Spatial-temporal modeling

- General modeling.
- Point-level modeling with continuous time
- Separable and nonseparable spatial-temporal models
- Dynamic spatio-temporal models

Due to the proliferation of data sets that are both spatially and temporally indexed, spatial-temporal modeling has received an dramatically increased attention in the last few years.

We will consider the case of point-level spatial data and areal data.

General modeling formulation

Consider the case of point-referenced data where time is discretized.

We can look at space-time indexed data $Y(x,t)$ in two ways:

- Writing $Y(x,t) = Y_x(t)$ as a spatially varying time series model.
- Write $Y(x,t) = Y_t(x)$ as a temporally varying spatial model.

We can create the singular value decomposition (or EOF) already discussed. Assume $T > n$, (where T is the number of observations over time, and n the number of location sites), we can write

$$Y = UDV = \sum_{l=1}^{T} d_l U_l V_l^T$$

where U is $n \times n$ orthogonal matrix with columns U_l, each

$U_l = (u_l(s_1), \ldots, U_l(s_n))$, and V is a $T \times T$ orthogonal matrix with columns V_l, each $V_l = (v_l(1), \ldots, v_l(T))^T$, D is a $n \times T$ matrix of the form $(\Delta, \mathbf{0})^T$ where Δ is $T \times T$ diagonal with diagonal entries D_l. Assume d_l's are arranged in decreasing order of their absolute values. $U_l V_l^T$ is referred to as the lth empirical orthogonal function. We can write

$$Y(s_i, t) = \sum d_l u_l(s_i) v_l(t).$$

Note that

$$YY^T = \sum_{l=1}^{T} d_l^2 U_l U_l^T$$

The first and second (maybe we need more than that) EOFs provide most of the information of the spatial structure.

EOFs are a useful exploratory tool to learn about the spatial structure of the data, but for full inference we need a full spatiotemporal model.

Model formulation

Denote $Y(s,t)$ measurement at location s and time t.

$$Y(s,t) = \mu(s,t) + \epsilon(s,t)$$

μ is the mean structure and ϵ the residual. If we write

$$\mu(s,t) = x(s,t)\beta(s,t)$$

where x are known covariates, and β are spatio-temporally varying coefficients. The error term ϵ can be rewritten as $w(s,t) + e(s,t)$ where e is a Gaussian white noise process and w is a mean-zero spatiotemporal process. Then, this would be a hierarchical model with a conditionally independent first stage given μ and w. The distribution:

Stage 1.

$$Y|\mu w \sim Normal(\mu + w, e)$$

instead of a normal it could be a member of the exponential family, with a link function g that it is $g(\eta(s,t)) = \mu(s,t) + w(s,t)$.

Spatiotemporal richness is captured by extending the model for $e(s,t)$. Assuming that time is discretized. We could have the 3 next different models that avoid specification of space-time interactions:

model 1:

$$e(s,t) = \alpha(t) + w(s) + \epsilon(s,t)$$

model 2:

$$e(s,t) = \alpha_s(t) + \epsilon(s,t)$$

model 3:

$$e(s,t) = w_t(s) + \epsilon(s,t)$$

$\epsilon(s,t)$ are iid normals. The first one provides an additive form in temporal and spatial effects. The second provides temporal evolution at each spatial location. The third provides a spatial structure at each time.

If t is continuous, we could model $\alpha(t)$ in the previous model, as one-dimensional stationary Gaussian process.

$$\alpha = (\alpha(1), \ldots, \alpha(t_m)) \sim N(0, \sigma_\alpha^2 \Sigma(\phi))$$

with $\Sigma(\phi)_{rs} = \rho(|t_r - t_s|; \phi)$.

Alternatively, we could model (in a discrete time setting)

$$\alpha(t) = \rho\alpha(t-1) + \eta(t)$$

where η are i.i.d. (if $|\rho| < 1$ we have an stationary AR model, but ρ is also allowed to be 1, that leads to an improper prior).

Regarding the $\alpha_s(t)$ components, we could model them as

$$\alpha_s(t) = \rho\alpha_s(t-1) + \eta_s(t)$$

$\eta_s(t)$ are i.i.d

The $w_t(s)$ are modeled as independent spatial processes.

For areal data we write,

$$Y_{it} = \mu_{it} + e_{it}$$

now $\mu_{it} = x_{it}\beta$ and

$$e_{it} = w_{it} + \epsilon_{it}$$

w_{it} are spatiotemporal random effects (with a CAR specification). Areal unit data are often non-Gaussian (e.g., sparse counts), we could view this model as a hierarchical model, and replace the first stage Gaussian specification with a for e.g. Poisson.

Assuming

$$Y(s,t) = \mu(s,t) + \epsilon(s,t)$$

With ϵ modeled using (models 1, 2 and/or 3). We obtain the likelihood function, $f(Y|parameters)$, which is normal. We can also obtain the predictive posterior distribution at some location (x_0, t_0), given the data $\mathbf{Y}$,

$$f(Y(s_0,t_0)|\mathbf{Y}) = \int f(Y(s_0,t_0)|\mathbf{Y}, parameters)p(parameters|Y).$$

Point-level modeling with continuous time

Suppose that $s \in \mathbb{R}^2$ and $t \in \mathbb{R}^+$, we seek to define a spatiotemporal process $Y(s,t)$. We need to specify a valid spatiotemporal covariance. It is not reasonable to use one of the known models (exponential, matern,...) in $\mathbb{R}^3$, because distance in space has nothing to do with "distance" in time.

A possibility is to change the temporal scale (by multiplying time by α), so distances in space and time can be used together in a 3-dimensional version of some of the known covariance models, e.g. C_0. We define the covariance

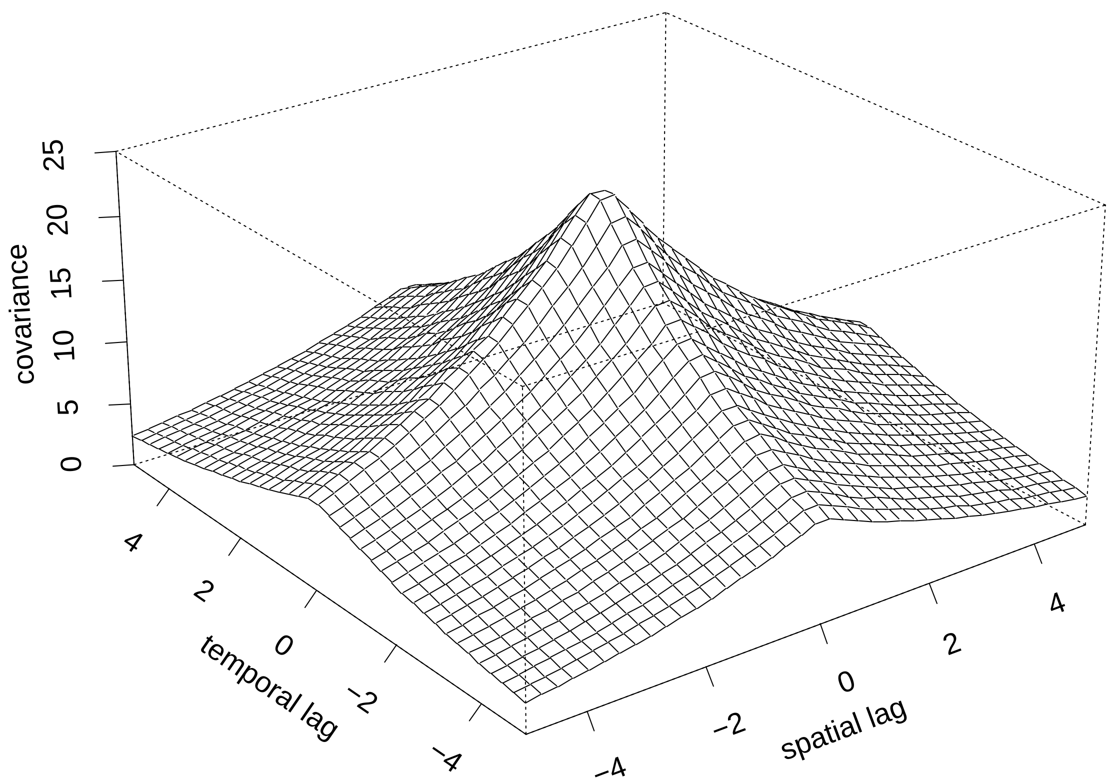
$$Cov(Z(s,t), Z(s',t')) = C_0(|(s, \alpha t) - (s', \alpha t')|)$$

Separable models

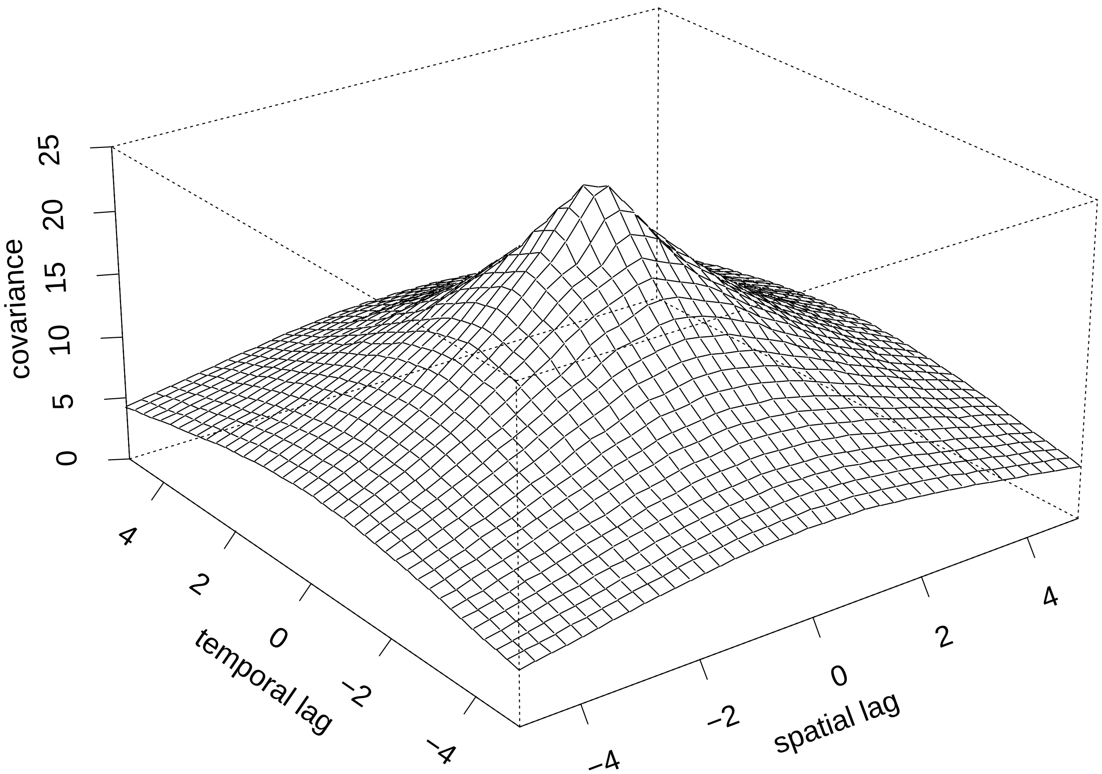
A spatial-temporal field $Z(\mathbf{s}, t)$, where $\mathbf{s}$ represent space and t time, is separable if $\text{Cov}\{Z(\mathbf{s}, t), Z(\mathbf{s}', t')\} = C_1(\mathbf{s}, \mathbf{s}')C_2(t, t')$ for some spatial covariance C_1 and temporal covariance C_2.

A class of nonseparable spatial-temporal models were proposed by Cressie and Huang (JASA, 99), Gneiting (JASA, 02), Stein (2003) and Chen, Fuentes and Davis (2004).

A separable covariance



A non−separable covariance



The spectral density, f, which is the Fourier transform of the spatial-temporal covariance function:

$$f(\boldsymbol{\omega}, \tau) = \frac{1}{(2\pi)^{d+1}} \int_{\mathbb{R}^d} \int_{\mathbf{r}} \exp(-i\boldsymbol{\omega}^T \mathbf{x} - i\tau t) C(\mathbf{x}, t) d\mathbf{x} dt, \tag{1}$$

and the corresponding covariance function is given by

$$C(\mathbf{x}, t) = \int_{\mathbb{R}^d} \int_{\mathbf{r}} \exp(i\boldsymbol{\omega}^T \mathbf{x} + i\tau t) f(\boldsymbol{\omega}, \tau) d\boldsymbol{\omega} d\tau. \tag{2}$$

For example, the Matérn spatial spectral density is given by

$$f(\boldsymbol{\omega}) = \gamma(\alpha^2 + |\boldsymbol{\omega}|^2)^{-\nu - d/2},$$

and its corresponding Matérn spatial covariance function is

$$C(\mathbf{x}) = \frac{\pi^{d/2}\gamma}{2^{\nu-1}\Gamma(\nu + \frac{d}{2})\alpha^{2\nu}} (\alpha|\mathbf{x}|)^{\nu} \mathcal{K}_{\nu}(\alpha|\mathbf{x}|),$$

where $\mathcal{K}_{\nu}(\alpha|\mathbf{x}|)$ is a modified Bessel function.

We propose the following spatial-temporal spectral density (Chen, Fuentes and Davis, 2004) that has a separable model as a particular case The spectral density f of Z changes with space and time to explain how the spatial temporal dependency varies on the domain of interest.

Locally (in a neighborhood of $\mathbf{s}_i = (\mathbf{x}_i, t_i)$, to allow lack of stationarity) we propose the following parametric model for f,

$$f_{\mathbf{s}_i}(\boldsymbol{\omega}, \tau) = \gamma_i(\alpha_i^2\beta_i^2 + \beta_i^2|\boldsymbol{\omega}|^2 + \alpha_i^2|\tau|^2 + \epsilon|\boldsymbol{\omega}|^2\tau^2)^{-\nu_i}, \tag{3}$$

if we have stationarity the parameters of the model do not change with location (e.g. it is simply α, rather than α_i), we have γ_i, α_i and β_i are positive, $\nu_i > \frac{d+1}{2}$ and $\epsilon \in [0, 1]$. The parameter α_i^{-1} explains the rate of decay of the spatial correlation. For the temporal correlation, the rate of decay is explained by the parameter β_i^{-1}, γ_i is a scale parameter.

When $\epsilon = 1$, the previous equation can be written as

$$\begin{aligned} f_{\mathbf{s}_i}(\boldsymbol{\omega}, \tau) &= \gamma_i(\alpha_i^2\beta_i^2 + \beta_i^2|\boldsymbol{\omega}|^2 + \alpha_i^2{\tau_i}^2 + |\boldsymbol{\omega}|^2\tau^2)^{-\nu_i} \\ &= \gamma_i(\alpha_i^2 + |\boldsymbol{\omega}|^2)^{-\nu_i}(\beta_i^2 + \tau_i^2)^{-\nu_i}. \end{aligned}$$

Therefore the corresponding spatial-temporal covariance is separable, both spatial component and temporal component are Matérn type covariances.

When $\epsilon = 1$, $\gamma_i = \alpha_i = \beta_i = d = 1$ and $\nu = 3/2$, a contour plot of the corresponding separable spatial-temporal covariance is given here. There are sharp ridges.

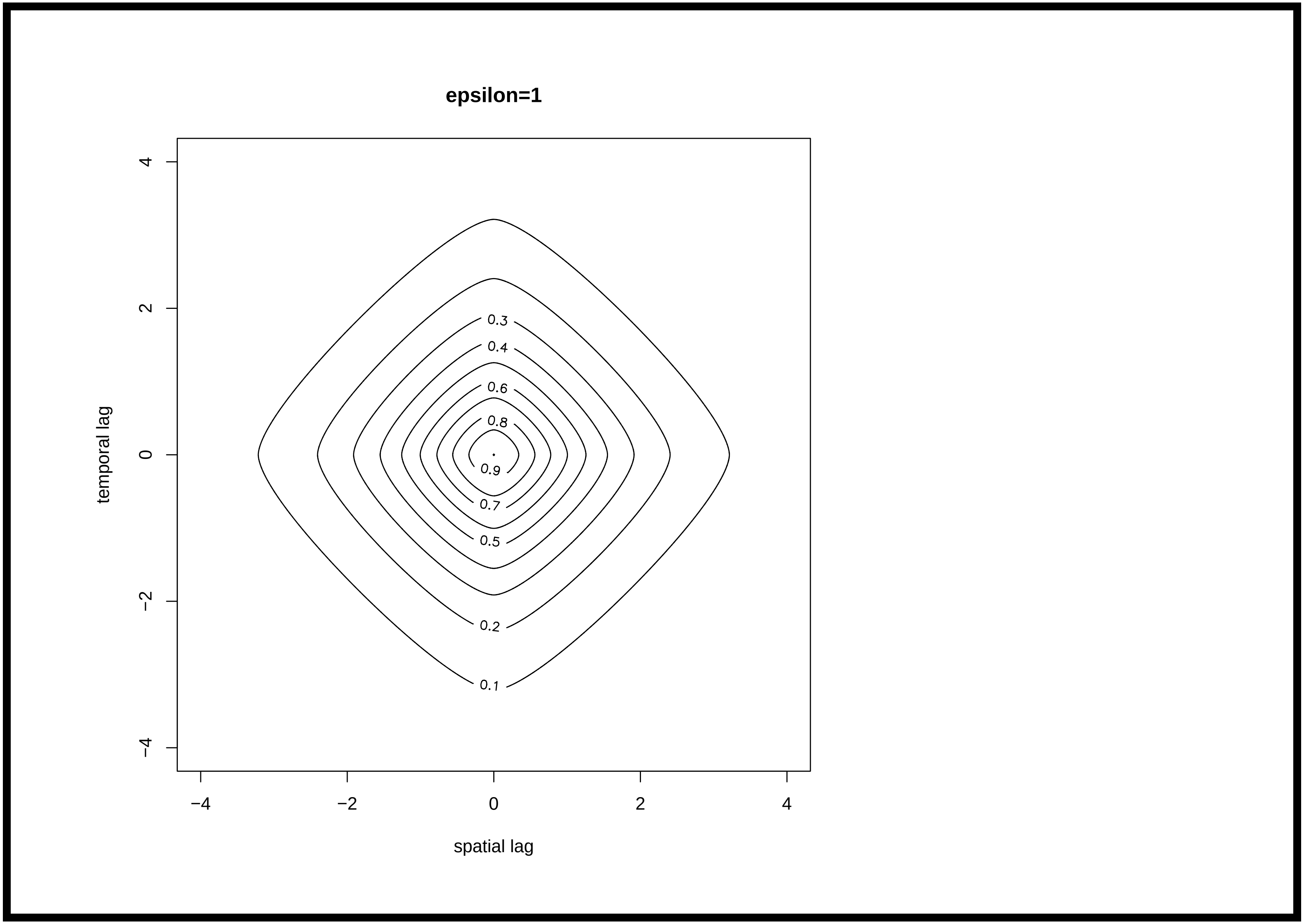

epsilon=1
temporal lag
spatial lag
4
2
0
−2
−4
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

Figure 3:

When $\epsilon = 0$,

$$f_{\mathbf{s}_i}(\boldsymbol{\omega}, \tau) = \gamma_i(\alpha_i^2\beta_i^2 + \beta_i^2|\boldsymbol{\omega}|^2 + \alpha_i^2|\tau|^2)^{-\nu_i}.$$

This is an extension of traditional Matérn spectral density. The parameter α_i^{-1} explains the rate of decay of the spatial correlation. For the temporal correlation, the rate of decay is explained by the parameter β_i^{-1}, γ_i is a scale parameter.

The corresponding spatial-temporal covariance is a 3-d Matérn type covariance with an extra parameter, which can be considered a conversion factor between the units in the space and time domains.

When $\epsilon = 0$,

$$C(\mathbf{x}, t) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} \left(\frac{\|(\mathbf{x}, \rho t)\|}{r}\right) \mathcal{K}_\nu \left(\frac{\|(\mathbf{x}, \rho t)\|}{r}\right),$$

- r is the range.

- σ^2 is the sill.
- $\nu > 0$ measures the smoothness of Z.
- ρ, which is new, is a scale factor to take into account the change of units between the spatial and temporal domains.
- This is a $d+1$ Matérn type covariance, but it takes into account the change of units between space domain and temporal domain.

When $\epsilon = 0$, $\gamma = \alpha = \beta = d = 1$ and $\nu = 3/2$, a contour plot of corresponding separable spatial-temporal covariance is given here. It does not have ridges.

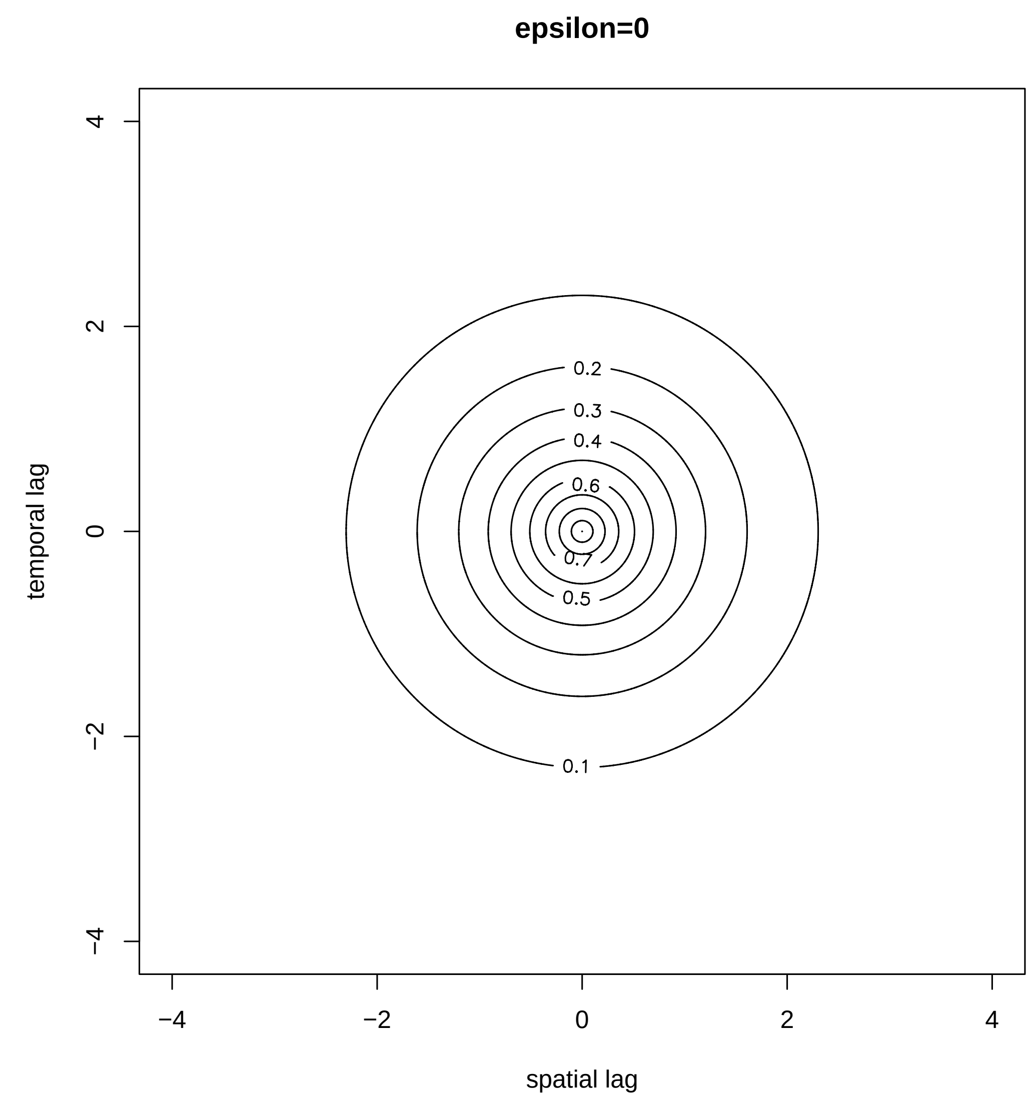


Figure 4:

Space-time covariances for different values of ϵ,

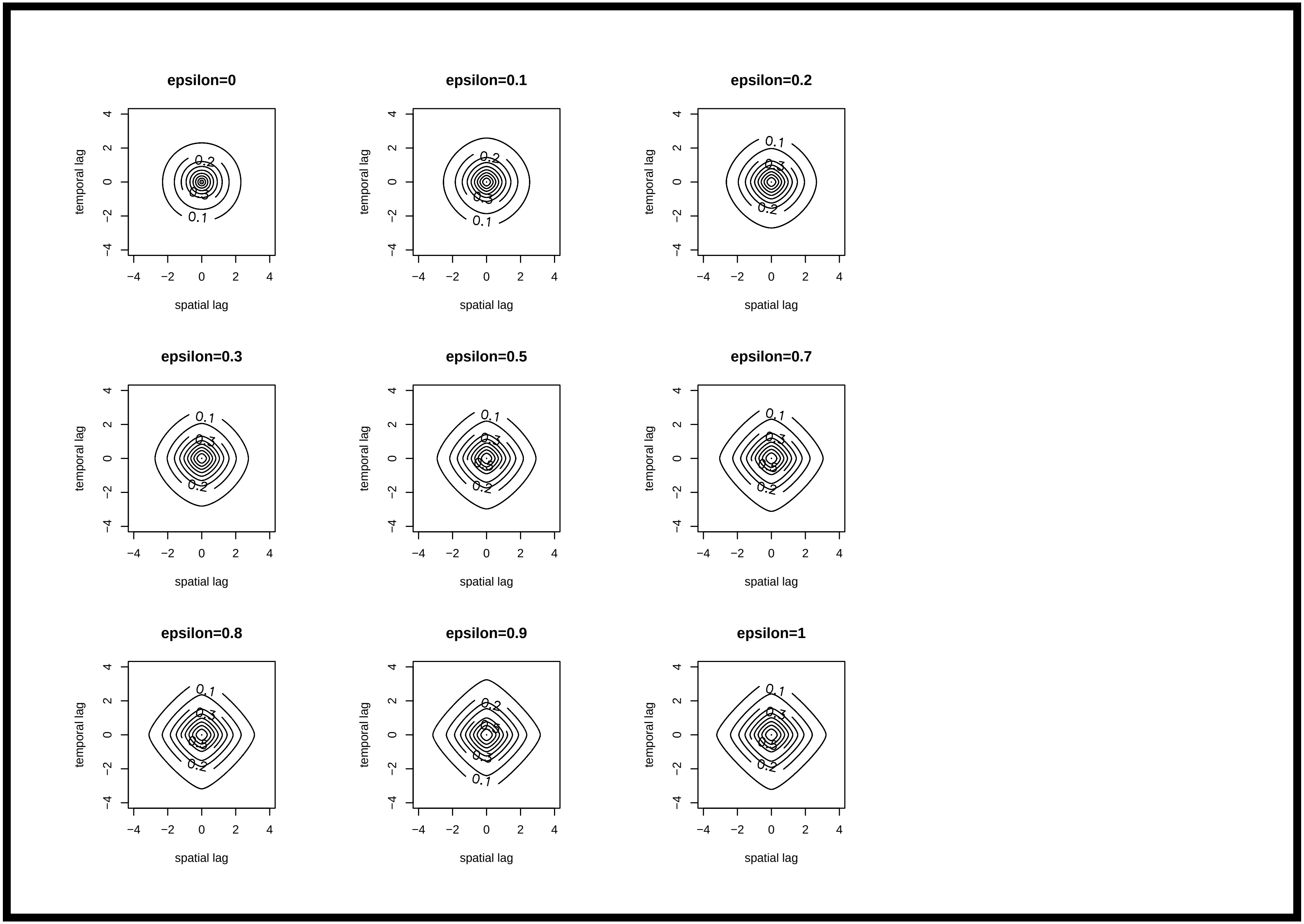

epsilon=0
epsilon=0.1
epsilon=0.2
epsilon=0.3
epsilon=0.5
epsilon=0.7
epsilon=0.8
epsilon=0.9
epsilon=1
temporal lag
spatial lag

Figure 5:

In summary, the new class spectral density in (3) is nonseparable for $0 \leq \epsilon < 1$, and separable for $\epsilon = 1$. Therefore, the parameter ϵ plays a role for separability. It controls the interaction between the spatial component and the temporal component. When ϵ equals 0 and 1, there are exact forms for the corresponding spatial-temporal covairances. Otherwise, the corresponding spatial-temporal covariance has to be computed numerically.

Cressie and Huang propose a generic approach to developing parametric models for spatial-temporal processes. The method relies heavily on spectral representations for the theoretical space-time covariance structure, and generalizes the results of Matern for pure spatial processes. In essence, Matérn constructs a number of parametric families for spatial processes by direct inversion of spectral densities. Cressie and Huang show that the same ideas can be used to construct families of spatial-temporal covariances.

First, Cressie and Huang represent the stationary spatial-temporal covariance $C(h, u)$ as

$$C(h, u) = \int \int e^{i(h^T \omega + u\tau)} g(\omega, \tau) d\omega d\tau \tag{4}$$

where $C(h, u)$ is a stationary spatial-temporal covariance function in which h represents a d-dimensional spatial vector and u is a scalar time component. The function $g(\omega, \tau)$, where ω is d-dimensional and τ is scalar, is the spectral density of the covariance function C. The function g may be written as a scalar Fourier transform in τ,

$$g(\omega, \tau) = \frac{1}{2\pi} \int e^{-iu\tau} h(\omega, u) du$$

with inverse

$$h(\omega, u) = \int e^{iu\tau} g(\omega, \tau) d\tau. \tag{5}$$

Putting (4) and (5) together,

$$C(h, u) = \int e^{ih^t\omega} h(\omega, u) d\omega. \tag{6}$$

The next step is to write

$$h(\omega, u) = k(\omega)\rho(\omega, u) \tag{7}$$

where $k(\omega)$ is the spectral density of a pure spatial process and $\rho(\omega, u)$ for each ω is a valid temporal autocorrelation function in u. Cressie and Huang remark that any smooth space-time covariance function can be written in the form (6) and (7), and they also impose the conditions:

(C1) For each ω, $\rho(\omega, \cdot)$ is a continuous temporal autocorrelation function, $\int \rho(\omega, u)du < \infty$ and $k(\omega) > 0$;

(C2) $\int k(\omega)d\omega < \infty$.

Under those conditions, the generic formula for $C(h, u)$ becomes

$$C(h, u) = \int e^{ih^T\omega} k(\omega)\rho(\omega, u)d\omega. \tag{8}$$

When $\rho(\omega, u)$ is independent of ω, (8) reduces again to a separable model. Cressie et al. developed seven special cases of (8). For example,

$$\rho(\omega, u) = \exp\left(-\frac{\|\omega\|^2 u^2}{4}\right) \exp\left(-\delta u^2\right), \qquad (\delta > 0)$$

$$k(\omega) = \exp\left(-\frac{c_0\|\omega\|^2}{4}\right), \qquad (c_0 > 0)$$

which lead to

$$C(h, u) \propto \frac{1}{(u^2 + c_0)^{d/2}} \exp\left(-\frac{\|h\|^2}{u^2 + c_0}\right) \exp\left(-\delta u^2\right). \tag{9}$$

The condition $\delta > 0$ is needed to ensure the condition (C1) is satisfied at $\omega = 0$, but the limit of (9) as $\delta \to 0$ is also a valid spatial-temporal covariance function, leading to the three parameter family

$$C(h, u) = \frac{\sigma^2}{(a^2u^2 + 1)^{d/2}} \exp\left(-\frac{b^2\|h\|^2}{a^2u^2 + 1}\right).$$

Cressie and Huang's approach is novel and powerful but depends on Fourier transform pairs in $\mathbf{r^d}$. Gneiting takes the approach of Cressie et al. and provides a very general class of valid spatial-temporal covariance models. The key result can be formulated as follows.

Let $\psi(t)$, $t \geq 0$, be a completely monotone function[a], and let $\phi(t)$, $t \geq 0$, be a positive function with a completely monotone derivative. Then

$$C(h,u) = \frac{\sigma^2}{\phi(|u|^2)^{d/2}} \psi\left(\frac{\|h\|^2}{\phi(|u|^2)}\right), \tag{10}$$

is a space-time covariance function, where $h \in \mathbf{r}^d$ represents a d-dimensional spatial vector and $u \in \mathbf{r}$ is a scalar time component.

[a] A continuous function $\psi(t)$, defined for $t \geq 0$, is said to be completely monotone if it possesses derivatives $\psi^{(n)}$ of all orders and $(-1)^n \psi^{(n)}(t) \geq 0$ for $t > 0$ and $n = 0, 1, 2, \cdots$.

For example, putting $\psi(t) = \exp(-ct^\gamma)$ and $\phi(t) = (at^\alpha + 1)^\beta$ in (10) leads to

$$C(h,u) = \frac{\sigma^2}{(a|u|^{2\alpha}+1)^{\beta d/2}} \exp\left(-\frac{c\|h\|^{2\gamma}}{(a|u|^{2\alpha}+1)^{\beta\gamma}}\right),$$

where $(h,u) \in \mathbf{r}^d \times \mathbf{r}$. The product with the purely temporal covariance function $\left(a|u|^{2\alpha}+1\right)^{-\delta}$, $u \in \mathbf{r}$, then gives the class

$$C(h,u) = \frac{\sigma^2}{(a|u|^{2\alpha}+1)^{\delta+\beta d/2}} \exp\left(-\frac{c\|h\|^{2\gamma}}{(a|u|^{2\alpha}+1)^{\beta\gamma}}\right),$$

where $(h,u) \in \mathbf{r}^d \times \mathbf{r}$. a and c are nonnegative scaling parameters of time and space, respectively; the smoothness parameters α and γ take values in $(0,1]$; $\beta \in [0,1]$, $\delta \geq 0$ and $\sigma^2 > 0$. A separable covariance function is obtained when $\beta = 0$.

Dynamic spatiotemporal models

Dynamic linear models are often referred to as state-space models in the time series literature. Let Y_t be a $m \times 1$ vector of observations at time t. θ_t is called the sate vector. Y is related to θ through the MEASUREMENT EQUATION. In general we do not observe θ. We have the following framework:

MEASUREMENT EQUATION:

$$Y_t = F_t \theta_t + \epsilon_t$$

where $\epsilon_t \sim N(0, \Sigma_t^{\epsilon})$.

TRANSITION EQUATION:

$$\theta_t = G_t \theta_{t-1} + \eta_t$$

where $\beta_t \sim N(0, \Sigma_t^{\eta})$. F_t and G_t are $m \times p$ and $p \times p$ matrices, they are called the system matrices and they might change over time.

We can compute the association accross time:

$$Cov(\theta_t, \theta_{t-1}) = G_t Var(\theta_{t-1})$$

and

$$Cov(Y_t, Y_{t-1}) = F_t G_t Var(\theta_{t-1}) F_{t-1}^T.$$

Formulation for spatiotemporal models

$$Y(s,t) = \mu(s,t) + \epsilon(s,t)$$

where $\epsilon(s,t) \sim N(0, \sigma^2)$.

$$\mu(s,t) = x(s,t)\tilde{\beta}(s,t)$$

$$\tilde{\beta}(s,t) = \beta_t + \beta(s,t)$$

$$\beta_t = \beta_{t-1} + \eta_t$$

where $\eta_t \sim N_p(0, \Sigma_\eta)$ and

$$\beta(s,t) = \beta(s,t-1) + w(s,t)$$

We introduce a model of linear coregionalization for the p-multivariate process $w(s,t)$, i.e.

$$w(s,t) = A\, v(s,t)$$

with $v(s,t) = (v_1(s,t), \ldots, v_p(s,t))$. The v_l are serially independent Gaussian processes with correlation function ρ_l. Thus, the cross-covariance of w is a linear combination of the ρ_l's. The ρ_l's are usually assumed to be separable space-time covariances.

Thus,

$$Cov(w(s,t), w(s',t')) = \sum_{j=1}^{p} \rho_j(s-s', t-t') T_j$$

where $T_j = a_j a_i$, and a_k the kth column of A.

A Bayesian hierarchical model may be complete by prior specification, for example:

β_0 has a normal prior, and $\beta(,0) \equiv 0$.

Σ_η, Σ_w have IW (inverse Wishart), σ^2 has inverse Gamma.

Instead of Σ_w having an IW, we could use the coregionalization model, with T having an IW prior.