International Workshop on Spatio-Temporal Modelling
Santiago de Compostela
30 Jun - 2 Jul 2010

Spatio-temporal covariance modelling of CO concentrations in Madrid

G. Fernández-Avilés[1,*], J.M. Montero[1] and J. Mateu[2]

[1] *Departamento de Estadística, Facultad de Ciencias Jurídicas y Sociales de Toledo, Universidad de Castilla La Mancha, 45071, Toledo, Spain; Gema.FAviles@uclm.es, Jose.MLorenzo@uclm.es*
[2] *Departamento de Matemáticas, Universitat Jaume I, Campus Riu Sec, E-12071 Castellón, Spain; mateu@mat.uji.es*
[*] *Corresponding author*

Abstract. *We present a space-time analysis of CO levels in Madrid (Spain). Several covariances ranging from separable to nonseparable structures are fitted using weighted composite likelihood methods. Prediction and goodness-of-fit are evaluated through RMSE.*

Keywords. *Environmental pollutant; Spatio-temporal covariances; Weighted composite likelihood.*

1 Introduction

Madrid is the third-most populous municipality in the European Union, and it is suffering a rapid suburbanization process where population and jobs are moving out of the central city. This process produces an imbalanced mobility pattern and more car dependency. In the last decade the number of vehicles in Madrid has increased by 5.6%, and in 2009 the number of unities was 1917382. This implies 1202.5 vehicles per km. and 683.5 vehicles per 1000 inhabitants (municipal register of Madrid). One million drivers enter and leave daily the city. So, car pressure is increasing as well as its negative environmental impacts. Nevertheless, air pollution in Madrid may be also attributed to other factors as manufacturing and heating systems during winter, among others. Currently, in Madrid there are working more than 1200 coal boilers. Due to the health effects of air pollution, and that in recent years the major threat to clean air inside the cities is posed by traffic emissions, carbon monoxide (CO) continues to be one of the main worries of Madrid citizens and authorities. CO is a toxic gas formed as a product of incomplete combustion in the burning of fossil fuels, and can cause harmful health effects by reducing oxygen delivery to the body's organs (like the heart and brain) and tissues. The main sources of CO in most parts of Madrid are motor vehicle exhaust emissions (91.4% in 2009). In Madrid City, as in other big cities around the world, there exists a monitoring net that measures hourly CO and other important pollutants. Obviously, monitoring stations are located according to the purpose they pursuit, but it is also true that in many occasions municipal regulations constitutes a serious restriction to their optimal location. This is the reason why estimation of CO levels in sites of significant traffic congestion, and particularly at busy intersections on inner-city streets are needed. Going beyond the traditional spatial strategies, here CO levels are modelled with spatio-temporal techniques, as not only space but time, and the interaction space-time, are core factors in air pollution prediction tasks. Additionally, RMSE provided by both spatial and spatio-temporal strategies are also compared.

2 Methods

It is clear that including time in classic interpolation strategies has notable advantages. Time is a crucial factor in pollution prediction tasks, as it is its interaction with space. This is the reason why space-time predictions of pollution are considered optimal from a geostatistical perspective. In the spatio-temporal case a previous requirement to the spatial analysis is the modelling of the temporal dynamics, by taking into account the strong seasonality existing in environmental data. This is done by specifying a time series equation following the additive model $Y_t = S_t + YSA_t$, where Y_t represents the original time series data, S_t is the unobserved seasonal component, and YSA_t is the seasonal adjusted time series. Usually, the seasonal adjusted time series can also be decomposed as $YSA_t = Trend_t + I_t$ where the unobserved $Trend_t$ component shows the long-run evolution of the series (associated with the motions in the low frequencies) and I_t is the irregular component.

When the seasonal pattern is not stable but it changes over time, we need to specify a stochastic model to estimate the seasonal component S_t. A possibility widely used for environmental data is the use of the STL procedure which applies weighted regressions based on neighbourhood observations. This is a non-parametric regression technique which allows smooth changes in the seasonal pattern. On the other hand, this procedure can also be applied to model the dynamic dependencies of the seasonal adjusted component as an alternative to ARMA models. Nevertheless, results tend to be quite similar in both cases. As spatial and spatio-temporal covariance functions, we used: (i) *The exponential spatial model*: $C(\mathbf{h}) = exp(-c||\mathbf{h})$, with $c > 0$; (ii) *The separable exponential spatio-temporal model*: $C(\mathbf{h},u,\theta) = exp(-c||\mathbf{h}|| - a|u|)$, with $a,c > 0$; (iii) *The Gneiting separable spatio-temporal model*: $C(\mathbf{h},u,\theta) = \frac{1}{(a|u|^{2\alpha}+1)} exp\left(-c||\mathbf{h}||^2\gamma\right)$; (iv) *The Gneiting non-separable spatio-temporal model*: $C(\mathbf{h},u,\theta) = \frac{1}{(a|u|^{2\alpha}+1)} exp\left(\frac{-c||\mathbf{h}||^2\gamma}{(a|u|^{2\alpha}+1)^{\beta\gamma}}\right)$.

3 Data analysis

The data used in this paper have been provided by the Atmosphere Pollution Monitoring System of Madrid municipality. They represent hourly measures at 23 fixed operative monitoring stations during 2008. Figure 1 shows the locations of the air quality monitoring stations. Most of the monitoring stations are located in the urban center and relatively few in the peripheral sites. Note the reasonable coverage of the domain under study by the monitoring stations since most of Madrid population is concentred in the urban center. Data have been daily averaged for each hour, and taking daily averages as inputs the average of CO for the labour days of every week in 2008 was computed. Therefore, the final data consisted of 52 time measures at 23 spatial locations.

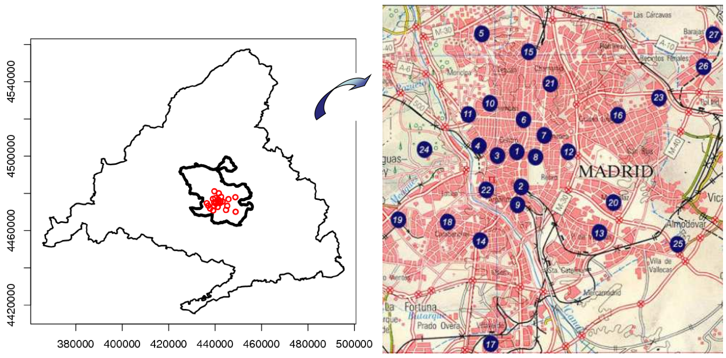

Figure 1: *Left:* Location of Madrid Municipality; *Right:* Locations of the monitoring stations

Due to the large number of observations (28704), maximum likelihood procedure became infeasible. This is the reason why weighted composite likelihood (WLS) ([1, 2, 3]) was used for the estimation of the covariance parameters. From a preliminary study, we identified the optimum distance $\mathbf{d} = (6617, 20)'$, where the spatial distance is expressed in meters and the temporal distance in hours. Prediction was calculated for each station eleven months ahead by using the simple kriging predictor. As a goodness-of-fit measure, we used the root mean square error (RMSE) at each site, defined as $RMSE(\mathbf{s}) = (\sum_{t=1145}^{1248}(z(\mathbf{s},t) - \widehat{z}(\mathbf{s},t))^2/104)^{1/2}$.

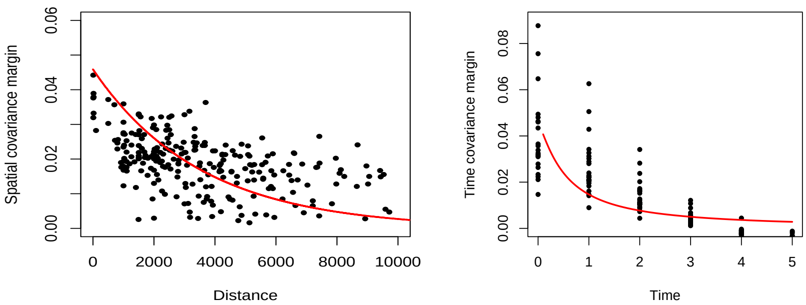

Figure 2: *Left:* Empirical and fitted marginal spatial covariance for the nonseparable Gneiting model; *Right:* Empirical and fitted marginal temporal covariance for the nonseparable Gneiting model

Table 1 shows that in any monitoring stations the spatio-temporal strategies overperform the spatial one. In general, the Gneiting nonseparable covariance structure fits best and provides reliable predictions. In particular, the proportional reduction in the RMSE obtained by using the nonseparable strategy is 43.9921% with respect to the spatial procedure, 8.3288% when comparing with the separable exponential space-time function, and 4.0404% when the comparison refers to the Gneiting separable case. We finally note that a comparison of space-time strategies with functional spatial techniques would be fair and would give a complete picture of the modelling possibilities.

Monitoring Station	**(a)**	**(b)**	**(c)**	**(d)**
MS 1: P. Recoletos	0.3854582	0.2139786	0.2044951	0.1955572
MS 3: Pl. del Carmen	0.3240636	0.1371976	0.1349203	0.1149335
MS 4: Pl. de España	0.3934171	0.1618391	0.1708875	0.1537030
MS 5: Barrio del Pilar	0.3311806	0.2360965	0.2191661	0.2278387
MS 6: Pl. Dr. Marañón	0.2816351	0.2015402	0.1897874	0.1797578
MS 7: Pl. M. Salamanca	0.1941061	0.1747552	0.1652564	0.1632985
MS 8: Escuelas Aguirre	0.2124621	0.1708966	0.1436747	0.1411525
MS 9: Pl. Luca de Tena	0.2094385	0.1804746	0.1621911	0.1659411
MS 10: Cuatro Caminos	0.4090921	0.2205546	0.2125661	0.2019521
MS 11: Av. Ramón y Cajal	0.3015753	0.2338579	0.2440335	0.2140196
MS 12: Pl. M. Becerra	0.3101455	0.1304668	0.1150154	0.1033097
MS 13: Vallecas	0.2708333	0.1487223	0.1908054	0.1380014
MS 14: Pl. Fdez. Ladrada	0.2727415	0.1833370	0.1814690	0.1720224
MS 15: Pl. de Castilla	0.3583467	0.2312207	0.1948311	0.2169114
MS 16: Arturo Soria	0.2380602	0.1965665	0.2013957	0.1824901
MS 18: General Ricardos	0.2784963	0.2208941	0.1841069	0.2021593
MS 19: Alto Extremadura	0.2196051	0.1433856	0.1321417	0.1381165
MS 20: Av. Moratalaz	0.2253499	0.1930202	0.1748129	0.1740558
MS 21: Isaac Peral	0.4121843	0.1317656	0.1541003	0.1287506
MS 22: Paseo Pontones	0.2713991	0.2082368	0.2062355	0.1975824
MS 23: Alcalá	0.4036888	0.2562729	0.2140545	0.2292948
MS 24: Casa Campo	0.4632078	0.1245605	0.1254493	0.1131509
MS 25: Santa Eugenia	0.1906614	0.1509338	0.1392225	0.1425540

Table 1: RMSE (log CO), (a) spatial exponential covariance, (b) spatio-temporal exponential covariance, (c) Gneiting [4, 5] separable spatio-temporal covariance, and (d) Gneiting [4, 5] nonseparable spatio-temporal covariance

References

[1] Bevilacqua, M., Gaetan, C., Mateu, J. and Porcu, E. (2007). Estimating space and space-time covariance functions: a weighted composite likelihood approach. *Technical Report*, 117-2007, Universitat Jaume I. Submitted.

[2] Bevilacqua, M. (2008). *Estimating space and space-time covariance functions: a weighted composite likelihood approach.* PhD. thesis.

[3] Bevilacqua, M., Mateu, J., Porcu, E., Zhang, H. and Zini, A. (2009). Weighted composite likelihood-based tests for space-time separability of covariance functions. *Statistics and Computing*. Technical Report, 131-2008, Universitat Jaume I.

[4] Gneiting, T. (2002). Stationary covariance functions for space-time data. *Journal of the American Statistical Association* **97**, 590-600.

[5] Gneiting, T., Genton, M.G. and Guttorp, P. (2005). Geostatistical space-time models, stationarity, separability, and full symmetry. *Technical Report* n. **475**, University of Washington.