Dynamic spatial models, including spatial time series [1]

Dani Gamerman
Instituto de Matemática - UFRJ
dani@im.ufrj.br

1 Dynamic linear models

This book has already presented many situations where the observation process under study contains the temporal dimension as well as the spatial dimension. This section is devoted to detail a popular and fairly general framework for handling these situations. It is based on firmly established models, called dynamic or state-space models, with a non-parametric flavor. They have proved a flexible tool to handle temporal correlation (West and Harrison, 1997) in a variety of different contexts. This section will also provide a number of contexts where these models can be applied in the context of spatial analysis.

Dynamic models are described via a p-dimensional latent process $\beta(\cdot)$ defined over time according to a temporal difference equation

$$\beta(t') = G(t',t)\beta(t) + w(t',t), \quad \text{with } w(t',t) \sim N(0, W(t',t)), \quad (t' > t) \quad (1)$$

where the transition matrix G conveys the deterministic part of the evolution and the system disturbance w is simply a stochastic component accounting for increased uncertainty (controlled by the disturbance variance W) over the temporal evolution. The model is completed with an initial specification for β at say $t = 0$. For temporally equidistant points, equation (1) can be simplified to

$$\beta_t = G_t\beta_{t-1} + w_t, \quad \text{with } w_t \sim N(0, W_t). \quad (2)$$

This will be assumed to be the case hereafter without loss of generality; treatment of non-equidistant times involves trivial changes that only clutter the notation. Note also that this gives rise to vector autoregressive (VAR) forms of order 1 when G_t is constant over time.

[1] This review is contribution to the Handbook of Spatial Statistics, eds. Gelfand et al.

Example 1. First order models

When $G = I_p$, the identity matrix of order p, the model is the random walk $\beta_t = \beta_{t-1} + w_t$ and therefore model (2) can also be referred to as generalized random walk. This is also referred to as first order models because they can be seen as the first order (Taylor expansion) approximation of an arbitrary underlying smooth function β_t. Note that, unlike Gaussian processes, this model is non-stationary with $Var(\beta_t)$ increasing with t.

Stationary processes may be also be obtained after replacing the random walk evolution matrix I_p by suitably chosen matrices P. This gives rise to vector autoregressive (VAR) forms of order 1. Special cases of interest are given by $P = \rho I_p$, with $|\rho| < 1$, and $P = diag(\rho_1, ..., \rho_p)$, with $|\rho_i| < 1$, for $i = 1, ..., p$. These models are attractive for their simplicity and low dimensionality but may be too restrictive.

Example 2. Second order models or Dynamic linear trend (LT)

Assume $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$ is a bivariate process and let $G_t = G_{LT} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, for all t. Then clearly β_2 is undergoing a univariate random walk and β_1 is being incremented each time by β_2. Thus, β_1 plays the role of an underlying level and β_2 plays the role of its increment. Typically only β_1 is present in the data description. Both components are varying locally around their prescribed evolutions and can accommodate local changes.

The disturbance variance matrix for this model is hereafter denoted by W_{LT}. It may take any positive definite form but there are good reasons to assume it as $W_{LT} = \begin{pmatrix} W_1 + W_2 & W_2 \\ W_2 & W_2 \end{pmatrix}$, for all t. In any case, it is recommended to assume the disturbance at the mean level to be larger than the disturbance of its increment. Thus, the 1st diagonal element of W_{LT} would be larger than its 2nd diagonal element.

Example 3. Seasonal models

Assume a seasonal pattern of length p is to be described. Let $\beta = (\beta_1, ..., \beta_p)^T$ and $G_t = \begin{pmatrix} 0 & I_{p-1} \\ 1 & 0 \end{pmatrix}$, for all t. Clearly, G_t is a permutation matrix and the evolution over time only rearranges β components by replacing its 1st component

by its 2nd component in the preceding time. Thus, allowing only the first β component to be present in the data description gives a form free pattern for the seasonality. This pattern is stochastic due to the presence of the disturbance term w_t.

Structured seasonal patterns may also be constructed. A single sine wave form is obtained by letting $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$ and $G_t = G_S = \begin{pmatrix} c & s \\ -s & c \end{pmatrix}$, where $c = \cos(2\pi/p)$ and $s = \sin(2\pi/p)$, for all t. This evolution matrix makes β_1 take the appropriate value in the sine wave for every next time.

Once again, allowing only the first β component to be present in the data description gives a sine wave form for the seasonality. The pattern is stochastic and can accommodate variations around the sine wave due to the disturbance term w_t. This is usually associated with an additional intercept in the model for the observations since the sine wave fluctuates around 0. Combination of harmonics is obtained by allowing extra pairs of β components with different lengths and completely general forms are obtained by incorporating $[p/2]$ harmonics. See West & Harrison (1997, ch. 7?) for details and Harvey (1989) for an alternative model for seasonal components.

The non-parametric nature of these models is easier to understand with the usual choice of a random walk. In this case, the process β is simply undergoing local changes without imposing any specific form for the temporal variation and as such is capable of (locally) tracking any smooth trajectory over time. The degree of smoothness is governed by the variances W. Models that depart from the random walk impose some structure in the mean of the process, as described in the examples above. Even for these models, the presence of the disturbance terms allows departures from this structure and accommodates data fluctuations around it.

These models can also be obtained by discretization of an underlying stochastic differential equation (Revuz and Yor, 1999), as those used above in this book for handling spatio-temporal processes in continuous time.

The typical set-up for the use of dynamic models is the context of temporally correlated data $Y_t = (Y_t(s_1), ..., Y_t(s_n))$, for $t = 1, ..., T$, where. It will be assumed that all temporal correlation present in the data is captured by an

underlying process β. Therefore, the observations are conditionally independent (given β) leading to the likelihood for β given by $l(\beta) = \prod_{t=1}^{T} p(y_t|\beta_t)$, where T is the last observed time.

The simple but important case of a normal linear models gives $p(y_t|\beta_t)$ as

$$Y_t = \mu_t + v_t, \quad \text{with } \mu_t = X_t\beta_t \text{ and } e_t \sim N(0, V_t), \tag{3}$$

for $t = 1, ..., T$. The variance matrix V_t of the observation error e_t may be specified using any of the models previously described in this book to handle spatial correlation. The main forms are Gaussian processes, typically used in continuous space.

Also, the error e_t can be further decomposed into $e_t = \eta_t + \epsilon_t$, as before. This decomposition eases the generalization towards non-normal observations. Assume the observational distribution is governed by parameter ξ_t. The spatially structured error term η_t is incorporated to the predictor $X_t\beta_t + \eta_t$ via $g(\mu_t) = X_t\beta_t + \eta_t$, for some differentiable function g (Diggle, Tawn and Moyeed, 1998). An important example is the exponential family with mean μ_t. The pure noise ϵ_t retains the description of unstructured observational variation.

The $n \times m$ matrix X_t plays the role of a design matrix containing values of the explanatory variables at time t. It is typically given by known functions of location with rows $X(s_1)^T, ..., X(s_n)^T$, not depending on time and is thus denoted hereafter simply by X. Therefore, for any given location s, the observational predictor (mean, in the normal case) is given by

$$\mu_t(s) = X(s)^T\beta_t \tag{4}$$

Thus, models are being decomposed into a deterministic part given by $X(s)^T\beta_t$ and an unexplained stochastic component e_t that may incorporate the spatial dependence. Note that in this dynamic setting the deterministic part of the model is only handling temporal correlation. One natural choice in the spatio-temporal setting is to let matrix X be a function of the spatial coordinates.

This approach was proposed by Stroud, Muller and Sansó (2001). They chose to define $X(s)$ as a linear combination of basis functions of the location s. This idea is applied in related contexts by many authors. Wikle and Cressie (1999) use the same decomposition in their dimension reduction approach. They

obtained it from a more general underlying processes, to be described later in this section. Sansó, Schmidt and Nobre (2008) also use this decomposition but without the error term[2].

These approaches have the common feature of considering X as a fixed function of space. This may be too restrictive to accommodate general spatial variation. Lopes, Salazar and Gamerman (2006) allow the columns $f_1, ..., f_m$ of X to vary stochastically according to independent Gaussian processes. The m-dimensional time-varying component β_t plays the role of m latent factor time series capturing the temporal variation of the data. Each of its m elements β_{tj} is associated to the observations through the space-varying vector f_j containing their loadings, for $j = 1, ..., m$.

All models above decompose models into 2 groups of components: one handling space and one handling time. Even tough these structures may combine into non-separable models, richer dependence between space and time is not allowed. A description of processes that combine spatial and temporal dependence into a single structure that can not be separated is provided in what follows.

The key aspect in the extension is to allow state parameters β_t to vary across space. This will obviously imply a substantial increase in the parameter dimensionality and may lead to identifiability problems. The solution to keep the problem manageable and the model identifiable is to impose restrictions over the parameter space. This can be achieved through likelihood penalization from a classical perspective or through prior specifications from a Bayesian perspective.

2 Space varying state parameters

From now on, it will be assumed that the state parameter $\beta_t(\cdot)$ varies also over space. In this setting, $\beta_t(\cdot) = \{\beta_t(s) : s \in \mathcal{D}\}$. Considering n locations s_i $(i = 1, ..., n)$ for spatially continuous observation processes, the vector $\beta_t = (\beta_{t1}, ..., \beta_{tn})$ can be formed with $\beta_{ti} = \beta_t(s_i)$ denoting the state parameter at

[2]They also consider the possibility that X models time (in which case it will recover its subscript t) and β_t models space (in which case it drops its subscript t).

time t and location s_i.

A simple form to account for spatial and temporal variation of the state parameter is to assume that $\beta_t(\cdot)$ can be decomposed as

$$\beta_t(s) = \bar{\gamma}_t + \gamma_t(s) \tag{5}$$

with a trend $\bar{\gamma}_t$ common to all locations and a spatio-temporal disturbance $\gamma_t(s)$ associated with its location. Paez et al. (2008) assumed that the common trend $\bar{\gamma}_t$ carries the temporal evolution according to (2). The spatio-temporal disturbance process $\gamma_t(\cdot)$ accounts for the spatial correlation through a multivariate Gaussian process, which they assume to be independent and identically distributed over time. They applied this process to the intercept and the regression coefficients of a dynamic model for pollutant measurements.

Their approach allows more generality in the description of state parameters. Despite the substantial increase in the nominal number of parameters, it achieves identifiability through the decomposition (5) and prior assumptions about $\bar{\gamma}_t$ and γ_t. Note that the temporal independence between the γ_t´s prevents any temporal correlation between them. So, their model still separates the spatial components $\bar{\gamma}_t$ from the temporal components γ_t. Thus, their model can be useful if no temporal dependance of the spatial variation is expected.

The simplest model that does not allow for explicit separation of space and time is the spatial random walk

$$\beta_t(s) = \beta_{t-1}(s) + w_t(s), \text{ for all } s, \tag{6}$$

where $\beta_t(\cdot)$ is a univariate process. In model (6), state parameters β_t evolve in forms that are seemingly independent in space. But spatial correlation is introduced via their respective disturbance processes $w_t(\cdot)$ through their joint distribution. It does it by assuming some form of a Gaussian process. It will typically have a geostatistical model form given in chapter 2 for spatially continuous data. The prior is completed with a Gaussian process prior for $\beta_1(\cdot)$.

Obviously, the spatial random walk can be defined for multivariate state parameters $\beta_t(\cdot)$. All it requires is an adequate multivariate representation of Gaussian processes. Some possibilities for doing it are analysed in the areal data

context by Gamerman, Moreira and Rue (2002) and described in the spatially continuous context by Gamerman, Salazar and Reis (2007).

The decomposition (5) can also be applied to (6). In this case, the disturbances $w_t(s)$ must also be decomposed into a purely temporal disturbance $\bar{w}_t$ and residual spatio-temporal disturbances as

$$\begin{array}{rcl} \beta_t(s) & = & \bar{\gamma}_t + \gamma_t(s), \text{ for all } s \\ \bar{\gamma}_t & = & \bar{\gamma}_{t-1} + \bar{w}_t, \\ \gamma_t(s) & = & \gamma_{t-1}(s) + w_t(s), \text{ for all } s \end{array} \quad . \tag{7}$$

This decomposition was used by Gelfand, Banerjee and Gamerman (2005). They analysed environmental data with a normal linear regression model. Identification is ensured by setting the mean of $\gamma_t(s)$ to be 0. The first component $\bar{\gamma}_t$ accounts for purely temporal variation. The second component $\gamma_t(s)$ accounts for the remaining temporal variation that was associated with space. Both components are assumed in (7) to evolve according to a random walk. Huerta, Sansó and Stroud (2004) also used it to model the effect of seasonal components in an environmental application.

More general forms can now be constructed by combining (7) with (2) leading to a general evolution

$$\begin{array}{rcl} \beta_t(s) & = & \bar{\gamma}_t + \gamma_t(s), \text{ for all } s \\ \bar{\gamma}_t & = & \bar{G}_t \bar{\gamma}_{t-1} + \bar{w}_t, \\ \gamma_t(s) & = & G_t \gamma_{t-1}(s) + w_t(s), \text{ for all } s. \end{array} \quad , \tag{8}$$

with transition matrices $\bar{G}_t$ and G_t and evolution disturbances $\bar{w}_t$ and $w_t(\cdot)$. When all transition matrices equal the identity matrix, the spatio-temporal random walk (7) is recovered.

These structures and their decomposition allow for many types of components. Thus, they give substantial flexibility to the modeller. This can be more easily appreciated with the illustrative example below.

Example 4. Assume that a simple linear regression model

$$Y_t(s) = \alpha_t(s) + \beta_t(s) z_t(s) + e_t(s), \text{ for } t = 1, ..., T \text{ and } s = s_1, ..., s_n \tag{9}$$

is considered with a single covariate $z_t(s)$ varying in space and time. Note that

both the intercept α and the (scalar) regression coefficient β are allowed to vary in space and time.

Assume also that the intercept can be completely described by a stochastic seasonal pattern that is common throughout the region of interest and can be locally described by a single wavelength of length p and has no additional spatio-temporal heterogeneity. This would imply that $\alpha_t(s) = \alpha_t$, for all (s,t). According to Example 3, its evolution is described with the help of additional time-varying parameter ξ_t as

$$\begin{pmatrix} \alpha_t \\ \xi_t \end{pmatrix} = G_S \begin{pmatrix} \alpha_{t-1} \\ \xi_{t-1} \end{pmatrix} + \begin{pmatrix} w_t^{\alpha} \\ w_t^{\xi} \end{pmatrix}, \text{ where } \begin{pmatrix} w_t^{\alpha} \\ w_t^{\xi} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, W_{\alpha} \right]$$

where G_S was defined in example 2 and W_α is a 2×2 covariance matrix.

Assume further that the regression coefficient has its common trend in the form of a random walk but the spatial variations around this common mean are thought to undergo a location-specific, linearly local trend. The conditions stated above imply that $\beta_t(s) = \bar{\gamma}_t + \gamma_t(s)$, for all (s,t)). The evolutions for $\bar{\gamma}_t$ and γ_t given respectively given by a univariate random walk $\bar{\gamma}_t = \bar{\gamma}_{t-1} + \bar{w}_t$ and by

$$\begin{pmatrix} \gamma_t(s) \\ \delta_t(s) \end{pmatrix} = G_{LT} \begin{pmatrix} \gamma_{t-1}(s) \\ \delta_{t-1}(s) \end{pmatrix} + w_t(s)$$

where G_{LT} was defined in example 1 and additional processes had to be introduced: $\delta_t(\cdot)$ is a univariate increment (over γ_t) process and $w_t(\cdot) = (w_t^{\gamma}(\cdot), w_t^{\delta}(\cdot))$ is a bivariate process. There are many possibilities for the latter. Paez et al. (2008) assumed the same spatial correlation fuction for both disturbance processes. If this is assumed then a Kronecker product representation is obtained for the covariance matrix for the components of $w_t(\cdot)$ at any given set of locations. Independence between the processes can also be assumed leading to a block diagonal representation for the covariance matrix for the components of $w_t(\cdot)$ at any given set of locations. Other forms of Gaussian processes are also possible (see Gamerman, Salazar and Reis, 2007).

The spatial relation between the regression coefficients in (8) enables inference about their values at unobserved locations. Consider a set of g unobserved locations $s_1^u, ..., s_g^u$ where the superscript u denotes unobserved. Define the col-

lection of state parameters β_t^u at these locations as $\beta_t^u = (\beta_t^u(s_1), ..., \beta_t^u(s_g))$. Then clearly the evolution equations defined for β_t can be readily extended to (β_t, β_t^u). The conditional distribution for $(\beta_t^u|\beta_t^u)$ can be obtained from this joint specification. If all disturbances and prior distributions are normally distributed, then simple calculations show that this conditional distribution is also normal. These calculations are made conditionally on the hyperparameters. Their integration can not typically be performed analytically. In this case, approximating methods such as MCMC algorithms must be applied.

Figure 1 shows an example of this interpolation for the spatio-temporal variation experienced by the regression component γ_t. This result comes from a study of the effect of precipitation on temperature (Gelfand, Banerjee and Gamerman, 2005) with monthly data over a single year. Relevance of spatial and temporal components in this regression setting is clear. For example, a more extreme spatial variation of the effect of temperature is observed in the months of more extreme weather.

In many situations, main interest rests in predicting unobserved data values. Data may not be observed because they are located at unobserved sites or at unobserved times. In either case, the predictive distribution for them conditional on all observed data must be obtained. The operation in space is referred to as interpolation or kriging. The operation in time is referred to as forecasting when interest lies in prediction into the future, hindcasting when interest lies in assessment of the performance of the model at previous times and nowcasting when interest lies in the immediate future.

In any case, prediction is conceptually easy to perform. The structural form of the model means that all spatial and temporal correlations are contained in the state parameters. Let y^u denote the unobserved data one wants to predict and β^u denote the regression coefficient present in the observation equation for y^u. Depending on the case of interest, β^u may contain values of the state at unobserved locations and/or times. In either case, predictive distributions conditional on hyperparameters θ are obtained through

$$p(y^u|\theta, D) = \int p(y^u|\beta^u, \theta, D)p(\beta^u|\theta, D)d\beta^u, \tag{10}$$

where D denotes the observed data.

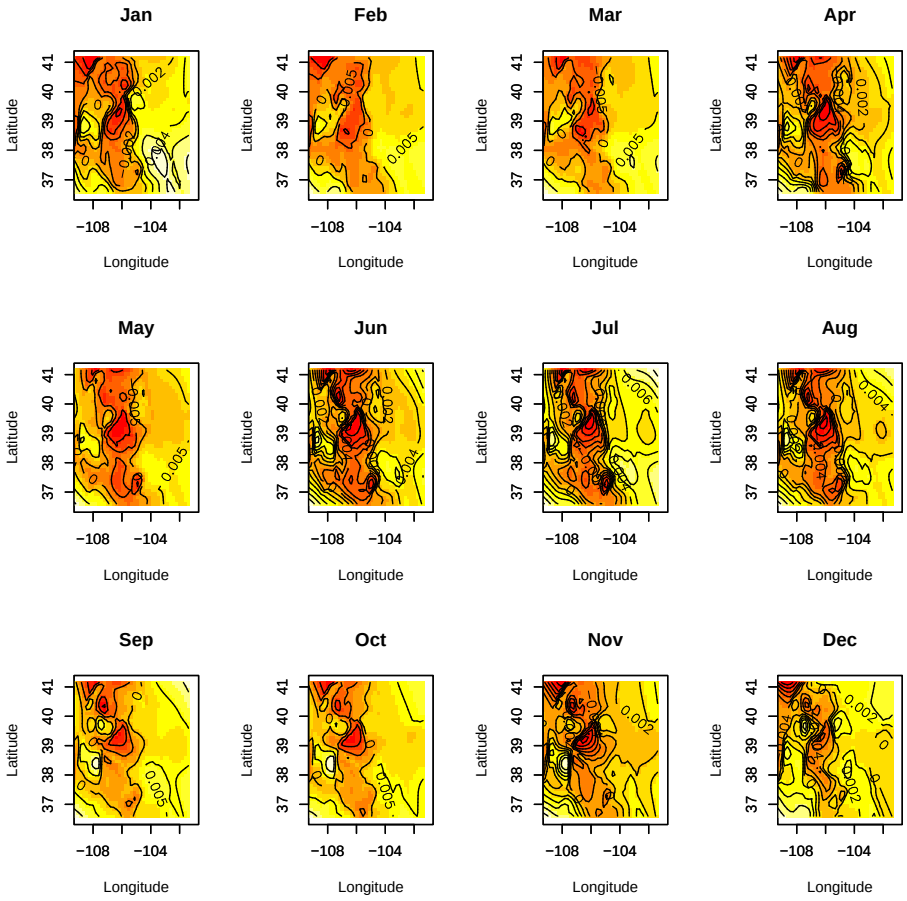

Figura 1: Posterior mean of the spatio-temporal variation γ_t of the regression coefficient of precipitation over temperature for a region of the State of Colorado, USA (reprinted from Gelfand, Banerjee and Gamerman, 2005).

The first density in the integrand is the observation equation and the second is the posterior distribution of β^u. The integration in (10) can be performed analytically in the case of normal evolution disturbances and normal observation errors. Elimination of hyperparameters is required to obtain the unconditional predictive distributions actually used for prediction. This operation is performed in very much the same way as (10) with β^u replaced by the hyperparameters. Namely, the predictive density $p(y^u|D)$ can be obtained via

$$p(y^u|D) = \int p(y^u|\theta, D)p(\theta|D)d\theta.$$

The integrand above contains the posterior density $p(\theta|D)$ of the hyperparameters and this is rarely available analytically. Thus, integration can only be performed approximately and approximating methods must be applied. In practice, MCMC/sampling methods are applied and integration with respect to β^u and θ is performed simultaneously. Figure 2 provides a visual summary of these prediction operations in the context of the application of Figure 1. Note that spatial extrapolation is more dispersed than temporal extrapolation for this application.

Models described in (8) retain the seemingly unrelated nature because their mean structure is location-specific and correlation across space is only provided through their unstructured error terms. Correlation across space can be imposed directly through the mean structure by forms such as

$$\beta_t(s) = \int k(u, s)\beta_{t-1}(u) + w_t(s), \tag{11}$$

where $k(u, s)$ is a kernel that provides the weights with which location u influences outcomes in location s for the next time. This evolution is considered by a number of authors in a number of different contexts (see, for example, Wikle and Cressie (1998)). When the integral can be well approximated by a discrete convolution over observed locations, then (11) falls into the general form (2) with the (i, j)th. entry of the transition matrix given by the values of $k(s_j, s_i)$, for $i, j = 1, ..., n$. Evolution (11) has been used only for the intercept but nothing prevents its use for more general state parameters such as regression coefficients or seasonal components.

The presentation of this Section is based on the Bayesian paradigm. Thus,

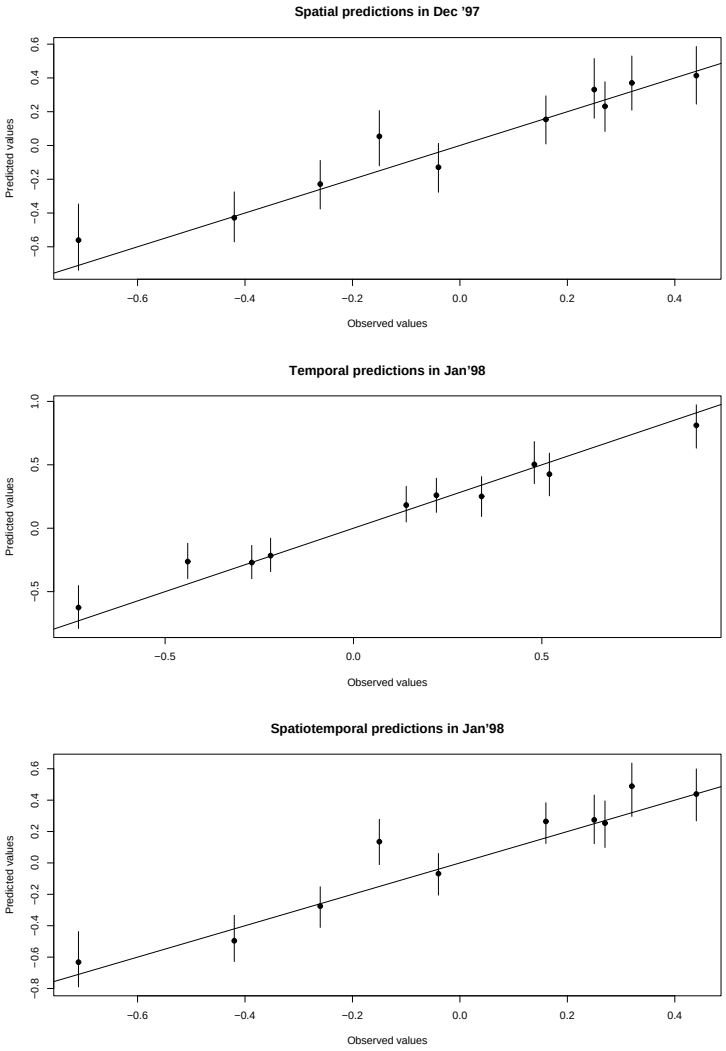

Figura 2: Predictive credible intervals for temperature values at a few unobserved locations (top panel), at a future time for a few observed locations (middle panel) and at a future time for a few unobserved locations (bottom panel) for a region of the State of Colorado, USA (reprinted from Gelfand, Banerjee and Gamerman, 2005).

prior distributions for the hyperparameters were also specified and inference was based on the posterior distribution. The classical paradigm may also be applied (Harvey, 1989). Its use can be illustrated in the context of prediction.

The classical approach is based on integrating out the state parameters in the operation described in (10). This gives rise to the integrated likelihood of the hyperparameters $l(\theta) = p(y^u|\theta, D)$. Maximum likelihood estimates can be approximately obtained by numerical operations. Confidence intervals and hypotheses testing can be performed but they require further information about the likelihood or about the sampling distribution of the maximum likelihood estimator. Once again, approximating methods based on asymptotic theory or on sampling techniques (such as bootstrap) must be applied to extract such information.

3 Applications

The class of models described above can be used in a number of contexts where space and time interact. The simplest and most obvious one is the context of a univariate component playing the role of mean of a spatio-temporal observation process, ie, the observation process is given by $y_t(s) = \mu_t(s) + e_t(s)$, where the mean response $\mu_t(s)$ is described by a dynamic Gaussian process (**??**). One generalization of this idea is achieved by considering the regression context with spatio-temporal heterogeneity that was described above. This way, not only the intercept but also the regression coefficients may vary in space-time stochastically.

In general terms, the ideas above can be used to incorporate temporal extensions to parameters of spatial models and also into spatial extensions to parameters of dynamic models. The former accommodates temporal dependence and the latter accommodates spatial dependence.

Spatial dependence was stochastically incorporated into dynamic factor models by Lopes, Salazar and Gamerman (2006). They used the loading matrix to achieve that. Lopes and Carvalho (2007) showed the relevance of including dynamics into the factor loadings. Thus, combination of these ideas may lead to fruitful possibilities and can be implemented with the class of models described

in this section. Salazar (2008) implemented these ideas. Figure 3 shows some promising results obtained with simulations.

Another natural application area for these ideas is point process modeling. This is the observation process where events occur in a given region and their location is registered. The usual approach in this setting is to define an intensity rate, that governs the occurrence of events. Under conditional independence assumption, Poisson processes become appropriate and the intensity rate suffices to completely characterize the process. A further stochastic layer can be introduced by allowing the intensity rate to be random (Cox, 1955). A popular choice for the intensity distribution is a log Gaussian process (Møller, Syversveen, and Waagepetersen, 1998). From a Bayesian perspective, this is equivalent to a Gaussian process prior for the logarithm of the intensity rate.

Point processes can also be observed over time. In this case, the intensity rate process is a time-continuous sequence of functions over space. This is a natural framework for application of the ideas above and for the use dynamic Gaussian processes as prior distributions for the intensity rate process. Reis (2008) explores this path in a number of examples and applications.

A similar representation for the intensity rate is entertained by Brix and Diggle (2001). They considered the time-continuous differential equation specification and performed classical estimation using moment-based estimators. Calculations were performed approximately by discretizations over space and time. Details are provided in Diggle (2009).

Another area for further exploration of dynamic Gaussian processes is spatial non-stationarity. Many of the models suggested to handle non-stationarity are built over generalizations of stationary Gaussian processes. The spatial deformations of Sampson and Guttorp (1992) and the convolutions of Higdon, Swall and Kern (1999) and Fuentes (2001) are among the most cited references in this area. Schmidt and O'Hagan (2002) and Damien, Sampson and Guttorp (2000) have independently casted the deformation problem into the Bayesian paradigm. See also Sampson (2009) for more details.

The approaches above make use of Gaussian processes for handling non-stationarity. The convolutions of Fuentes (2001) are based on kernel mixtures of Gaussian processes. The hyperparameters governing these processes may be

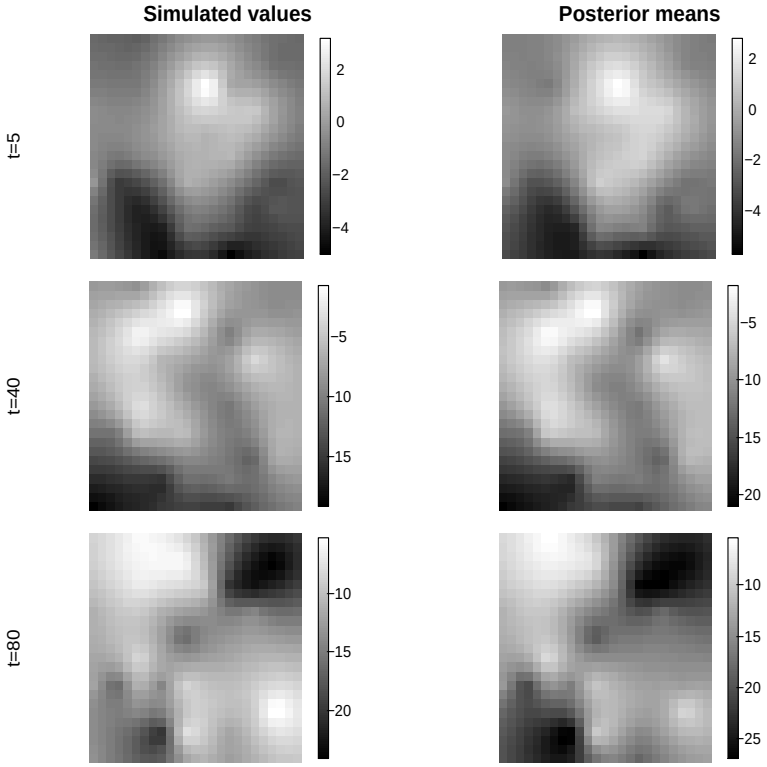

Figura 3: Results of simulation with a dynamic model with 2 factors and observation window of 30 locations and 80 equally spaced times. Loadings of the first factor: simulated values (left) and posterior means (right). The first, second and third rows refer to t = 5, 40 and 80, respectively. Posterior means for loadings at unobserved locations were obtained by Bayesian interpolation (reprinted from Gamerman, Salazar and Reis, 2007).

related over space. Gaussian process prior distribution is one of the choices in this setting and this was entertained by Fuentes (2001). The deformations approach requires the specification of a prior distribution in a Bayesian context or a penalization in a classical setting for the deformed space. In either case, a natural choice would be a Gaussian process.

There are cases when the observation process may span over a period of time. The spatial correlation structure may remain constant over the period. But the spatial non-stationarity may also vary over time. This is at the very least plausible and in many cases highly likely to occur. In this case, the time constancy is not longer valid. Alternatives must be sought to appropriately account for this variation. This is a natural setting for the consideration of dynamic Gaussian process. They can provide a starting exploration step. Due to their local behaviour, they are able to describe the variation with a non-parametric flavour, without imposing any specific form for the changes over time that these hyperparameter may experience. Once a specific time pattern is observed for the parameters, specific assumptions about this change can be incorporated into the model.

Consider, for example, the convolution of Gaussian processes to account for spatial non-stationarity of the data generating processes. Among the parameters defining the Gaussian processes is their sill parameters. They can be allowed to change not only over space but also over time. This would allow for many purposes: to borrow information across nearby locations and consecutive time periods and to smooth variations of this parameter over space and over time.

4 Further remarks

The material of the last subsection is of more speculative nature. The ideas described here are just beginning to be used. They involve use of elaborate model specifications that are far from easy to be estimated from the data. This poses yet another challenge in the use of this methodology. The most common approach these days from a Bayesian perspective is MCMC (see Gamerman and Lopes, 2006). Alternatives based on non-iterative approximations are also being proposed (Rue, Martino and Chopin, 2008).

This section was mostly based on the so-called latent approach where spatial and temporal dependences are incorporated in the model through the use of latent structures. Other possibilities are also avaliable. Spatial autoregression (SAR) provides a natural framework to contemplate spatial dependence directly at the observational level (Anselin, 1988). Addition of temporal components can be made separately in a different model block or jointly. Gamerman and Moreira (2004) have simply added a temporal autoregressive component to a multivariate SAR model. Kakamu and Polasek (2007) considered a SAR structure over (temporally) lagged variables thus inducing spatial and temporal dependence simultaneously. These are just a couple of the many possibilities available through this approach.

The purpose of this section is to draw attention to a tool that is flexible and can accommodate many patterns of spatio-temporal data variation. That are other areas that can become potential applications for this technology. It is hoped in this section that readers have their attention drawn to these modelling tools, find them useful and eventually try them in their own problems at hand.

References

Anselin, L. (1988). *Spatial Econometrics*.

Brix, A. and Diggle, P. J. (2001). Spatiotemporal prediction for log-Gaussian Cox processes. *Journal of the Royal Statistical Society Series B*, **63**, 823-841.

Cox, D.R. (1955). Some statistical models related with series of events. *Journal of the Royal Statistical Society Series B*, **17**, 129-164.

Damian, D., Sampson, P. D. and Guttorp, P. (2000). Bayesian estimation of semi-parametric non-stationary spatial covariance structure. *Environmetrics*, **11**, 161-176.

Diggle, P. J., Tawn, J. A. and Moyeed, R. A. (1998). Model-based geostatistics (with discussion). *Applied Statistics*, **47**, 299-350.

Diggle, P. J. (2009). Spatio-temporal point processes. In *Handbook of Spatial Statistics*, eds. Gelfand et al., chapter 5.

Fuentes, M. (2001). A high frequency kriging for nonstationary environmental processes. *Environmetrics*, **12**, 1-15.

Gamerman, D. and Lopes, H. F. (2006). *Monte Carlo Markov Chain: Stochastic Simulation for Bayesian Inference* (2nd edition). Chapman and Hall: London.

Gamerman, D. and Moreira, A. R. B. (2004). Multivariate spatial regression models. *Journal of Multivariate Analysis*, **91**, 262-281.

Gamerman, D., Moreira, A. R. B. and Rue (2003). Space-varying regression models: specifications and simulation. *Computational Statistics and Data Analysis*, **42**, 513-533.

Gamerman, D., Salazar, E. and Reis, E. (2007). Dynamic Gaussian process priors, with applications to the analysis of space-time data (with discussion). In *Bayesian Statistics 8* (eds. Bernardo, J. M. et al.), pp. 149-174. Oxford University Press.

Gelfand, A. E., Banerjee, S. and Gamerman, D. (2005). Spatial process modelling for univariate and multivariate dynamic spatial data. *Environmetrics*, **16**, 465-479.

Harvey, A. (1989). Forecasting, Structural Time Series Models and the Kalman Filter. University Press, Cambridge.

Higdon, D. M., Swall, J. and Kern, J. (1999), Non-Stationary spatial modeling. In *Bayesian Statistics 6* (eds. Bernardo, J. M. et al.), 761-768. Oxford University Press.

Huerta, G., Sansó, B. and Stroud, J. (2004). A spatiotemporal model for Mexico city ozone levels. *Applied Statistics*, 53, 231-248.

Kakamu, K. and Polasek, W. (2007). Cross-sectional space-time modeling using ARNN(p, n) processes. Technical report # 203, Institute for Advanced Studies, Vienna.

Lopes, H. F. and Carvalho, C. (2007) Factor stochastic volatility with time varying loadings and Markov switching regimes, Journal of Statistical Planning and Inference, 137, 3082-91.

Lopes, H. F., Salazar, E. and Gamerman, D. (2006). Spatial dynamic factor analysis. Technical Report, Universidade Federal do Rio de Janeiro.

Møller, J., Syversveen, A. and Waagepetersen, R. (1998). Log Gaussian Cox processes. *Scandinavian Journal of Statistics*, **25**, 451-482.

Paez, M. S., Gamerman, D., Landim, F. M. P. F. and Salazar, E. (2008).

Spatially-varying dynamic coefficient models. *Journal of Statistical Planning and Inference*, **138**, 1038-1058.

Reis, E. (2008). *Dynamic Bayesian Models for Spatio-temporal Point Processes*. Unpublished Ph.D. thesis (in Portuguese). Universidade Federal do Rio de Janeiro.

Revuz, D. and Yor, M. (1999). *Continuous Martingales and Brownian Motion* (3rd ed.). Springer.

Rue, H., Martino, S. and Chopin, N. (2008). Approximate Bayesian Inference for Latent Gaussian Models Using Integrated Nested Laplace Approximations. To appear in *Journal of the Royal Statistical Society Series B*.

Salazar, E. (2008). *Dynamic Factor Models with Spatially Structured Loadings*. Unpublished Ph.D. thesis (in Portuguese). Universidade Federal do Rio de Janeiro.

Sampson, P. D. (2009). Constructions for non-stationary spatial processes. In *Handbook of Spatial Statistics*, eds. Gelfand et al., chapter 2.

Sampson, P. D. and Guttorp, P. (1992). Nonparametric Estimation of Nonstationary Spatial Covariance Structure, Journal of the American Statistical Association, 87, 108-119.

Sansó, B., Schmidt, A. M. and Nobre, A. A. (2004). Bayesian spatio-temporal models based on discrete convolutions. To appear in *Canadian Journal of Statistics*.

Schmidt, A. M. and O'Hagan, A. (2003). Bayesian Inference for Nonstationary Spatial Covariance Structures via Spatial Deformations. *Journal of the Royal Statistical Society Series B*, **65**, 743-758.

Stroud, J., Muller, P. and Sansó, B. (2001). Dynamic models for spatiotemporal data. *Journal of the Royal Statistical Society, Series B*, **63**, 673-689.

West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models* (2nd ed.) Springer.

Wikle, C. and Cressie, N. (1999). A dimension-reduced approach to space-time Kalman filtering. *Biometrika*, **86**, 815-829.