

A Bayesian estimation of a separable spatio-temporal model with autoregressive temporal component

Stima bayesiana di un modello spazio-temporale separabile con componente temporale autoregressiva

Silvia Bozza

Dipartimento di Statistica
Università degli Studi di Venezia
silvia@vivaldi.dst.unive.it

Riassunto: In questo articolo, partendo da un processo stocastico spazio-temporale, si propone un modello bayesiano per l'analisi della struttura di dipendenza spazio-temporale. L'approccio seguito produce una stima congiunta non parametrica della covarianza spaziale e temporale, dove in particolare la struttura di dipendenza temporale è modellata attraverso un approccio gerarchico.

Keywords: Separability; Space-Time models; Bayesian hierarchical models; MCMC.

1. Introduction

Space-time models have gained an increasing popularity in the last decade which can be explained from one side with the numerous applications related with the environmental and health sciences, from the other side with the increasing computational powers nowadays available. Most of the literature on space-time modelling is motivated by specific applications and each model appears to be unique and have little in common with others, but in general we can recognize two distinct ways of approaching a space-time problem: the prioritizing and the conjoint method. The prioritizing method models the spatial and temporal components separately: it begins by modelling one dimension and then making the parameters of that model depend on the coordinates of the other dimension (Waller *et al.* (1997)). Prioritizing models are typically hierarchical. The conjoint method models the spatial and temporal components jointly, but modelling a massive joint covariance matrix can become very complicated. Separability or additivity of the spatial and temporal dimensions, when admissible, simplify greatly the problem (Gneiting and Schlather (2001)). The hypothesis of separability in particular allows the spatial and temporal dimensions to be written as product of a purely spatial component and a purely temporal component, for which there are standard and flexible classes of models. Separability is a strong assumption since it underlines that all the locations in space have got the same temporal covariance structure, but we believe that separability allows a non-parametric estimation of the spatio-temporal dependence structure. Many applications in fact are based on classes of stationary spatio-temporal covariance functions, which have the great advantage that allow prediction in correspondence of every location in the space and every instant of time. Nevertheless, while there are heuristic arguments underlying assumptions of stationarity in time series, there is little reason to expect spatial covariance to be stationary over the spatial scales of interest. A Bayesian hierarchical model is proposed on section 2, while a specific application is presented on section 3.

2. A Bayesian semi-parametric hierarchical model

Let's denote with $Y(s, t)$ the observable matrix, where $s \in \mathbb{R}^2$ denotes the locations in space and $t \in \mathbb{R}$ the instant of time. If we stuck the columns of Y into a vector, then we can write $y = (y'_1, \dots, y'_t)'$. Assume that $E(y) = 0$ and $cov(y, y') = \Omega \otimes \Delta$, where $\Omega(t, t)$ is a purely temporal covariance component and $\Delta(s, s)$ is a purely spatial covariance component. Then, under hypothesis of normality, the likelihood will be:

$$f(\Omega, \Delta | Y) \propto \frac{1}{|\Omega|^{\frac{s}{2}} |\Delta|^{\frac{t}{2}}} \exp \left\{ -\frac{1}{2} tr [\Omega^{-1} Y' \Delta^{-1} Y] \right\}. \quad (1)$$

The prior densities for the unknown spatial and temporal covariance matrices are chosen from the space of all positive definite symmetric matrices, that is they're of type Wishart inverse. The Bayesian choice adopted guarantees the existence of proper posteriors, but with the drawback that they depend too much on the prior assumptions, and this can be explained by the great number of parameters to be estimated compared to the available observations. Now, supposing to have a fixed number of locations in the space in correspondence of them there are collected observations at regular intervals of time, instead of allowing the prior for the temporal covariance matrix being anything, we might choose an inverted Wishart distribution which in mean reflect the structure of a given model for time series. The scale matrix of the distribution will be then completely specified by the parameters of time series model. We therefore introduce a Bayesian hierarchical model which is a Bayesian statistical model where the prior distribution is decomposed in conditional distributions. The prior modelling can therefore be decomposed in two parts: the first one corresponding to the structural information about the model, the second one to the more subjective information (Robert (1994)). In the case of interest, at the first level there are chosen priors for the unknown temporal and spatial covariance matrices. The scale of the inverted Wishart for Ω is defined conditionally on the parameters which define the autocovariance function of a given time series model. At the second level there are chosen weak priors for the unknown parameters of the time series model.

3. An application with an autoregressive process of order 1

We have simulated a data set with $i = 12$ locations located over a regular grid on the space, in correspondence of them there are collected data for a period of $j = 100$. The spatial dependence is allowed to be anything, though it is thought to decrease with distance, meanwhile the temporal dependence is thought to be coherent in mean with the autocovariance structure of an autoregressive process of order 1 (Hamilton (1994)). At the first level of the hierarchy we'll have:

$$p(\Delta) = IW_s(R, n_\delta) \quad (2)$$

$$p(\Omega) = IW_t(T(\phi, \sigma^2), n_\omega) \quad (3)$$

where R and T are the scale matrices, ϕ is the autoregressive parameter, σ^2 the innovation variance, n_δ and n_ω are the degrees of freedom of the inverted Wishart which are chosen high in order to reduce the variability of the Wishart ($n_\delta = 36$, $n_\omega = 300$).

The second level parameters are assumed to be independent. For the autoregressive parameter it is chosen a Beta distribution, with parameters a and b fixed in order to have a

weak prior ($a = 1.5, b = 1.5$). The respect of the stationary region, $|\phi| < 1$, is guaranteed by the transformation: $\phi = 2\eta - 1, \eta \sim Be(a, b)$. It is straightforward proving that $p(\eta) \propto p(\phi)$. The prior for the innovation variance must naturally have positive domain. We choose a Gamma distribution with fixed parameters α and β ($\alpha = 3.5, \beta = 3.5$). The conjoint posterior looks quite intractable:

$$p(\Delta, \Omega, \phi, \sigma^2 | Y) = f(Y | \Delta, \Omega) p(\Delta, \Omega | \phi, \sigma^2) p(\phi) p(\sigma^2) \quad (4)$$

$$\propto \frac{1}{|\Delta|^{\frac{n\delta+t}{2}} |\Omega|^{\frac{n\omega+s}{2}}} \exp \left[-\frac{1}{2} \text{tr} \Omega^{-1} Y' \Delta^{-1} Y \right] \sigma^{2(\alpha-1)} e^{-\beta\sigma^2}$$

$$\exp \left[-\frac{1}{2} \text{tr} \Delta^{-1} R - \frac{1}{2} \text{tr} \Omega^{-1} T(\phi, \sigma^2) \right] \left(\frac{\phi+1}{2} \right)^{a-1} \left(\frac{1-\phi}{2} \right)^{b-1}.$$

We need to implement techniques of type Markov chain Monte Carlo to determine the relevant posterior distributions. Since we have chosen Gaussian distributions with conjugate priors, the derivation and implementation of the full conditional distributions needed for the Gibbs sampling are, with one exception, quite straightforward. The full conditional distributions for the first level parameters of the hierarchy are of type inverted Wishart so we do not have any difficulties in sampling from them (Ripley (1987)). The full conditional of the innovation variance is still a Gamma, but we have decided to constrain this parameter to be equal to 1, to solve a problem of non identifiability that follows from the Kronecker product structure of the covariance, we have in fact: $\Omega \otimes \Delta = c\Omega \otimes c^{-1}\Delta$. Being the innovation variance a second level parameter, this is not an excessive arbitrary choice. The exception concerns the full conditional of the autoregressive parameter which is rather complicated with the consequence that we are not able to sample directly from it. In order to sample a value for ϕ we add in the Gibbs sampling a random walk Metropolis Hastings step (Gilks *et al.* (1996)) with a normal distribution as candidate generating density.

We have generated a chain 20000 iterations long, with a burn-in of 10000 iterations. Results of the Gibbs-sampler algorithm are given on table 1. For the temporal and spatial covariance matrix we have decided to represent the distance between the theoretical value and sampled one. The trace-plot for the spatial covariance matrix, Fig. 1, and the temporal covariance matrix, Fig. 2, and the relative autocorrelations, Fig. 3-4, suggest we have obtained convergence.

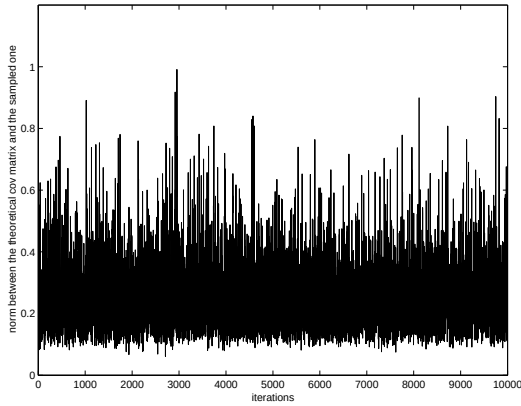


Fig. 1: Trace-plot delta

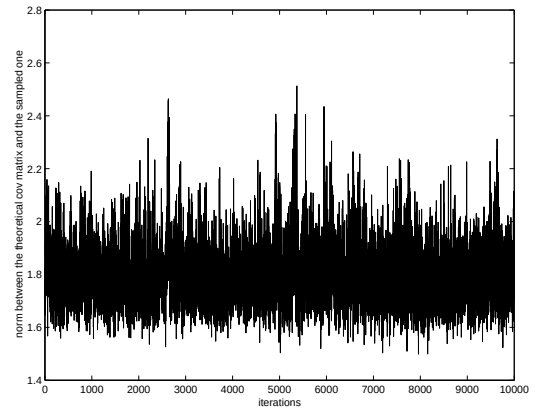


Fig. 2: Trace-plot omega

Model parameter	Th. value	Prior mean	MCMC post mean	Standard error
ϕ	0.7	0	0.6912	1.5332e-006
$\ \Delta_{th} - E(\Delta)\ /\ \Delta_{th}\ $	-	0.1010	0.0503	2.0011e-005
$\ \Omega_{th} - E(\Omega)\ /\ \Omega_{th}\ $	-	0.9094	0.2591	1.0839e-005

Table 1: Gibbs-sampler results. $t = 100$

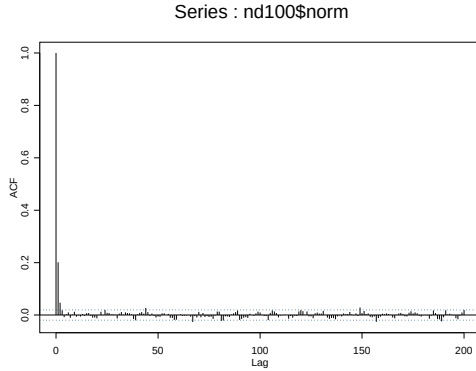


Fig. 3: Acf delta

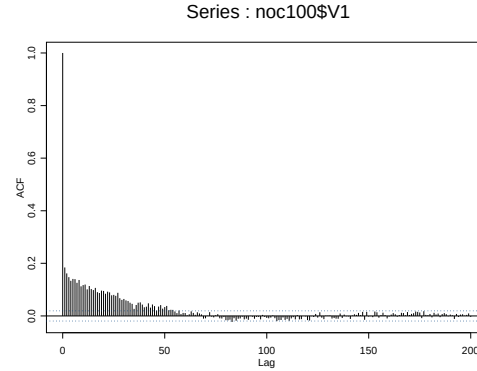


Fig. 4: Acf omega

The model proposed can obviously be applied for processes of higher order. It is clear that as the observations in time grow, the number of parameters to estimate increase, but we point out that the inverted Wishart introduced has got a mean which is specified by a little number of parameters, and a variance kept little with an opportune choice of the degrees of freedom. The estimations obtained have shown to be consistent. In practice, sampling from an inverted Wishart of considerable high dimensions will be computationally onerous. However, by utilizing efficient algorithms for the inversion of matrices, significant speed improvements can be realized.

References

- Gilks W., Richardson S. and Spiegelhalter D. (1996) *Markov chain Monte Carlo in practice*, Chapman & Hall.
- Gneiting T. and Schlather M. (2001) *Space-time covariance models*, New York, Wiley, 1–15, Encyclopedia of Environmetrics, to appear.
- Hamilton J. (1994) *Time series analysis*, Princeton University Press, New Jersey.
- Ripley B. (1987) *Stochastic simulation*, Wiley, N.Y.
- Robert C. (1994) *The Bayesian Choice. A Decision-Theoretic Motivation*, Springer Verlag.
- Waller L., Carlin B., Hong X. and Gelfand A. (1997) Hierarchical spatio-temporal mapping of disease rates, *J. Amer. Statist. Soc.*, 92, 438, 607–617.