Bayesian Univariate Space - Time Hierarchical Models for Mapping Pollutant Concentrations in the Municipal Area of Taranto

Lorenza Cretarola[1], Giovanna Jona Lasinio[1],
Serena Arima[3], Alessio Pollice[2]

1 Dipartimento di Statistica, Probabilità e Statistiche applicate
Università di Roma "La Sapienza"

2 Dipartimento di Scienze Statistiche "Carlo Cecchi"
Università degli Studi di Bari

3 Dip. di Studi Geoeconomici, Linguistici, Statistici e Storici per l'Analisi Regionale
Università di Roma "La Sapienza"

1 Introduction

The main goal of this work is the proposal and critical analysis of some univariate hierarchical space-time Bayesian models for mapping the concentration of three pollutants in the air shed of the city of Taranto.
The municipal area of Taranto (southern Italy) is 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 other environmental impacting activities are more deeply integrated within the urban area and have to do with the presence of a large commercial harbor 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 ([9]). 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 up to 25 monitoring stations on the whole ([9]).
Pollutants continuously monitored by the stations include sulphur dioxide (SO_2), nitrogen oxide (NO_X) and nitrogen dioxide (NO_2), carbon monoxide (CO), benzene, PM_{10} and ozone. This study is focused on PM_{10}, SO_2 and NO_2 concentrations.
Only six monitoring stations were considered, those producing the longest time series for the three pollutants. In figure1 their locations are reported. NO_2 is a brownish, highly reactive gas that is present in all urban atmospheres. This pollutant can irritate the lungs, cause bronchitis and pneumonia, and lower resistance to respiratory infections. The major mechanism for the formation of NO_2 in the atmosphere is the oxidation of nitric oxide, which is produced by most combustion processes. PM_{10} (particles measuring 10μm or less) identifies those particles likely to be inhaled by humans and to enter deep into alveoli, thereby potentially impacting on health. PM_{10} sources are connected to both human and natural activities. A significant portion of PM_{10} is generated from a variety of human activities including agricultural operations, industrial processes, combustion of wood and fossil fuels, construction and demolition activities, and entrainment of road dust into the air. Natural (nonanthropogenic or biogenic) sources also contribute to the overall PM_{10} problem. These include windblown dust, sea salt and wildfires. SO_2 is a colorless, reactive gas. It is released primarily from burning fuels that contain sulfur (like coal, oil and diesel fuel). Stationary sources such as coal- and oil-fired power plants, steel mills, refineries, pulp and paper mills, and nonferrous smelters are the largest releasers.
In this paper we propose a hierarchical spatio-temporal modeling approach to describe and map daily mean concentrations of PM_{10}, SO_2 and NO_2. The model structure describes explicitly the spatial and temporal relationships within the data and those between pollutants and meteorological variables. This feature allows a better understanding of the pollutants diffusion and generating processes that is crucial to infer of their sources and effects on human health.

Besides, this work is part of a much broader plan, realized with the intent of comparing two different approaches: three separate univariate analyses, one for each pollutant, combining spatial and temporal aspects and a multivariate approach, wherein all three pollutants are modelled simultaneously. The same data have been analyzed in [8], there the authors adopted a multi-step procedure based on the combination of a multivariate hierarchical spatio-temporal model within a Bayesian framework proposed by [7] and an external missing data imputation procedure based on spatial interpola-

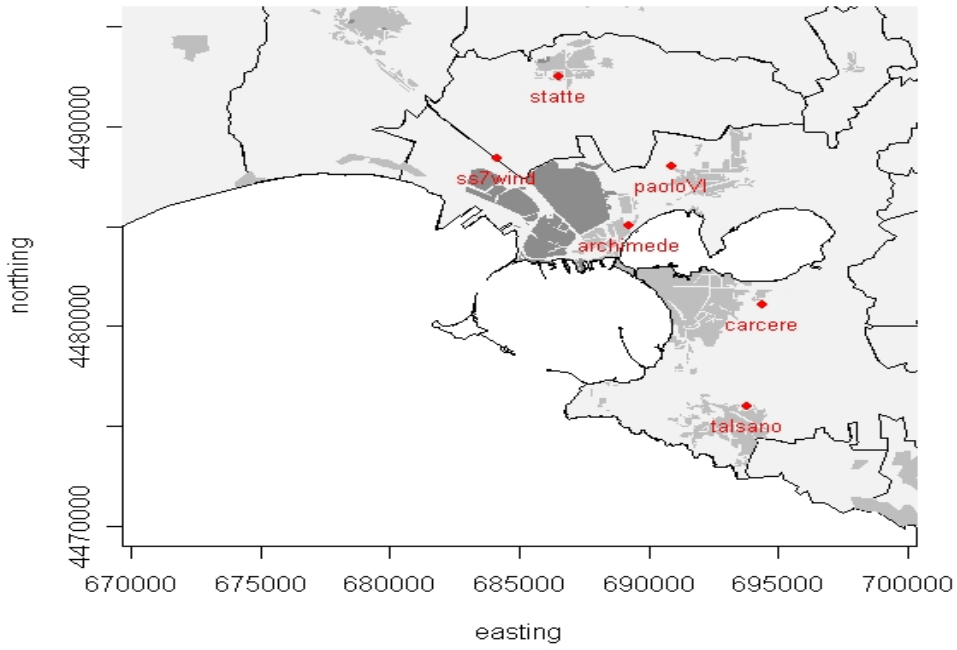

Figure 1: Dislocations of the six monitoring stations

tion, the latter carried on in the Bayesian framework too. The multivariate model, characterized by the use of time varying weather covariates and a semi-parametric spatial covariance structure, does not allow a complete and satisfactory handling of estimates' uncertainty, as it requires a preliminary sparse missing data imputation step outside the model estimation procedure. Here we use a full Bayesian approach as it allows to include missing data imputation within the model estimation step. Estimates are obtained via Monte Carlo Markov Chain (MCMC) routines implemented in WinBUGS.
Our starting point is the model proposed by [10] that we modify and extend to fit our situation. The paper is organized as follows. In section 2 the model is described. Some exploratory analyses and detailed data description together with the final models for each pollutant are illustrated in section 3. Results are discussed in section 4 and some concluding remarks are given in section 5

2 The general model

Stage 1: *observed data model*
Suppose that a pollutant (Y) has been observed at S spatial locations and T time points, along with a set of q meteorological variables. Let Y_{ts} denote the observed level of pollutant on day t (t = 1,...,T) at spatial location s (s = 1...,S), let (C_{1s}, C_{2s}) be the spatial coordinates of site s and let $\mathbf{X}_t$ be the q-dimensional vector of meteorological variables on day t.
At the first level of hierarchy, conditional on the mean (μ_{ts}) and the measurement error variance (σ_s^2), observations are modelled as :

$$Y_{ts}|\mu_{ts}, \sigma_s^2 \sim N(\mu_{ts}, \sigma_s^2) \tag{1}$$

and

$$\mu_{ts} = \gamma_1 C_{1s} + \gamma_2 C_{2s} + \mathbf{X}_t^{'}\boldsymbol{\beta} + \theta_t + \epsilon_{ts} \tag{2}$$

Parameters γ_1 and γ_2 capture the large scale linear spatial trend, while vector $\boldsymbol{\beta}$ captures the dependence of pollutant levels on the covariates. θ_t represents a temporal random effect and the vector $\boldsymbol{\epsilon}_{t.} = (\epsilon_{t1}, \epsilon_{t2}, ..., \epsilon_{tS})$ describes the spatial random effects at time t.

Stage 2(a): *temporal model*
The time dynamic is represented as a random walk:

$$\theta_t = \theta_{t-1} + \omega_t, \qquad \omega_t \sim N(0, \sigma_\theta^2) \tag{3}$$

Stage 2(b): *spatial model*
We assume that the spatial and temporal processes are separable and that at each time t, the vector $\boldsymbol{\epsilon}_{t.} = (\epsilon_{t1}, \epsilon_{t2}, ..., \epsilon_{tS})$ is a zero mean, isotropic Gaussian process with $S \times S$ correlation matrix Σ

$$\boldsymbol{\epsilon}_{t.}|\sigma_\epsilon^2, \Sigma \sim MVN(\mathbf{0}_S, \sigma_\epsilon^2\Sigma) \tag{4}$$

The sill parameter σ_ϵ^2 plays the role of the zero-distance variance. The *ss'* entry of the correlation matrix represents the correlation between sites s and *s'* assumed to be exponential:

$$\Sigma_{ss'} = exp\ (-\phi d_{ss'}) \tag{5}$$

where $d_{ss'}$ is the distance between sites s and *s'*.
Stage 3: *hyperpriors*
Model hierarchy is completed by prior specification for the hyperparameters.

A Gaussian prior is assumed for the regression coefficients γ_1, γ_2 and β_i ($i = 1,...,q$). The final settings for the three models are reported in detail in the appendix.

3 Data processing and variable selection

The data set contains validated data for years 2005-2007, available for only 6 monitoring stations managed by the Apulia regional government, all equipped with analogous instruments either reporting hourly, two-hourly or daily measurements. Hourly observations of several meteorological variables (including temperature, relative humidity, pressure, rain, solar radiation, wind speed and direction) are also available for the same time period and for 3 weather monitoring stations. Our main objective is to integrate pollution and meteorological data in order to summarize the behaviour of pollution diffusion processes over the area of the municipality for the study period (1st January 2005 - 31 December 2007). As already mentioned we focus on three pollutants PM_{10}, NO_2 and SO_2.

Preliminary data analysis involved addressing quite a few data problems: first we obtained a homogeneous time scale for all monitoring stations transforming the data into daily averages. Normalizing transformations were then applied in order to reach approximate marginal Gaussianity: the square roots of the logs of SO2 and the logs of PM10 and NO2 daily averages were considered. In Table 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).

	Archimede	**Carcere**	**PaoloVI**	**SS7wind**	**Statte**	**Talsano**
PM_{10}	321(29%)	98(9%)	144(13%)	184(17%)	199(18%)	23(2%)
SO_2	183(17%)	109(10%)	176(16%)	206(19%)	93(8%)	25(2%)
NO_2	209(19%)	120(11%)	202(18%)	214(20%)	159(14%)	71(7%)

Table 1: Missing daily averages (%)

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 3 stations data. As a first step one of the three stations was chosen as the main source of data. More reliable pressure and solar radiation measurements recorded by each of the other two monitors were

considered. Then daily averages were obtained by arithmetic mean (temperature, relative humidity, pressure), geometric mean (wind speed, solar radiation), circular mean (wind direction), mode (wind direction - quadrants), maximum (wind speed), sum (rain). Missing daily values were imputed by averaging hourly data recorded 12h before and after the gap. Only rain levels were imputed as averages of those recorded at the other two stations. Not all variables were considered like possible covariates for the construction of the models. Their relevance as covariates was verified by fitting linear regression models: conditional OLS estimates were obtained for the normalized pollutant concentration levels at the 6 sites with weekday and month calendar variables and all meteorological covariates as explanatory variables. Concentration levels were overall significantly affected by the effects of weekday, calendar month, temperature, humidity, rain, maximum wind speed and wind direction quadrant. To these, we added the spatial coordinates of sites.

4 Results

In the proposed model the missing data problem can be easily solved as missing data can be seen as a further set of parameters to be estimated via MCMC.

We carried on separate analyses for each pollutant. Starting from the model described in section 2 we verified MCMC convergence for several model structures. Two separate chains of 50000 iterations starting from overdispersed initial values were run for each model. A thinning interval of 25 and a burn-in period of 20000 iterations were applied. Convergence was assessed by visual inspection of the chains sample trace plots, and by computing the Gelman and Rubin, Geweke and Raftery and Lewis statistics.

At first, temperature, humidity, rain and maximum wind speed were used as covariates for the three pollutants. After several long runs with unsatisfactory results, in particular for the convergence of the variance parameters chains, the addition of wind direction categorized into quadrants lead to convergence of the estimation process for the mean terms of the models for PM_{10} and NO_2. For the SO_2 model also the calendar month was added. All three models showed convergence problems as far as the time dynamic was concerned. We then modified the time dynamic by replacing the random walk in (3) with the following autoregressive model:

$$\theta_t = \phi_1 \theta_{t-1} + \omega_t, \qquad \omega_t \sim N(0, \sigma^2_\theta) \tag{6}$$

This choice allowed the convergence of the estimation algorithm for the three models.

The final models point out the role of covariates in determining pollutants levels. Increases in the rain amount and maximum wind speed reduce PM_{10}, on the contrary temperature and relative humidity have positive coefficients, in accordance with the PM_{10} production process encouraged by high temperatures during warmer seasons. Also the wind direction is significantly related to the PM_{10} concentration, suggesting the presence of a transport phenomenon of particulate. Fitting the same model to NO_2 shows no significant influence of the rain amount and wind direction, while increases in temperature and relative humidity contribute to the pollutant's production, following a winter to summer reduction of its level. The model for SO_2 highlights how the rain amount and the calendar month have no influence on the pollutant level, while temperature, relative humidity and wind speed and direction contribute significantly. Spatial coordinates play the same role for PM_{10} and SO_2, showing a positive however very small correlation with the pollutants level, suggesting the presence of a positive gradient in the NE direction complying with the possible effect of the sea breeze on the reduction of the levels of these two pollutants. While as far as NO_2 is concerned the spatial behavior is less clear, little variation can be found in the SW direction.

4.1 Prediction

The last part of this work is dedicated to obtain spatial predictions of concentration levels for the three pollutant at unmonitored locations using the proposed model. Prediction of the pollutant level at site s' and time t is obtained by sampling from the posterior predictive distribution $p(\mu_{s't}|Y)$ whose components are:

$$\mu_{ts'}|Y = (\gamma_1|Y)C_{1s'} + (\gamma_2|Y)C_{2s'} + X_t^{'}(\beta|Y) + (\theta_t|Y) + (\epsilon_{ts'}|Y). \quad (7)$$

Justification of the additive form of the predictive distribution is contained in [10]. Samples from the predictive distribution $p(\mu_{s't}|Y)$ are obtained via MCMC. The predictive distribution is used to interpolate the daily normalized pollutant fields on a 400 points grid. These additional prediction locations belong to a 14×31 square lattice with 700m cell side, covering the whole area of interest. We obtained 200 simulations at each of the 400 grid-points on each of the 1095 days. Daily expectations and simulations summaries (means, standard errors, upper and lower 95% credibility interval limits) for the grid points closest to the six monitoring stations (see figure 2) are considered as the final output for the evaluation of the modeling

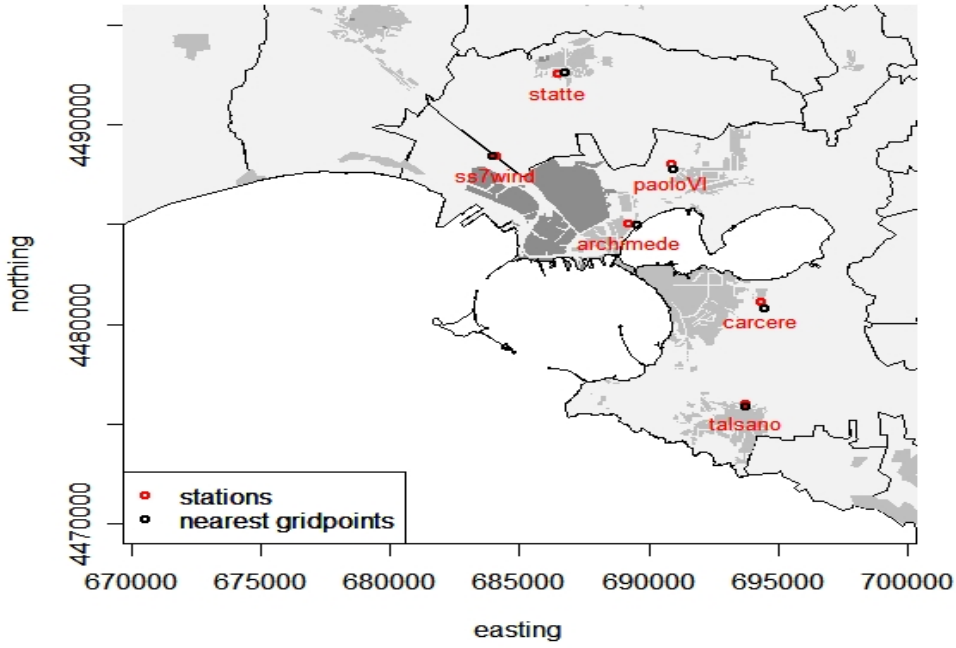

Figure 2: Locations of the six monitoring stations and of the nearest grid points

strategy. In figure 3, the time dynamic of predicted and observed values together with the 95% credibility bounds are reported. The large majority of observed normalized daily concentrations fall inside the corresponding credibility intervals, showing an over-all compliance of the observed data with the simulations from the estimated predictive distribution.

In order to assess model's ability in predicting pollutants levels we consider the following validation statistics:

- root mean squared error (RMSE)

$$RMSE_s = \sqrt{(MSE_s)} = \sqrt{\frac{\sum_{i=1}^{T}(Y_{ts} - \hat{Y}_{ts})^2}{T}},$$

  where Y_{ts} represents the observed normalized pollutant concentrations at time t and monitoring location s and $\hat{Y}_{ts}$ represents predictions at time t and the nearest grid-point;

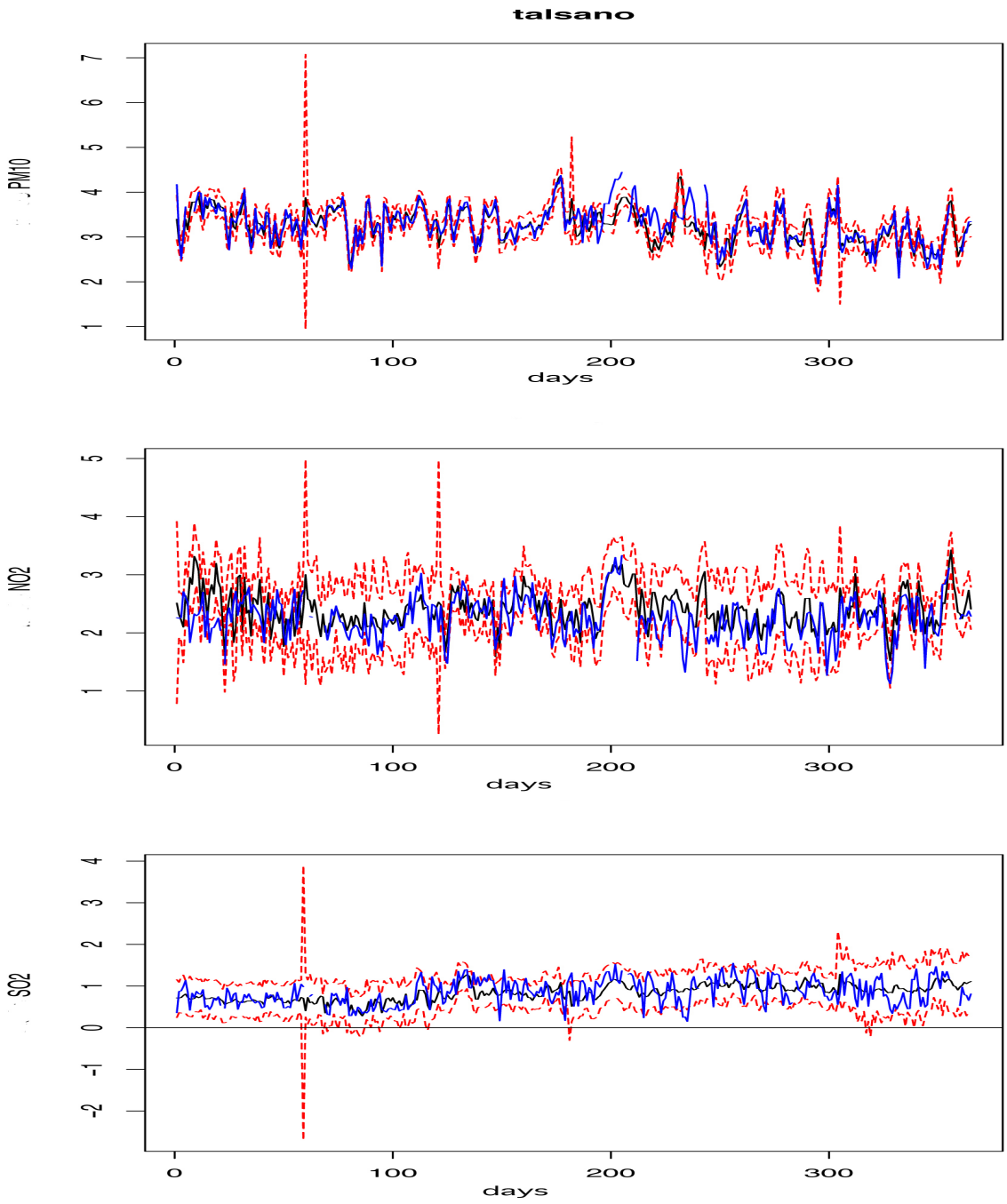

Figure 3: Normalized observed pollutant concentrations (blue line) for the Talsano monitoring station and those predicted at the nearest gridpoint (black line); dotted red lines are 95% credibility intervals. Year 2007.

- CR_1 (Carroll and Cressie, 1996)

$$CR_1 = \frac{1}{S}\sum_{s=1}^{S}\left\{\frac{1}{T}\sum_{t=1}^{T}\frac{(Y_{ts}-\hat{Y}_{ts})}{\hat{\sigma}_{ts}}\right\}$$

  allows us to verify the unbiasedness of the predictors, it should be as close as possible to 0;

- CR_2 (Carroll and Cressie, 1996)

$$CR_2 = \frac{1}{S}\sum_{s=1}^{S}\left\{\frac{1}{T}\sum_{t=1}^{T}\left(\frac{(Y_{ts}-\hat{Y}_{ts})}{\hat{\sigma}_{ts}}\right)^2\right\}^{\frac{1}{2}}$$

  verifies the accuracy of the mean squared prediction error and should be as close as possible to 1.

Tables 2, 3, 4 show RMSE values for the three pollutants, computed at the monitoring stations and over different time windows.

Years	**Root Mean Squared Error**					
	Archimede	Carcere	Paolo VI	SS7Wind	Statte	Talsano
2005	0.4476	0.3801	0.5920	0.5955	0.3225	0.4905
2006	0.3427	0.3695	0.4570	0.4499	0.2428	0.3545
2007	0.3526	0.3062	0.4044	0.3555	0.1770	0.2570
2005 - 2006 - 2007	0.3874	0.3557	0.4966	0.4844	2578	0.3813

Table 2: RMSE's for the six monitoring stations over four time windows: SO_2

Years	**Root Mean Squared Error**					
	Archimede	Carcere	Paolo VI	SS7Wind	Statte	Talsano
2005	0.5881	0.5688	0.6638	0.5357	0.5226	0.5702
2006	0.6603	0.6554	0.6400	0.5422	0.5191	0.5437
2007	0.6326	0.4188	0.4558	0.5581	0.7772	0.3339
2005 - 2006 - 2007	0.6276	0.5617	0.6021	0.5478	0.6039	0.4933

Table 3: RMSE's for the six monitoring stations over four time windows: NO_2

The models prediction performance in terms of pollutants levels is satisfactory, RMSE values are small for all sites and time windows, as RMSE's are

Years	Root Mean Squared Error					
	Archimede	Carcere	Paolo VI	SS7Wind	Statte	Talsano
2005	0.5152	0.3722	0.2778	0.4510	0.2201	0.1179
2006	0.5180	0.2164	0.2038	0.6041	0.2499	0.1064
2007	0.4114	0.2853	0.3231	0.4884	0.2975	0.2495
2005 - 2006 - 2007	0.4939	0.2997	0.2699	0.5172	0.2584	0.1862

Table 4: RMSE's for the six monitoring stations over four time windows: PM_{10}

Year	Index	SO_2	NO_2	PM_{10}
2005	CR_1	-0.2350	-0.1385	0.0163
	CR_2	1.7630	1.3678	2.2205
2006	CR_1	-0.0003	-0.1474	-0.0931
	CR_2	1.6729	2.1789	2.3856
2007	CR_1	-0.0442	0.0909	0.6749
	CR_2	1.4732	2.3566	3.0347
2005 - 2006 - 2007	CR_1	-0.0912	-0.0798	0.1885
	CR_2	1.6526	2.0071	2.6039

Table 5: CR_1 and CR_2 computed between monitoring stations and nearest grid points, over several time windows, for all pollutants

on the same scale as the normalized input data. In Table 5 CR_1 and CR_2 values are reported. The best pollutants' level prediction is obtained for SO_2 in 2006, for NO_2 and PM_{10} in 2005. While an overall slight tendency to overstimation is shown when analyzing the three years at once for SO_2 and NO_2 ($CR_1 < 0$).

To analyse the model behaviour with respect to the time dynamic, we examine the autocorrelation functions for observed concentrations and those predicted at the nearest grid point ACF's and PACF's are shown in figure 4 and 5 respectively for the Archimede monitoring station. It must be noticed that observed ACF's and PACF's estimates are obtained from time series with a large number of missing data and can thus be unreliable. However results are satisfactory on the whole. The autocorrelation structure of observed time series is well reproduced by all three models, while more discrepancies

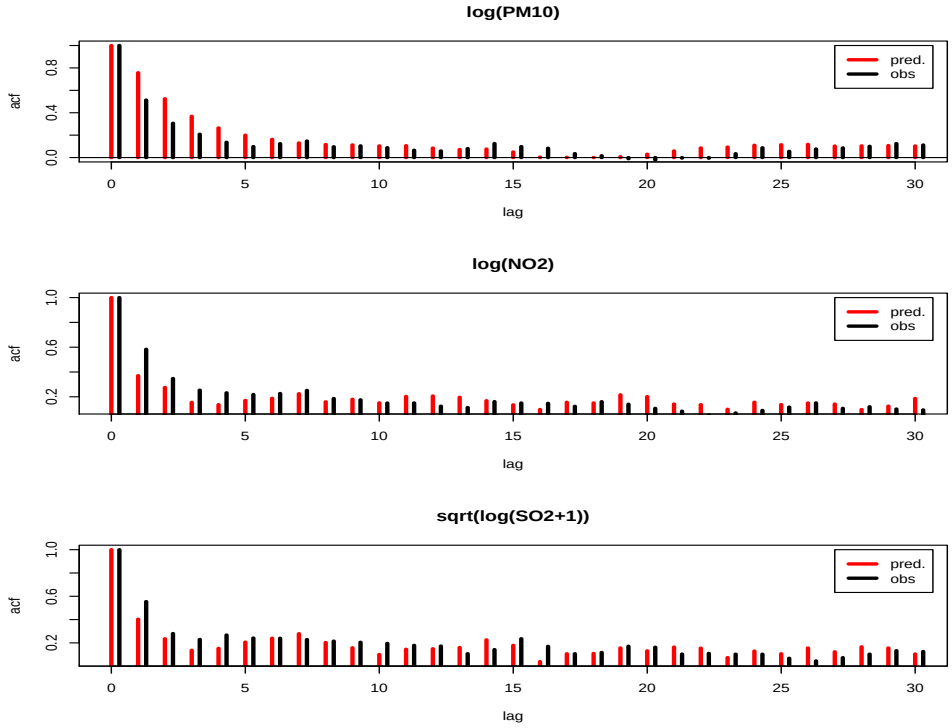

Figure 4: ACF's of normalized pollutants and of does predicted at the nearest grid point for the Archimede monitoring station (red bars)

can be found in the partial autocorrelation.

As far as the spatial prediction is concerned we report some examples of maps for each pollutant and the corresponding 95% credibility intervals (figures 6 - 11).

In the PM_{10} and SO_2 maps the influence of the wind direction appears clearly, while NO_2 flatter surface seems less influenced by this meteorological condition. This is in accordance whith what was previously noticed concerning the significance of spatial coordinates and the influence of the sea breeze. All maps show very little spatial variation, however they present highest pollutants levels where expected, according to experts opinion. Examining credibility intervals for the three pollutants shows that while estimates of PM_{10} and NO_2 concentrations are acceptable, this is not the case for the normalized SO_2. The credibility intervals are wider then those obtained for the other two pollutants and their lower limit is often negative, which is quite unrealistic. This maybe due to a lack of the model in capturing the large number of zeros recorded for SO_2.

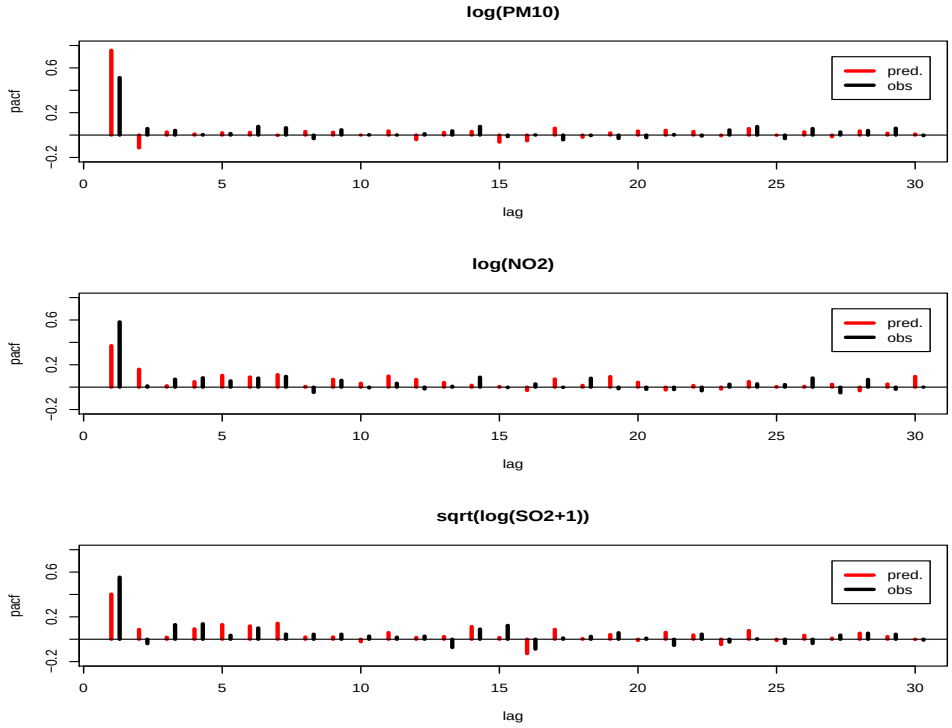

Figure 5: PACF's of normalized pollutants and of does predicted at the nearest grid point for the Archimede monitoring station (red bars)

5 Concluding remarks

In this work we analyze the behavior of a full Bayesian separable space-time hierarchical model adapted to predict normalized pollutant concentrations (PM_{10}, NO_2, SO_2) on a fine grid spanning the Taranto municipal area. The main advantages of the approach consist in its capability to easily handle missing data, properly reproduce the time dynamic and capture spatial information from the data. From the physical point of view the predicted maps have acceptable interpretation, however results are not completely satisfactory for SO_2. Furthermore, results obtained for the three univariate models are not really comparable with those in [8], where a multivariate approach is considered. As the multivariate hierarchical model requires a considerable computational effort to be estimated, we believe that further investigations are required giving more attention to the computational efficiency of the estimation procedure.

References

[1] Biggeri, A., Bellini, P., Terracini, B. *Metanalisi italiana degli studi sugli effetti a breve termine dell'inquinamento atmosferico*. Epidemiologia e Prevenzione, 28.

[2] Cocchi, D., Greco, F. , Trivisano, C. (2007) *Hierarchical space-time modelling of PM_{10} pollution*. Atmospheric Environment 41, pp. 532-542.

[3] Cressie, N. (1993) *Statistics for spatial data, revised edition*. John Wiley & Sons, New York.

[4] Gamerman, D. (1997) *Markov Chain Monte Carlo. Stochastic simulation for Bayesian inference*. Chapman e Hall.

[5] Geweke, J., Berger, J. O., Dawid, A. P., Smith A. F. M. (1992) *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments*. Bayesian Statistics 4.

[6] Gilks, W.R., Richardson, S., Spiegelhalter, D.J. (1996) *Markov Chain Monte Carlo in Practice*. Chapman e Hall, pp. 116-118.

[7] Le, N.D., Zidek, J.V. (2006) *Statistical Analysis of Environmental Space-Time Processes*. Springer.

[8] Pollice, A., Jona Lasinio, G. (2009) *A multivariate approach to the analysis of air quality in a high environmental risk area*. Working Paper Graspa n.32, www.graspa.org.

[9] 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.

[10] Shaddick, G., Wakefield, J. (2002) *Modelling daily multivariate pollutant data at multiple sites*. Appl. Statist. 51, part 3, pp. 351-372.

[11] Wikle, C. K., Berliner, L.M. , Cressie, N. (1998) *Hierarchical Bayesian space-time models*. Environmental and Ecological Statistics 5 , pp. 117-154.

Appendix: Priors settings

In this appendix we report the priors and their settings as implemented in WinBugs for the estimation of the three final models. In the mean term of the model, described in equation (2), γ_1, γ_2 are the coefficients of spatial coordinates and $\beta_{temp}, \beta_{relum}, \beta_{wind}, \beta_{rain}$ and $\beta_{winddir_i}, i = 1, \ldots, 4$ are the coefficients of temperature, relative umidity, wind speed, rain and wind direction respectively. For the SO_2 we added the calendar month whose coefficients are denoted by $\beta_{month_j}, j = 1, \ldots, 12$. The error term (one for each monitoring station) has precision parameter $\tau_s = \frac{1}{\sigma_s^2}$.
The temporal structure described in (6), has parameters ϕ_1 and $\tau_1 = \frac{1}{\sigma_\theta^2}$. The exponential spatial structure in (4) is ruled by two parameters: the range ϕ_{sp} and the sill $\sigma_{sp}^2 = \frac{1}{\tau_{sp}}$ of the covariance function.
For computational reasons we adopted lognormal distributions to generate precisions (τ_1, τ_{sp} and τ_s). This choice is purely computational and linked to the way WinBugs generates gamma and lognormal variates. The latter allow for a a more effective and simpler tuning of the initial values and, with the proper choice of hyperparameters, generates values that mimics gamma variates.
In the three models the ϕ_1 parameter has been generated from a normal distribution with very small precision centered on the maximum likelihood estimate of a AR(1) obtained by stacking the six monitoring stations recordings in a single series.

PM_{10} and NO_2

$$\gamma_1 \sim N(0, 0.0001) \qquad \gamma_2 \sim N(0, 0.0001) \tag{8}$$

$$\beta_{temp} \sim N(0, 0.0001) \qquad \beta_{relum} \sim N(0, 0.0001) \tag{9}$$

$$\beta_{wind} \sim N(0, 0.001) \qquad \beta_{rain} \sim N(0, 0.001) \tag{10}$$

$$\beta_{winddir_i} \sim N(0.0, 0.0001) \qquad i = 1, \ldots, 4 \tag{11}$$

$$\tau_s \sim LogN(0, 0.22); \quad s = 1, \ldots, 6 \tag{12}$$

The AR(1) hyperparameters for the NO_2 are set as:

$$\phi_1 \sim N(0.7068, 0.00001) \tag{13}$$

$$\tau_1 \sim LogN(0, 0.22) \tag{14}$$

for the PM_{10} AR(1) model component we have:

$$\phi_1 \sim N(0.7438, 0.01) \tag{15}$$

$$\tau_1 \sim LogN(0, 0.22) \tag{16}$$

$$\phi_{sp} \sim N(0, 0.0001)\mathcal{I}(0, 1) \tag{17}$$

$$\tau_{sp} \sim LogN(0, 0.22) \tag{18}$$

SO_2 As far as the SO_2 model is concerned we have the same hyperparameters structure as in the PM_{10} and NO_2 models with respect to all the shared covariates, the calendar month coefficients are set as:

$$\beta_{month_j} \sim N(0.0, 0.0001), \quad j = 1, \ldots, 12 \tag{19}$$

and AR(1) parameters are:

$$\phi_1 \sim N(0.63, 0.00001) \tag{20}$$

$$\tau_1 \sim LogN(0, 0.22) \tag{21}$$

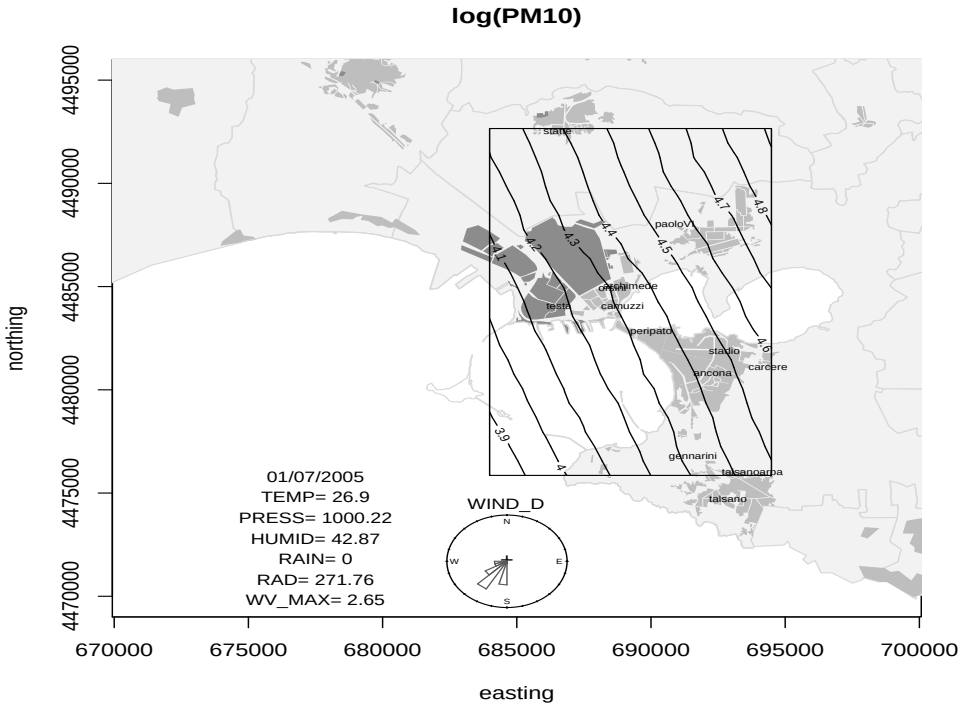

Figure 6: Log PM_{10} July 1, 2005 map

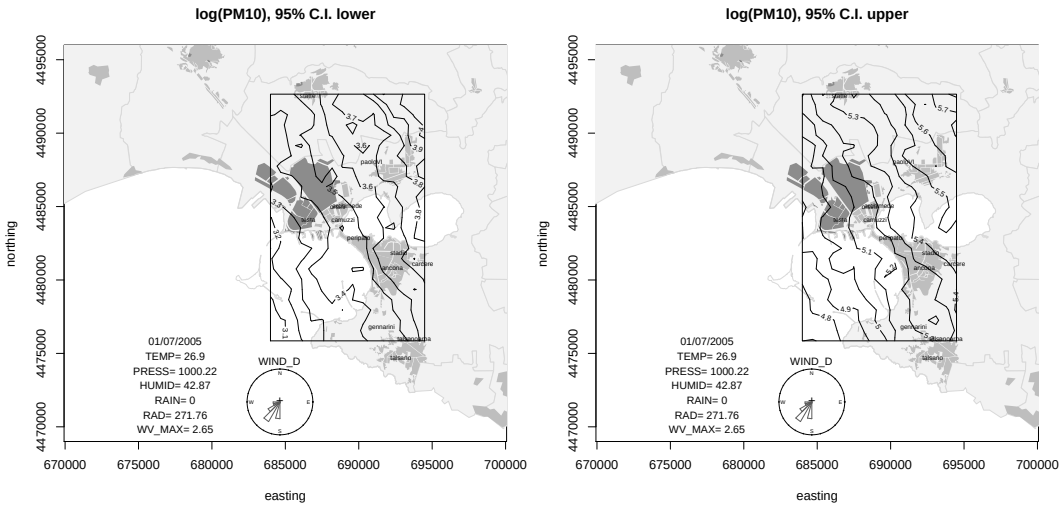

Figure 7: Log PM_{10} July 1, 2005, 95% Credibility Interval maps

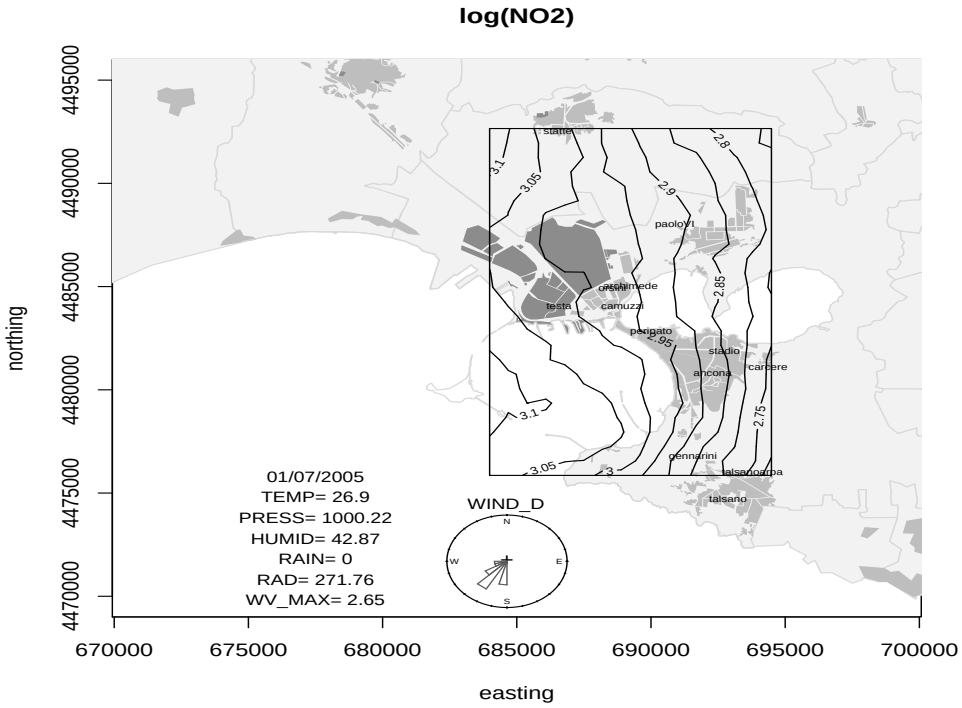

Figure 8: Log NO_2 July 1, 2005 map

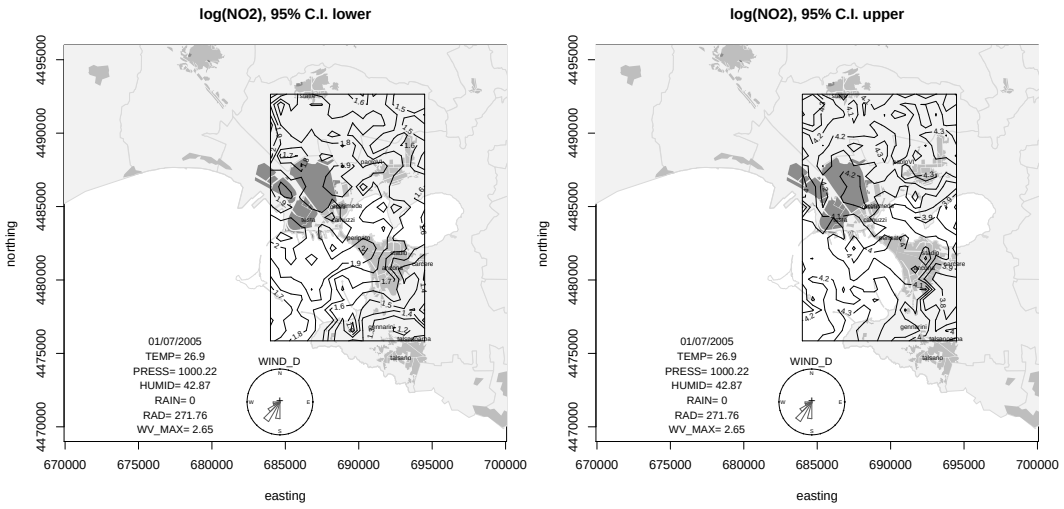

Figure 9: Log NO_2 July 1, 2005, 95% Credibility Interval maps

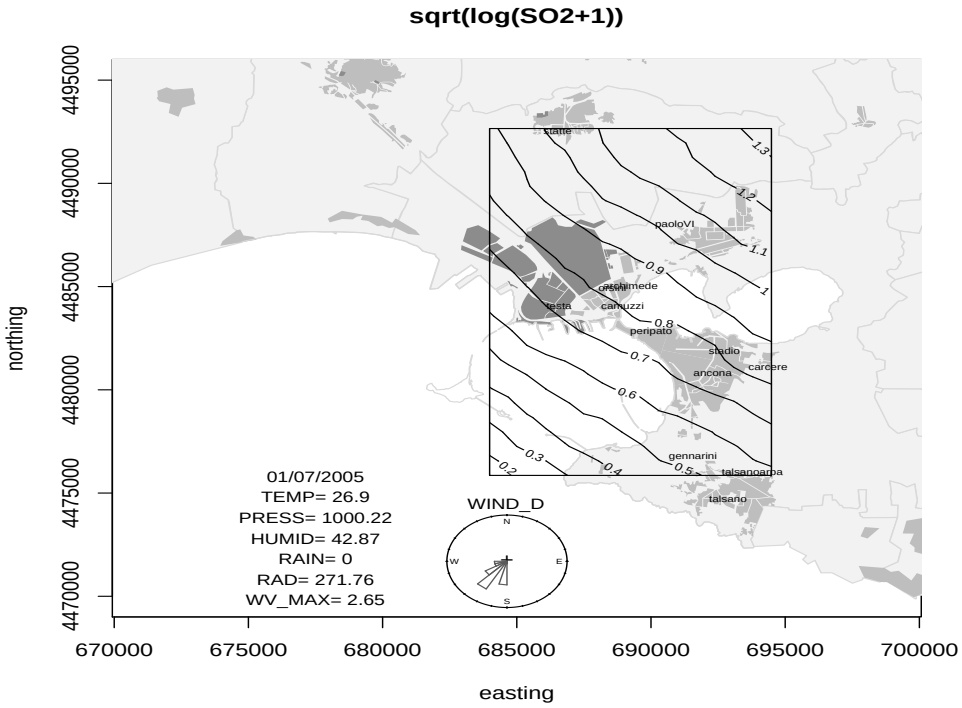

Figure 10: Transformed SO_2 concentration December 16, 2007 map

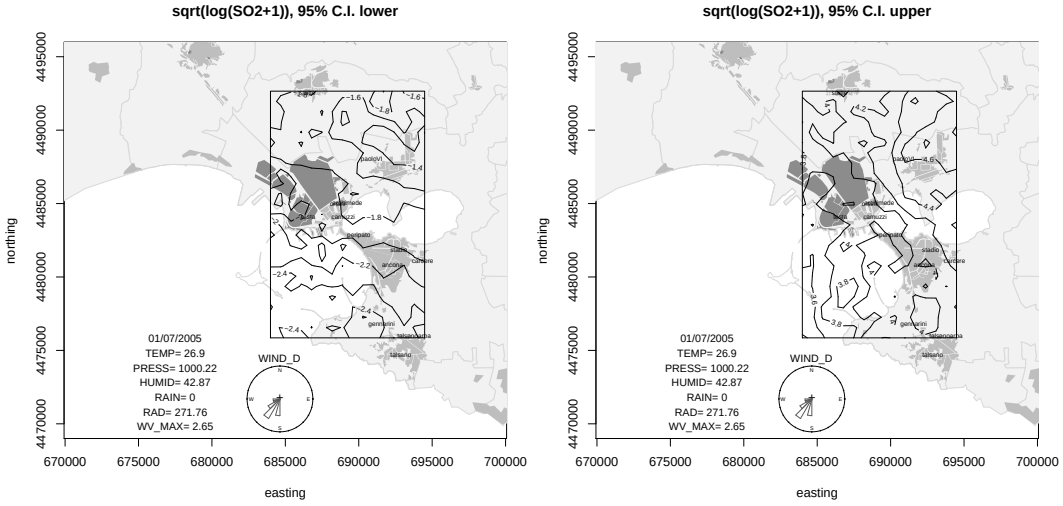

Figure 11: Transformed SO_2 July 1, 2005, 95% Credibility Interval maps