Tapering spatio temporal models [1]

Alessandro Fassó, Francesco Finazzi, Moreno Bevilacqua
DIIMM, University of Bergamo, alessandro.fasso@unibg.it

Abstract: We consider regression models for multivariate spatio temporal data. We view the data as a time series of spatial processes and work in the setting of dynamic models. In order to add flexibility we consider regression models with spatio temporal varying coefficients. Spatial dependence among the different measurements is attained considering the linear model of coregionalization. Since spatio-temporal data are tipically of large dimension we propose to perform estimation both through maximum likelihood by means of the EM algorithm and a modified version of it exploiting the covariance tapering likelihood function.

Keywords: Air quality assessment, Covariance tapering, EM algorithm,r maximum likelihood estimation.

1 Introduction

The increasing availability of datasets on multivariate spatio-temporal data parallels the need for statistical models which are flexible enough for covering the underlying complexity and can be estimated by means of well founded inferential techniques. The dynamic coregionalization model, recently proposed by Fassó and Finazzi (2011a), has these advantages as it allows modelling of complex multivariate spatio-temporal dynamics and performing maximum likelihood parameter estimation by means of the EM algorithm.

Due to the advancement of technology, massive amounts of data are often observed at a large number of spatial locations in environmental sciences. For this reason recent literature focused on geostatistical analysis of large multivariate spatio-temporal datasets. See for instance Bevilacqua *et al.* (2011) and Cressie and Johannesson (2008). This is because spatial problems with modern data often overwhelm traditional implementations of spatial statistics, such as maximum likelihood estimation. In this paper, in order to estimate multivariate regression spatio temporal models for EU air quality assessment, we consider an approximation of the estimation method proposed by Fassó and Finazzi (2011a) by considering the covariance tapering approach. The key idea is that the use of covariance tapering allows to manage large multivariate spatio temporal data.

[1]This research is part of Project EN17, Methods for the integration of different renewable energy sources and impact monitoring with satellite data", funded by Lombardy Region under Frame Agreement 2009.

2 Dynamic coregionalization model with varying coefficients

We consider the following observation equation for the multivariate spatio-temporal random process $Y(s,t) = (Y_1(s,t), \ldots Y_q(s,t))'$ at time $t = 1, \ldots, T$ and site $s \in D \subset R^2$:

$$Y(s,t) = X_1(s,t)\beta + X_2(s,t) K_2 Z(t) + X_3(s,t) K_3 W(s,t) + \varepsilon(s,t), \qquad (1)$$

where X_1, X_2, X_3 are matrices of known covariates and K_2, K_3 are matrices of constants . $Z(t)$ is a $p-$variate Markovian component, $W(s,t)$ is a $r-$variate Gaussian random field and $\varepsilon(s,t)$ is a $q-$variate Gaussian white noise in space and time.

The $p-$dimensional latent temporal state $Z(t)$ has the Markovian dynamics $Z(t) = GZ(t-1) + \eta(t)$, with G a stable transition matrix and $\eta \sim N(0, \Sigma_\eta)$. The $k-$dimensional Gaussian random field is described by coregionalization model of c components

$$W(s,t) = \sum_{j=1}^{c} W^{[j]}(s,t)$$

where each $W^{[j]}(s,t)$, for fixed t, is a latent zero-mean Gaussian process with covariance and cross-covariance matrix function $\Gamma^{[j]} = cov\left(W_i^{[j]}(s,t), W_{i'}^{[j]}(s',t)\right) = V_j \rho^{[j]}\left(h, \bar{\theta}^{[j]}\right)$, $1 \le i, i' \le r$, $1 \le j \le c$. Each V_j is a coefficents matrix and each $\rho^{[j]}$ is a valid correlation function and $h = \|s - s'\|$ is the Euclidean distance between s and s'. All spatial processes above are purely spatial processes in the sense that are uncorrelated over different time points. Finally, $\varepsilon_i(s,t) \sim N(0, \sigma_{\varepsilon,i})$, $i = 1, \ldots q$ is the measurement error which is white-noise in space and time. The parameter set to be estimated is $\Psi = (\beta, \sigma_\varepsilon; G, \Sigma_\eta; \theta; V) = (\Psi_Y, \Psi_Z, \Psi_W)$ where $\beta = (\beta_1, ..., \beta_q)'$, $\sigma_\varepsilon = (\sigma_{\varepsilon,1}, ..., \sigma_{\varepsilon,q})'$, $\theta = (\theta_1, ..., \theta_c)'$ and $V = (V_1, ..., V_c)'$.

3 Estimation method

At each time t, each $Y_i(s,t)$ is observed at n_i sites $S_i = (s_{i,1}, ..., s_{i,n_i})$. The sets in $S = (S_i, ..., S_q)$ are not constrained and can be disjoint. The observed vector at time t is then $Y(S,t) = (y_1(S_1,t), ..., y_q(S_q,t))'$ a vector of dimension $N = \sum_{i=1}^{q} n_i$.

Due to the Markovian assumption and to the space-time separability property of the model, and setting $Y = (Y(S,1), \ldots, Y(S,T))'$, $Z = (Z_0, Z_1, ..., Z_T)'$, with $Z_t = Z(t)$ and $W^{[j]} = \left(W_1^{[j]}, ..., W_T^{[j]}\right)'$, $j = 1, .., c$, and $W = \left(W^{(1)}, ..., W^{(c)}\right)'$, the complete-data log-likelihood function $L(\Psi; Y, Z, W)$ takes the nice additive form:

$$l(\Psi; Y, Z, W) = l(\Psi_Y; Y \mid Z, W) + l(\Psi_Z; Z) + \sum_{j=1}^{c} \sum_{t=1}^{T} l\left(\Psi_W; W_t^{[j]}\right)$$

where

$$l(\Psi_Z; Z) = l(\Psi_Z; Z_0) + \sum_{t=1}^{T} l(\Psi_Z; Z_t \mid Z_{t-1})$$

The involved distributions are all of Gaussian type. Specifically for the latent variable

$$\begin{aligned} Z_0 &\sim N_p(\mu_0, \Sigma_0) \\ (Z_t \mid Z_{t-1}) &\sim N_p(GZ_{t-1}, \Sigma_\eta) \\ W_t^{[j]} &\sim N_N\left(0, \Sigma^{[j]}\right), \ 1 \leq j \leq c \end{aligned}$$

Note that $\Sigma^{[j]} = \Sigma^{[j]}(\Psi_W)$ is a matrix of dimension $R \times R$. with $R = \sum_{i=1}^{r} n_i$

Estimation can be performed adapting the EM- algorithm as proposed in Fassó and Finazzi (2011a) and modified by Fassó and Finazzi (2011b).

Here we propose a modification of this alghorithm to take into account the problem of large dataset. Specifically the modification is based on the covariance tapering likelihood idea Kaufmann *et al.* (2008) that is we consider the tapered complete data likelihood:

$$l_{TAP}(\Psi; Y, Z, W) = l(\Psi_Y; Y \mid Z, W) + l(\Psi_Z; Z) + \sum_{j=1}^{c}\sum_{t=1}^{T} l_{TAP}\left(\Psi_W; W_t^{[j]}\right)$$

where $l_{TAP}\left(\Psi_W; W_t^{[j]}\right)$ is defined as:

$$l_{\mathrm{TAP}}(\Psi_W; W_t^{[j]}) = -\frac{1}{2}\log|\Sigma^{[j]} \circ T(d)| - \frac{1}{2} W_t^{[j]\prime}([\Sigma^{[j]} \circ T(d)]^{-1} \circ T(d)) W_t^{[j]} \qquad (2)$$

This is the multivariate version of the tapering likelihood proposed by Kaufman *et al.* (2008). In their approach for the univariate case, certain elements of the covariance matrix are set to zero multiplying it element by element by a correlation matrix coming from a compactly supported isotropic correlation function.

Here $T(d)$ is a sparse cross-correlation matrix coming from a valid model of matrix valued correlation function with compact support and $\circ$ is the Schur product. A simple model for the isotopic case is the following: let $\rho(h, d)$ a compact support correlation function (one of the Wendland (1995) class for instance) and let B a $r \times r$ positive definite matrix of coefficients, then $\rho(h, d)B$ is a $r \times r$ valid model of matrix valued correlation function with compact support. The associated matrix is $T(d) = B \otimes H(d)$ where $H(d) = \{\rho(||s_i - s_j||, d)\}_{i,j=1}^{n}$. The 'tapered' matrix $\Sigma^{[j]} \circ T(d)$ is still positive definite and sparse matrix algorithms can be used to evaluate an approximated likelihood efficiently. The intuition behind this approach is that correlations between pairs of distant sampling locations are often nearly zero, so little information is lost in taking them to be independent.

References

Bevilacqua, M., Gaetan, C., Porcu, E.,Mateu, J.. (2011), Estimating space and space-time covariance functions for large data sets: a weighted composite likelihood approach. *Journal of the American Statistical Association*, Accepted for publication.

Cressie, N., and Johannesson, G. (2008), Fixed rank kriging for very large spatial data sets, *Journal of the Royal Statistical Society*, Series B, 70, 209-226..

Fassó A., Finazzi F., (2011a) Maximum likelihood estimation of the dynamic coregionalization model with heterotopic data. *Environmetrics* Accepted for publication.

Fassó A., Finazzi F. (2011b) Spatio temporal models with varying coefficients for European air quality assessment. Proceedings of Spatial 2, Spatial Data Methods for Environmental and Ecological Processes 2nd Edition, Foggia, Sept. 1-2, 2011.

Kaufman, C., Schervish, M., and Nychka, D. (2008), Covariance tapering for likelihood- based estimation in large spatial datasets, *Journal of the American Statistical Association*, 103, 1556-1569.

Wendland, H. (1995), Piecewise Polynomial, Positive Definite and Compactly Supported Radial Functions of Minimal Degree, *Advances in Computational Mathematics*, 4, 389-396.