 concentrations in the Taranto area

Alessio Pollice and Giovanna Jona Lasinio

GRASPA Working paper n.31, July 2008

Spatio-temporal analysis of PM10 concentrations in the Taranto area

Alessio Pollice
Dipartimento di Scienze Statistiche "Carlo Cecchi"
Università degli Studi di Bari

Giovanna Jona Lasinio
Dipartimento di Statistica, Probabilità e Statistiche applicate
Università di Roma "La Sapienza"

1 Introduction

An analysis of air quality data is provided for the municipal area of Taranto, characterized by high environmental risks due to the massive presence of industrial sites with environmental impacting activities along the NW boundary of the city conurbation. Such activities include iron production (one of the largest plants in Europe), oil-refinery, cement production, fuel storage, power production, waste materials management, mining industry and many others. Some more environmental impacting activities are more deeply integrated within the urban area and have to do with the presence of a large commercial harbour and quite a few military plants (a NATO base, an old arsenal and fuel and munitions storages). These activities have effects on the environment and on public health, as a number of epidemiological researches concerning this area reconfirm (Primerano et al., 2006). In the context of an agreement between Dipartimento di Scienze Statistiche - Università degli Studi di Bari and ARPA Puglia, air quality data for the municipal area of the city of Taranto were provided, belonging to different monitoring networks pertaining to the regional and municipal government and counting 25 monitoring stations on the whole (AA.VV., 2007). Pollutants continuously monitored by some of the stations include sulphur dioxide (SO2), nitrogen oxide (NOx) and dioxide (NO2), carbon monoxide (CO), benzene, PM10 and ozone. At present validated data are available for only one common yearly operating period corresponding to year 2005.

The present study is focused on particulate matter as measured by PM10 concentrations. PM10 is an ubiquitous pollutant with adverse effects on human health, typically in highly industrialized areas: correlations between pollutant concentrations and the presence of chronic respiratory diseases are established according to several epidemiological analyses (Biggeri et al., 2004). For this study we can rely on 13 stations monitoring PM10, all equipped with analogous instruments based on the Beta absorption technology, either reporting hourly, two-hourly or daily measurements (Menegotto, 2006). Hourly observations of several meteorological variables (pressure, humidity, wind speed and direction, temperature etc.) are also available for 6 weather monitoring stations. Our main objective is to define a suitable statistical protocol aiming to build reliable daily surface estimates of the PM10 concentration over the entire urban area.

Preliminary data analysis involved addressing quite a few data problems: first we obtained a homogeneous time scale for all monitoring stations transforming PM10 data into daily averages. A log transformation was then applied to reach approximate marginal Gaussianity. In Tab.1 a summary of the missing data situation is reported. Missing data are due to both different operational periods of the stations (*staircase missingness*) and occasional malfunction of the sensors (*sparse missing data*); as a consequence an adequate choice between different missing data imputation strategies was required. Two different validation protocols were applied by the networks managers: 5 instruments controlled by the regional government (ARPA) are considered to be more reliable than those validated by the municipal government (GECOM). The latter 8 instruments often appear to overestimate PM10 concentration levels (Fig. 1b), this behavior being only partly attributable to the more peripherical location of the ARPA sensors (Fig. 1a). A *calibration* procedure producing some adjustment of the GECOM data to allow for data comparability was thus deemed necessary and investigated. Finally available weather data are characterized by gaps and unreliable measurements; a unique daily weather database at the city level was then obtained combining the 6 stations data as follows:

- The daily average values of temperature, relative humidity and pressure, the daily geometric means of the solar radiation (excluding measurements from 19:00 to 6:00) and the daily total amounts of rain were

Station	Starting date	% missing
Ancona	01/01/2005	1.10
Camuzzi	01/01/2005	2.19
Carcere	01/01/2005	1.64
Gennarini	01/01/2005	9.32
Stadio	01/01/2005	9.59
Talsano	01/01/2005	9.04
Talsano (A)	01/01/2005	2.74
Testa	01/01/2005	2.74
Paolo VI	15/01/2005	9.86
Peripato	15/01/2005	25.75
Orsini	08/02/2005	17.81
Archimede	07/04/2005	29.04
Statte	07/04/2005	34.79

Table 1: Operating periods starting dates and percentages of missing daily averages (ARPA stations in italic).

considered for the more reliable Machiavelli weather monitoring station. Gaps in the first four series were filled by averaging data available 24h before and after the gap. The pressure series was discretized into two levels, by the threshold value of 975 mbar. Averages of the daily total amounts of rain available for 4 other monitoring stations were used to impute the missing values of the rain series;

- The daily geometric means and maxima of the wind velocity, the daily circular means and the daily prevalent quadrants of the wind direction (SE, SW, NW, NE) were obtained for data coming from the S.Vito monitoring station.

After this initial exploratory stage of the analysis, spatio-temporal modeling of log daily PM10 values is performed within a Bayesian hierarchical framework proposed by Le and Zidek (2006), characterized by the use of time varying weather covariates and a semi-parametric spatial covariance structure. This is one of the few spatio-temporal statistical models for which several applications to PM10 concentration data are available (Le and Zidek 2006, Zidek et al. 2002).

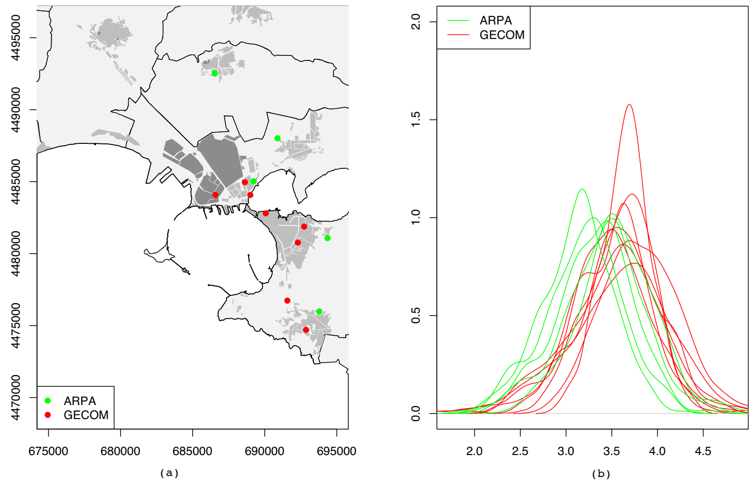

Figure 1: Map of the monitoring stations of the two networks (a) and density estimates of log-average daily PM10 concentrations (b).

The paper is organized as follows. Missing data imputation and calibration are addressed in section 2. Section 3 explores the dependence on weather covariates and the temporal and spatial behaviour of the data. The modeling approach is briefly described in section 4 and in section 5 some results are reported. Section 6 is devoted to the discussion of the proposed strategy and to some concluding remarks.

2 Imputation and calibration

Missing data is a ubiquitous problem in evaluating long-term experimental measurements, such as those associated with air quality monitoring. Spatio-temporal modeling often implies that such gaps in the measured data are filled or imputed. On the other side statistical calibration is often referred to as the process of adjusting the output of a measurement instrument to agree with the values of some specified standard and is intended as a reverse process to regression (Osborne, 1991). So far, no standardized method has been accepted and imputation and calibration methods used are largely dependent on the researchers' choice.

The objective of the method to be described in this section is to obtain a "clean" database by imputing missing values and adjusting data recorded at presumably overestimating (GECOM) stations. Here the basic idea is to preserve and exploit the spatial correlation of the observed PM10 concentrations, recursively estimating daily spatial interpolation models in order to predict missing and overestimated data (Pollice and Jona Lasinio, 2008). Hierarchical Bayesian models embracing properly defined spatial autocorrelation structures can admit any pattern of missing measurements in a partially observed spatial process, as this approach provides a predictive distribution that can be used for imputation. The usual LME model, specified in two hierarchical levels, is chosen as the daily spatial interpolation model (Diggle and Ribeiro, 2007):
Level I - daily data process: Y is a p-dim GRF representing PM10 log daily mean concentrations

$$Y|\beta, \phi, \tau, \sigma^2 \sim N_p\left(\beta, V_y\left(\frac{\tau^2}{\sigma^2}, \phi\right)\right)$$

Level II - prior specification:

- diffused priors for β and σ^2
- discrete priors on a specified reference grid for covariance structure parameters $\tau^2_{rel} = \tau^2/\sigma^2$ and ϕ

Due to the nonstandard prior structure, the predictive distribution has to be computed by numerical approximation: values of the covariance structure parameters τ^2 and ϕ simulated from their marginal discrete posterior distributions are plugged in the t-type predictive distribution obtained for the fully conjugate case.

The function `krige.bayes` in the `R` library `geoR` is used for the implementation of the following procedure making use of two daily spatial kinds of models specified as Bayesian LME's, namely *prediction models* and *estimation models*. Preliminarily, to properly set the prediction model priors for covariance structure parameters ϕ and τ^2_{rel}, a unique daily estimation model is fitted to available data and posterior estimates are obtained. Within a leave-one-out scheme daily spatial prediction models are then fitted and used to predict each observation to be treated. Priors are daily updated by posterior estimates obtained by the estimation model on the previous day. The spatial variation is thus believed to follow a sort of order 1 time dependence, with daily covariance parameter estimates depending stochastically on those of the day before. This recursive posterior-to-prior model estimation step is repeated updating observations to be treated until convergence is reached.

Letting y be the vector of log-concentrations for a specified days and J the set of indices denoting the monitoring stations to be treated, the whole procedure can be summarized in the following iterative algorithm:

step 0 A discrete uniform prior is chosen for τ^2_{rel} on the interval (0,1) with 0.1 increments, while ϕ is allowed to vary in a discrete sequence between 1 and 7 km with 0.5km incremental value and a reciprocal prior. For day 1 fit the estimation model to vector y where data corresponding to the stations to be treated are omitted. For days 2 to 365 fit the estimation model to vector y of the previous day, where data corresponding to the treated stations (z) are substituted. Obtain daily posterior estimates of ϕ and τ^2_{rel}.

step 1 For $i \in J$ let $y_{(i)}$ be obtained by omitting station i in the vector of daily observations y. Iteratively predict each y_i from $y_{(i)}$ using posterior estimates of ϕ and τ^2_{rel} obtained in the previous step for the prior specification of the prediction models. Store predicted values in vector z and substitute them to corresponding values in y.

step 2 Store the current z values in z_{old} and repeat step 1 to obtain a new z.

step 3 If $|z_{old} - z| < \varepsilon$ ($\varepsilon = 0.0001$) or the iterations number is ≥ 100 stop, otherwise repeat step 2 until convergence.

The entire procedure was investigated and compared to other approaches in Pollice and Jona Lasinio (2008). The algorithm is computationally efficient and convergence is reached largely within the maximum number of iterations (100) for all the 365 days, suggesting its use for longer time series as well. Notice that this technique shows a good capability towards spatial variation reconstruction and time dynamic preservation: the shift obtained

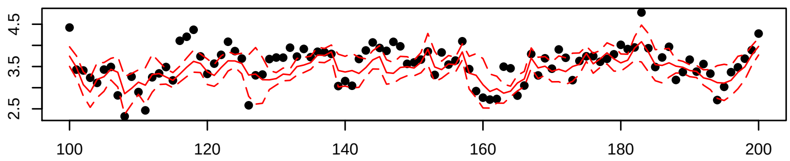

Figure 2: Log-average daily PM10 concentrations for the Testa GECOM monitoring station before (black dots) and after imputation and calibration (red line), with 95% credibility intervals (red dotted lines): days 100 to 200, i.e. 10/4/2005-19/7/2005.

by the adjustment procedure does not alter the dynamics observed in the time series (Fig. 2). Credibility intervals based on the 2.5% and 97.5% percentiles of the simulated predictive distribution prove to be quite narrow and the inspection of the graphs doesn't produce any evidence for the daily IC's size to vary as a function of the number of stations to be treated. Similar results were obtained for all the eight stations belonging to the GECOM network, the details being available from the authors on request.

3 Exploring time and spatial patterns

In order to identify a suitable model structure to predict log-average daily PM10 concentrations we briefly investigate the relationship between meteorological covariates obtained as described in §1 and the pollutant concentrations. In Table 2 linear correlations between log-average PM10 concentrations and continuous meteorological covariates at each monitoring station are reported. Clear indications of the effect of each covariate are evident and agree with known physical facts; the linear relations shown are not very strong but cannot be neglected.

As mentioned in §1 the pressure and the wind direction were transformed into categorical factors, with categories corresponding to high and low pressure and the 4 main wind directions. Boxplots of PM10 concentrations at each monitoring site (not shown) show higher log-PM10 levels when pressure rises above 975mbar and when the wind blows from the south-east direction. Also a meaningful relation between the pollutant concentration

Station	Temperature	Relative Umidity	Solar Radiation	Total Rain	Maximum Wind Speed
Ancona	0.40	−0.21	0.31	−0.27	−0.32
Camuzzi	0.49	−0.33	0.41	−0.26	−0.23
Carcere	0.29	−0.09	0.17	−0.23	−0.33
Gennarini	0.38	−0.21	0.31	−0.28	−0.33
Orsini	0.49	−0.34	0.43	−0.25	−0.22
PaoloVI	0.44	−0.25	0.39	−0.23	−0.37
Peripato	0.46	−0.28	0.37	−0.27	−0.28
Stadio	0.41	−0.22	0.31	−0.27	−0.32
Talsano	0.36	−0.20	0.31	−0.28	−0.33
TalsanoF	0.25	−0.12	0.26	−0.28	−0.36
Testa	0.46	−0.28	0.38	−0.26	−0.28
Archimede	0.48	−0.41	0.44	−0.19	−0.05
Statte	0.25	−0.07	0.24	−0.20	−0.39

Table 2: Linear correlation between log-PM10 at each monitoring station and meteorological continuous covariates

and the month of the year is found at all stations. The relevance of the covariates was further verified fitting a unique linear regression model by pooling the 13 log-PM10 series: all effects were found highly significant.

For PM10 concentrations a strong daily dependence is expected, due to the high atmospheric lifetime of smaller size particles (Cocchi et al., 2007). Adjusting the data according to the procedure proposed in section 2 does not alter the observed autoregressive time-correlation structure: after the sparse missing data imputation and calibration step we are left with a database characterized by staircase missingness, accounting for different activity periods (Tab. 1). Alternatively a complete database is obtained by imputing also the initial gaps in the data according to the above described procedure. In both cases the time series are characterized by a strong daily time correlation structure, remarkably consistent across all sites. In this section we present some analyses applied to data characterized by staircase missingness, but results are substantially similar for both datasets. Autoregressive and other unpublished analyses lead to the adoption of a single AR(3) model fitted across all 13 monitoring sites (Fig. 3).

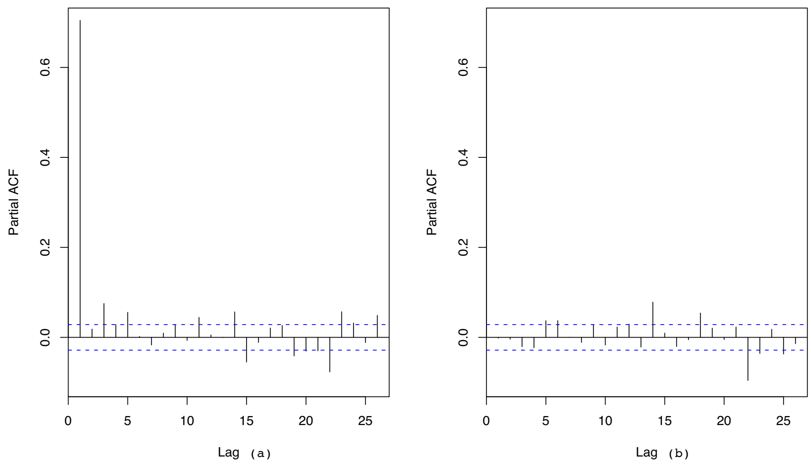

Figure 3: PACF's of spatially pooled normalized and adjusted daily mean concentrations before (a) and after (b) subtracting the estimated AR(3) model.

Residuals of a single AR(3) model estimated by pooling the 13 time series of PM10 daily log-average concentrations after imputation and adjustment don't show a significant correlation at lower lags (Fig. 3 (b)). The variation in the residuals can thus be expected to arise from variation due to space only. The same procedure was applied to remove the autoregressive variation from the log-average concentrations series before imputation and calibration. A comparison of the average variogram estimators (Sahu and Mardia, 2005) of the two series shows that the kriging-based method in §2 tends to preserve the spatial variation adjusting the GECOM data by values matching the fitted spatial structure (Fig. 5 (a) and (b)) and thus leads to a substantial reduction of the nugget effect (Fig. 4). This is a desirable feature of the procedure as we expect that a large part of the measurement error is due to the different calibration of the two networks.

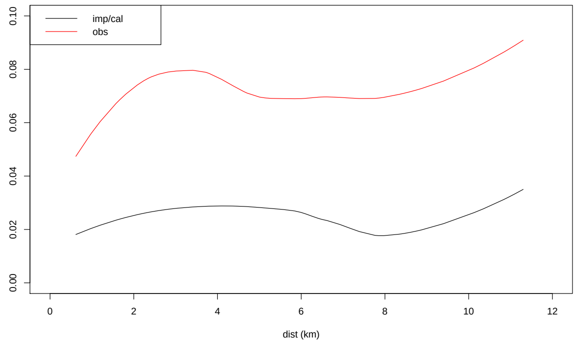

Figure 4: Smooth loess curves interpolating empirical average variogram clouds for time detrended observed and adjusted data.

In order to separately model temporal and spatial variability we verify the absence of the so called *spatial correlation leakage* (Zidek et al., 2002) clearly shown in Fig. 5 (b) and (c): the procedure in §2 reconstructs the spatial correlation on the basis of the existing spatial information, this leads to an overall increase in the correlogram values in passing from Fig. 5 (a) to Fig. 5 (b) where smaller values corresponding to higher distances are evident; in Fig. 5 (c) the subtraction of the AR(3) temporal trend does not imply an overall decrease in the correlogram (the spatial structure is not diminished).

4 Spatial interpolation

The spatial predictive distribution was obtained by the Bayesian kriging-based model proposed by Le and Zidek (2006) and characterized by the use of time varying covariates and a semi-parametric nonstationary covariance structure. Among its attractive features this approach includes the possibility of accounting for data from k blocks of stations having different operational periods and the availability of a multivariate extension for multiple pollutants. It is specified in the form of a Bayesian hierarchical model:

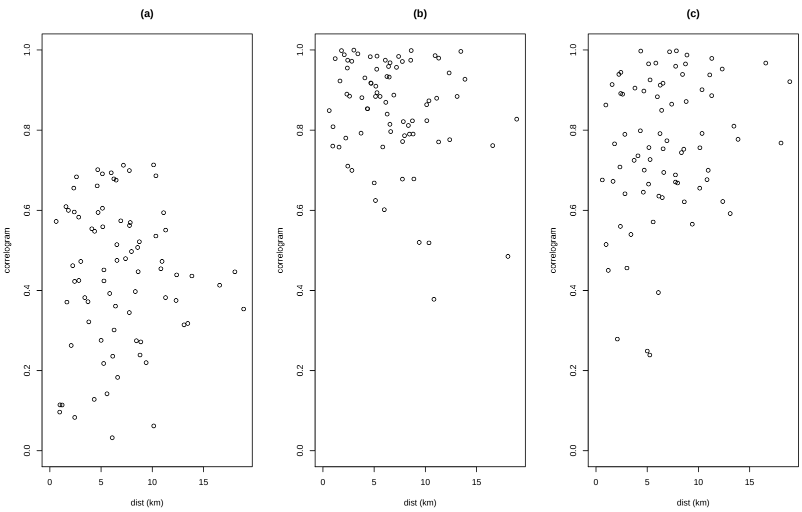

Figure 5: Correlograms for: (a) observed log-daily average PM10 concentrations, (b) observed log-daily averages after sparse missing data imputation and calibration of stations belonging to the GECOM network, (c) residuals of adjusted log-daily averages after the subtraction of the AR(3) temporal trend.

Level I - daily data process:

$$Y|Z,B,\Sigma \sim N_s(ZB,\Sigma)$$

Here Y is a s-dim GRF representing PM10 log daily mean concentrations at s gauged and ungauged sites, Z is a $s \times sr$ matrix containing replicates of the r-dimensional vector of daily regressors (constant across sites, e.g. daily weather data at the city level), while regression coefficients in vector B are admitted to vary over sites. Conditionally on the corresponding covariates in Z, replications of Y are assumed to be independent over time.

Level II - conjugate prior specification:

$$B|B_0,\Sigma,F \sim N_{rs}(B_0,F^{-1} \otimes \Sigma)$$

$$\Sigma|\Psi,\delta \sim \mathrm{GIW}(\Psi,\delta)$$

where F^{-1} is the between covariates variance component of B and GIW stands for the generalized inverted Wishart distribution with multiple degrees of freedom parameters $\delta = (\delta_1, \ldots, \delta_k)$, representing uncertainty associated with k different operational periods.

Due to the previous conjugate specification, the explicit expression of the predictive distribution is obtained as a product of matric t-distributions depending on hyperparameters B_0, F, Ψ and δ. Such hyperparameters are estimated by the following two-step procedure:

step 1 At the gauged sites (monitoring stations) parameter estimates are obtained by EM marginal likelihood maximization (empirical Bayes/type-II MLE);

step 2 At the ungauged sites (grid points) the respective covariance and cross-covariance components of Σ are obtained by the Sampson-Guttorp method (Sampson & Guttorp, 1992). The method is based on constructing a thin-plate splines smooth mapping between locations in the geographic space, where stationarity of the random field is not assumed, to locations in a (virtual) new space where isotropy is assumed. Multidimensional scaling is used to obtain new locations for which the isotropy assumption is appropriate and an isotropic variogram model is fitted using the observed correlations and distances in the new space. The smooth mapping function, together with the isotropic variogram model enables to estimate the spatial dispersion between the stations and the ungauged sites.

The estimate of the spatial covariance is used to obtain the spatial predictive distribution. Its expectation or the mean of a specified number of simulations at selected grid points can be used to interpolate the daily log PM10 fields.

The method has some clear theoretical advantages including the consideration of a very flexible spatial covariance structure and explicit expressions of posterior distributions enabling to avoid computationally cumbersome MCMC estimates. Computations are also made easy by a suite of R functions implementing the above estimation/prediction framework, available at `http://enviRo.stat.ubc.ca`. For the sake of completeness we report that the need for sparse missing data imputation and for filtering the time variability due to the conditional time-independence assumption increases the multi-step feature of the whole procedure with a consequent loss of control over the its overall variability.

5 Some results

In compliance with the conditional temporal independence assumption, the time detrended daily residuals are used to estimate the model described in §4, with weather covariates obtained as in §1 and accounting for the seasonal effect of the calendar month. We begin by discussing the PM10 concentration data following the staircase missing data structure highlighted in Tab. 1, with four blocks of stations having different starting times. Fig. 6 shows the result of applying the Sampson and Guttorp method: the deformation of the geographic space appears to be consistent with the presence of the sea in the south-western part of the study area (for details on the interpretation see Sampson and Guttorp, 1992). A compromise between the complexity of the mapping and the fit to the parametric variogram model in the isotropic virtual space leads to the choice of the amount of smoothing ($\lambda = 0.0005$). On the other hand, the resulting predictions are not particularly sensitive to the choice of the smoothing parameter (our experience with several attempts and Sun et al, 1998).

The predictive distribution obtained by the estimated spatial covariance is used to interpolate the daily time detrended log PM10 fields on a 434 points grid. These additional prediction locations belong to a 14×31 square lattice with 600m cell side, covering the whole area of interest. The predictive distribution is used to obtain expectations and 1000 simulations at each of the 434 grid-points on each of the 365 days. The estimated AR(3) component of §3 is then added to such interpolated residuals, completing the construction of the spatial predictor. Daily expectations and simulations summaries (means, standard errors, upper and lower quantiles, extremes) at each grid-point are considered as the final output of the modeling strategy and used to assess its behavior and to describe the spatio-temporal diffusion of the pollutant. According to the Bayesian posterior predictive p-values paradigm, daily 90% credibility intervals are obtained by the 5-th and 95-th percentiles of the 1000 simulations from the predictive distribution. Adjusted log-PM10 concentrations at each monitoring station are compared with predictions at the nearest grid-point.

As shown in Fig. 7 some difficulties arise with the prediction of initial

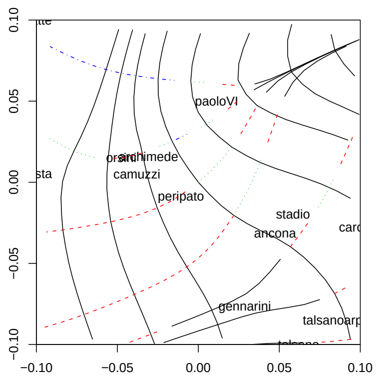

Figure 6: Biorthogonal grid characterizing the deformation of the geographic space obtained to reach approximate isotropy ($\lambda = 0.0005$).

missing data, not only due to the high variability, but mainly to the evidence of unreliable estimates as shown by the simulation summaries for the peripherical Statte station (Fig. 7 top left and Fig. 6 top left). Boundary effects can be excluded as changes in the amount of spatial smoothing and in the dimension of the prediction grid do not cause any substantial improvement upon this situation. Other unpublished attempts based on different data produced imputations characterized by an analogous high variability and very flat prediction surfaces for days with missing data. Predictions are very accurate when data for all the 13 stations become available (Fig. 7 right): the 90% credibility intervals look quite narrow and contain the observed values in the majority of days.

As an alternative, initial missing data are imputed by the procedure in §2 and spatial predictions are obtained as in §4 using the AR(3) model described in 3. The estimated temporal pattern and spatial covariance structure show only slight changes with respect to the results presented for the previous case. As we show in Fig. 9 the iterative strategy proposed in §2 produces imputations characterized by small variability, also for data missing at the beginning of the series. As expected, the neglection of the variability due to missing data imputation avoids the uncontrolled increase in the variability

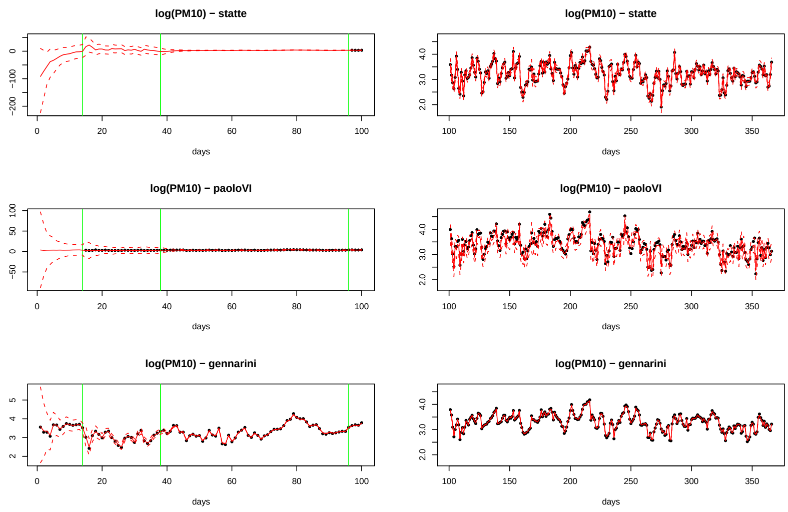

Figure 7: Adjusted log-average PM10 concentrations (black dots) for three monitoring stations and those predicted at the nearest grid points (red line). Staircase missing data pattern included in the prediction model. Dotted lines are 90% credibility intervals, green vertical lines indicate operating period starting dates (15/1/2005, 8/2/2005 and 7/4/2005).

of initial predictions (Fig. 8). Indeed this artifice does not allow to properly take into account the loss of information implied by the presence of missing values. Maps for the initial 15 days, when the number of missing data is larger, are acceptable although flatter than those for the remaining days. For all days spatial predictions show a strong connection with the wind direction (Fig. 10). Model fitting is very satisfactory in terms of RMSE computed with respect to the nearest grid point (Fig.11(a)) and simulations from the predictive distributions are highly concentrated, given the low dispersion values shown in Fig.11(b).

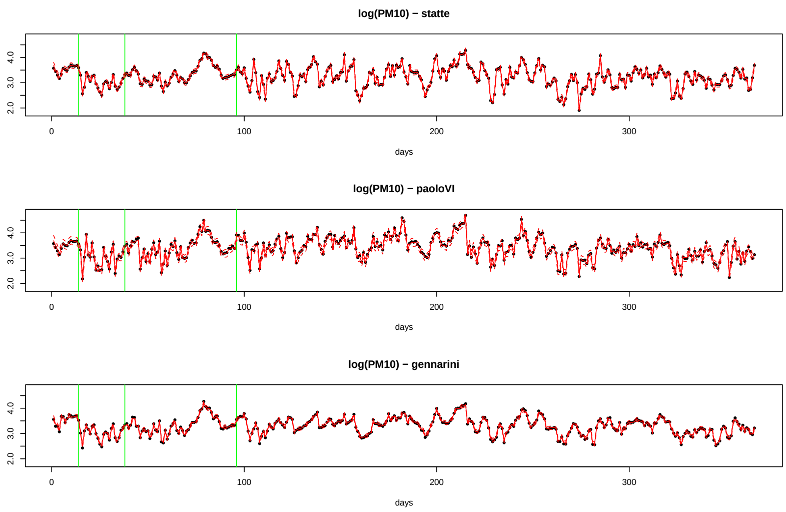

Figure 8: Adjusted log-average PM10 concentrations (black dots) for three monitoring stations and those predicted at the nearest grid points (red line). Staircase missing data imputed by the procedure in §2. Dotted lines are 90% credibility intervals, green vertical lines indicate operating period starting dates (15/1/2005, 8/2/2005 and 7/4/2005).

Most "hot-spots" are found in the vicinity of the iron plant (darker grey area in the maps) and the nearby paolo VI monitoring station. Peaks move south when wind blows from the north-west direction and they are often found near the harbor. Consistently with the empirical knowledge of the PM10 behavior, often lower concentrations are found when it rains considerably (more then 5mm). Maps of 90% credibility intervals based on simulations (not shown), return daily evaluations of plausible values intervals and of the estimates quality (together with the simulations standard deviations).

6 Discussion and concluding remarks

In this study daily estimates of the PM10 concentration surfaces based on 13 monitoring stations were obtained in order to identify areas of higher concentration (hot spots), possibly related to specific anthropic activities.

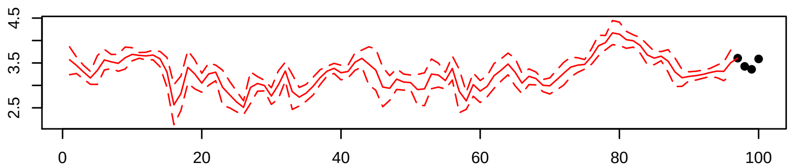

Figure 9: Imputed log-average daily PM10 concentrations for the Statte monitoring station (red line), with 95% credibility intervals (red dotted lines): days 1 to 100, i.e. 1/1/2005-10/4/2005.

PM10 concentrations data were obtained from 5 instruments controlled by the regional government (ARPA) (most reliable) and 8 managed by the municipal government (GECOM) (less reliable). Preliminary analysis involved addressing several data problems, mostly linked to the treatment of missing data and to the calibration between the two networks. In section 2 we dealt with this problem proposing a Bayesian kriging-based technique, using a hierarchical model proposed by Diggle and Ribeiro (2007) enriched by a time-recursive prior structure (for a different approach see Fassò et al., 2007). Imputed/calibrated values seem consistent with the experts empirical knowledge of the PM10 behavior in the area.

Spatio-temporal modeling was then performed within a Bayesian hierarchical framework proposed by Le and Zidek (2006) and briefly described in sections 3 and 4. This approach is characterized by the use of time varying weather covariates and a semi-parametric spatial covariance structure.

The proposed missing data/calibration treatment and the necessary removal of the temporal trend produce a composite estimation strategy for which it is particularly difficult to asses the estimates precision. Indeed ignoring this aspect may seriously affect the final uncertainty evaluation and the use of an integrated model should be considered. On the other side the proposed approach is computationally efficient, unlike many more general Bayesian models involving complex MCMC simulation-based estimation procedures.

We applied this modeling strategies to two different situations: with staircase missingness and imputing initial missing data. As far as the initial period is concerned the staircase missingness approach showed serious problems especially with the peripherical Statte station, returning very unre-

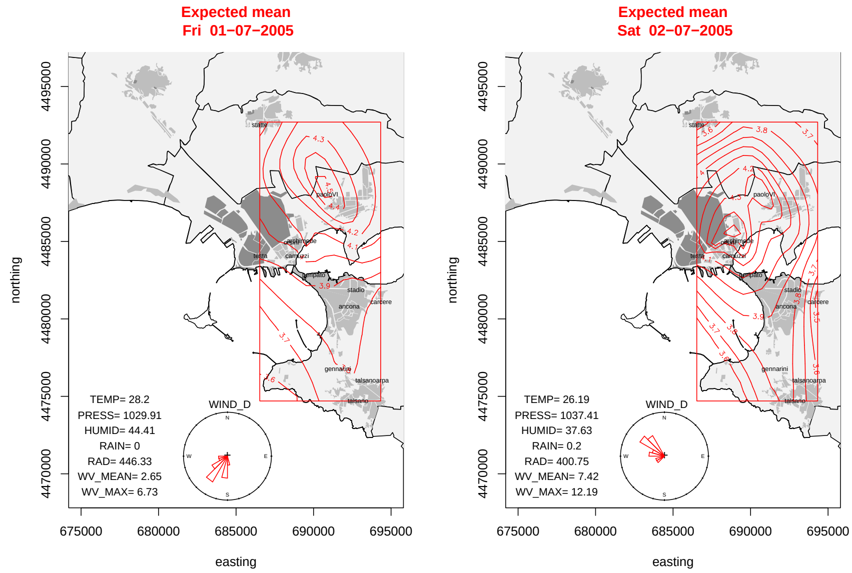

Figure 10: Means of 1000 simulations from the predictive distribution on a 14×31 square grid for two consecutive days.

liable estimates and driving all surface estimates to fairly unreliable results. When using imputed data, surfaces are quite flat in the initial period as not much variability is found in the available data. This point can be properly clarified only when longer time series will be made available as this could also be a typical PM10 winter behavior in the area.

In general terms the proposed protocol returns coherent and satisfactory results with a reasonable computational effort.

References

BIGGERI, A., BELLINI, P., TERRACINI, B. (2004) Metanalisi italiana degli studi sugli effetti a breve termine dell'inquinamento atmosferico. *Epidemiologia e Prevenzione*, **28**.

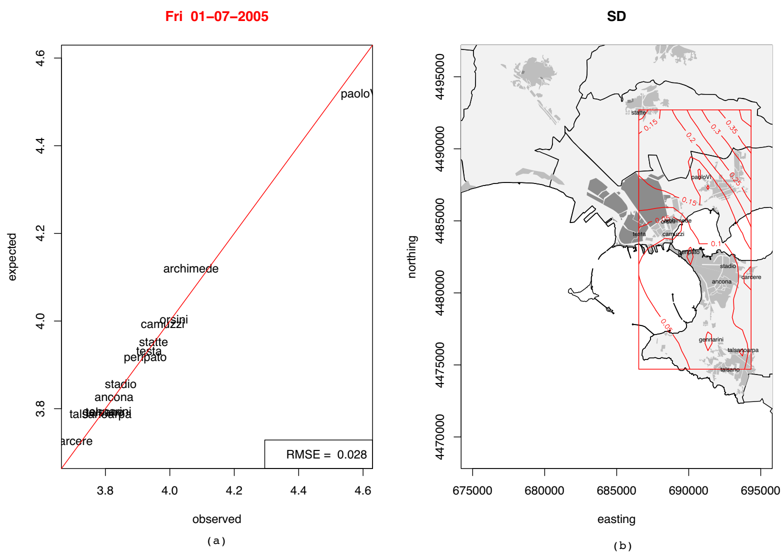

Figure 11: Log-average PM10 concentrations observed on 1/7/2005 and those predicted at the nearest grid point (a). Standard deviations of 1000 simulations from the predictive distribution on a 14×31 square grid for the same day (b).

COCCHI, D., GRECO, F., TRIVISANO, C. (2007) Hierarchical space-time modelling of PM10 pollution. *Atmospheric Environment*, **41**, 532-542.

DIGGLE, P.J., RIBEIRO, P.J. (2007) *Model-based Geostatistics*, Springer.

FASSÒ, A., CAMELETTI, M., NICOLIS, O. (2007) Air quality monitoring using heterogeneous networks. *Environmetrics*, **18**, 245-264.

LE, N.Z., ZIDEK, J.V. (2006) *Statistical Analysis of Environmental Space-Time Processes*, Springer.

MENEGOTTO, M. (2006) Qualità dell'aria nei comuni dell'area ad "elevato rischio di crisi ambientale" di Taranto - Rapporto annuale 2005. *Technical report*, ARPA Puglia.

POLLICE, A., JONA LASINIO, G. (2008) Two approaches to imputation and adjustment of air quality data from a composite monitoring network. *GRASPA working paper*, **30** (www.graspa.org).

PRIMERANO, R., MENEGOTTO, M., DI NATALE, G., GIUA, R., NOTARNICOLA, M., ASSENNATO, G., LIBERTI, L. (2006) Episodi acuti di inquinamento da PM10 nell'area ad elevata concentrazione industriale di Taranto, poster presented at *Secondo Convegno Nazionale sul Particolato Atmosferico - PM2006*, Florence, 10-13 september 2006.

SAHU, S.K., MARDIA, K.V. (2005) A Bayesian Kriged Kalman model for short-term forecasting of air pollution levels. *Applied Statistics*, **54**, 1, 223-244.

SAMPSON P., GUTTORP P. (1992) Nonparametric estimation of non stationary spatial structure. *JASA*, **87**, 108-119.

ZIDEK, J., SUN, L., LE, N., ÖZKAYNAK, H. (2002) Contending with space-time interaction in the spatial prediction of pollution: Vancouver's hourly ambient PM10 field. *Environmetrics*, **13**, 595-613.

SUN, W., LE, N., ZIDEK, J., BURNETT, R. (1998) Assessment of a Bayesian multivariate interpolation approach for health impact studies. *Environmetrics*, **9**, 565-586.