Multivariate Spatial Modeling

There are many situations where we observe a multivariate response that is space referenced. We expected that there will be spatial association for each component of the process as well as dependencies between the different components. Thus, there are two dependence structures that need to be modeled. This can be achieved separately for each producing a **separable process** or not.

There are many situations where we observe a multivariate response that is space referenced. We expected that there will be spatial association for each component of the process as well as dependencies between the different components. Thus, there are two dependence structures that need to be modeled. This can be achieved separately for each producing a **separable process** or not.

We denote the process as $Y(s)$, where $s \in S \subset \mathbb{R}^d$ and $Y(s) \in \mathbb{R}^p$. We describe the process through the joint distributions of any finite collection of locations in S.

For a multivariate Gaussian process we need to specify the **cross-covariance function**

$$C(s, s') = \text{cov}(Y(s), Y(s')) \in \mathbb{R}^{p \times p}$$

As $\text{cov}(Y_j(s), Y_{j'}(s'))$ need not be equal to $\text{cov}(Y_{j'}(s), Y_j(s'))$, $C(s, s')$ need not be a symmetric matrix. Of course $C(s, s)$ is the covariance matrix of the vector $Y(s)$, which is symmetric.

Multivariate Gaussian Processes

For a multivariate Gaussian process we need to specify the **cross-covariance function**

$$C(s, s') = \text{cov}(Y(s), Y(s')) \in \mathbb{R}^{p \times p}$$

As $\text{cov}(Y_j(s), Y_{j'}(s'))$ need not be equal to $\text{cov}(Y_{j'}(s), Y_j(s'))$, $C(s, s')$ need not be a symmetric matrix. Of course $C(s, s)$ is the covariance matrix of the vector $Y(s)$, which is symmetric.

$C(s, s')$ is a valid cross-covariance function if for an arbitrary number and choice of n locations in S, the resulting $np \times np$ matrix is positive definite.

Multivariate Gaussian Processes

For a multivariate Gaussian process we need to specify the **cross-covariance function**

$$C(s, s') = \text{cov}(Y(s), Y(s')) \in \mathbb{R}^{p \times p}$$

As $\text{cov}(Y_j(s), Y_{j'}(s'))$ need not be equal to $\text{cov}(Y_{j'}(s), Y_j(s'))$, $C(s, s')$ need not be a symmetric matrix. Of course $C(s, s)$ is the covariance matrix of the vector $Y(s)$, which is symmetric.

$C(s, s')$ is a valid cross-covariance function if for an arbitrary number and choice of n locations in S, the resulting $np \times np$ matrix is positive definite.

Stationarity and isotropy are achieved if $C(s, s')$ depends, respectively, only on $s - s'$ and $||s - s'||$.

Hierarchical Representation

The previous model can be represented hierarchical with the observation equation

$$Y(s_i) \sim N_p(\mu(s_i) + v(s_i), D), \quad i = 1, \ldots, n$$

and the process equation

$$v = (v(s_1) \ldots v(s_n))' \sim N\left(0, \sum_{j=1}^{p} R_j \otimes T_j\right)$$

where $[R_j]_{ik} = \rho_j(s_i - s_k)$.

Integrating v out we have

$$p(Y|\beta, D, \rho, T) = N\left(\mu, \sum_{j=1}^{p} R_j \otimes T_j + I_n \otimes D\right)$$

We can write

$$p(v(s)) = p(v_1(s))p(v_2(s)|v_1(s))\dots p(v_p(s)|v_1(s),\dots,v_{p-1}(s))$$

In the case $p = 2$ we have

$$v_1(s) = a_{11}w_1(s), \quad p(v_2(s)|v_1(s)) = N\left(\frac{a_{21}}{a_1 1}v_1(s), a_{22}^2\right)$$

There is a big computational advantage in working with the conditional parameterization. Instead of doing calculations with a $np \times np$ covariance matrix, we need to handle p covariance matrices of dimensions $n \times n$. In the presence of additive white noise and general covariates the conditional and multivariate formulations of the model are not equivalent.

Multivariate Predictive Process

The predictive Gaussian process extends to the multivariate case, using a multivariate parent process w at m locations, as

$$\tilde{w}(s)\mathrm{cov}(w(s), w^*)\mathrm{var}^{-1}(w^*)w^* = c(s)C^{-1}w^*.$$

The resulting cross-covariance function is

$$\Gamma_{\tilde{w}}(s, s') = c(s)C^{-1}c(s')$$

If the parent process is obtained via coregionalization then its cross-covariance function is

$$\Gamma_w(s, s') = A\mathrm{Diag}(\rho_j(s, s'))A'.$$

The resulting covariance matrix for $\tilde{w}$ is

$$\mathrm{var}(\tilde{w}) = (I_n \otimes A)\tilde{\Sigma}\Sigma^{-1}\tilde{\Sigma}'(I_m \otimes A')$$

where

$$\Sigma = \mathrm{BD}(\rho_j(s_i^*, s_j^*)) \in \mathbb{R}^{mp \times mp}$$

and

$$\tilde{\Sigma} = \mathrm{Diag}(\rho(s_i, s_j^* a)) \in \mathbb{R}^{np \times mp}$$

The Kronecker product and the block diagonal structure allow for fast calculations of the $np \times np$ covariance matrix.

A modified version of the multivariate predictive process, to account for short scale variability, can be obtained in a similar manner to the univariate case.

Spatially-Varying Coregionalization

A non-stationary extension of the model of coregionalization is

$$Y(s) = A(s)w(s),$$

that corresponds to

$$T(s) = A(s)A(s)', \quad \text{with} \quad c(s, s') = \sum_j \rho_j(s, s')a_j(s)a_j(s').$$

We can make the space variability dependent on a covariate $x(s)$ by letting $T(s) = g(x(s))T$. Alternatively, we can define a Wishart process with ν degrees of freedom and scale matrix $\Gamma\Gamma'$ as

$$T(s) = \Gamma Z(s)Z(s)'\Gamma'$$

where $Z(s)$ is a $p \times \nu$ matrix whose columns are ν independent replicates of vectors with independent Gaussian process with correlations $\tau_j, j = 1, \ldots, p$.

Separable Models

The simplest way of specifying a valid cross-covariance is to take a valid spatial correlation function $\rho(s, s')$ and a valid covariance matrix $T \in \mathbb{R}^{p \times p}$. Then

$$C(s, s') = \rho(s, s')T$$

this is a **separable** model.

For a set of locations $s_1, \ldots, s_n$ the resulting covariance matrix for $(Y(s_1), \ldots, Y(s_n))$ is

$$\Sigma = R \otimes T = \begin{bmatrix} \rho(s_1, s_1)T & \cdots & \rho(s_1, s_n)T \\ \vdots & \vdots & \vdots \\ \rho(s_n, s_1)T & \cdots & \rho(s_n, s_n)T \end{bmatrix} \in \mathbb{R}^{np \times np}$$

We notice that $\Sigma^{-1} = R^{-1} \otimes T^{-1}$ and $|\Sigma| = |R|^p |T|^n$

We can write the joint density of $Y = (Y(s_1), \ldots, Y(s_n))$ as

$$|R|^{p/2}|T|^{n/2} \exp\left\{-\frac{1}{2}\mathrm{tr}\left(T^{-1}YR^{-1}Y'\right)\right\}$$

avoiding computations for $np \times np$ matrices

We can write the joint density of $Y = (Y(s_1), \ldots, Y(s_n))$ as

$$|R|^{p/2}|T|^{n/2}\exp\left\{-\frac{1}{2}\mathrm{tr}\left(T^{-1}YR^{-1}Y'\right)\right\}$$

avoiding computations for $np \times np$ matrices

A separable $C(s, s')$ is symmetric. Moreover, for a stationary correlation function ρ we have that

$$\frac{\mathrm{cov}(Y_l(s), Y_{l'}(s+h))}{\sqrt{\mathrm{cov}(Y_l(s), Y_l(s+h))\mathrm{cov}(Y_{l'}(s), Y_{l'}(s+h))}} = \frac{T_{ll'}}{\sqrt{T_{ll}T_{l'l'}}}$$

regardless of s and h. Also, if ρ is an isotropic correlation function, then the spatial range will be identical for all the components of Y.

These three features are undesirable consequences of the separability assumption

Suppose we have a spatial covariate $X(s)$ and a spatial response $Y(s)$ and we want to predict the value of $Y(s)$ at a location where $X(s)$ has been observed. We assume that

$$\begin{pmatrix} X(s) \\ Y(s) \end{pmatrix} \sim N(\mu(s), T), \ \ \forall s \in S$$

Suppose for simplicity that $\mu(s) = (\mu_1, \mu_2)'$. We are interested in the posterior distribution of the regressor $E(Y(s)|x(s))$, given the observations of Y and X.

we have that $p(y(s)|x(s), \beta_0, \beta_1, \sigma^2) = N(\beta_0 + \beta_1 x(s), \sigma^2)$ where

$$\beta_0 = \mu_2 - \frac{T_{12}}{T_{11}}\mu_1, \ \ \beta_1 = \frac{T_{12}}{T_{11}}, \ \ \sigma^2 = T_{22} - \frac{T_{12}^2}{T_{11}}$$

We fit the separable model

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N_{np}(\mu, R(\phi) \otimes T)$$

using an inverse Wishart prior on T. This will produce posterior samples for μ_1, μ_2, ϕ and T. From these we can obtain samples of $E(Y(s)|x(s))$.

When the value of $x(s)$ is not observed we can use the spatial structure to obtain samples of it. In practice we introduce a latent variable that is included in the sampling scheme.

Consider the process $Y(s) = Aw(s)$ where the components of $w(s)$ are i.i.d. stationary spatial processes with mean 0, variance 1 and correlation function $\rho(h)$. The $E(Y(s)) = 0$ and $C_Y(s, s') = \rho(s - s')AA'$. So, letting $AA' = T$ we obtain the separable model. Without loss of generality we assume that A is lower triangular.

Consider the process $Y(s) = Aw(s)$ where the components of $w(s)$ are i.i.d. stationary spatial processes with mean 0, variance 1 and correlation function $\rho(h)$. The $E(Y(s)) = 0$ and $C_Y(s, s') = \rho(s - s')AA'$. So, letting $AA' = T$ we obtain the separable model. Without loss of generality we assume that A is lower triangular.

More generally, let $w_j(s)$ have mean μ_j, variance 1 and correlation function $\rho_j(h)$. Then $E(Y(s)) = A\mu$ and

$$C_Y(s, s') = \sum_{j=1}^{p} \rho_j(s - s')T_j$$

where $T_j = a_j a_j'$, with a_j the j-th column of A. $\sum_j T_j = T$. This model corresponds to a stationary, non-separable process.

A General Multivariate Spatial Model

We can use the corregionalization idea within a more general framework. Thus

$$Y(s) = \mu(s) + v(s) + \varepsilon(s)$$

where

$$\mu_j(s) = X_j'\beta_j, \;\; v(s) = Aw(s), \;\; E(w_j(s)) = 0, \;\; \varepsilon(s) \sim N(0, \mathrm{diag}(\tau_j^2))$$

The lower triangular structure of the matrix A allows for a representation of the distribution of $v(s)$ as the product univariate conditional distributions.

Hierarchical Representation

The previous model can be represented hierarchical with the observation equation

$$Y(s_i) \sim N_p(\mu(s_i) + v(s_i), D), \quad i = 1, \ldots, n$$

and the process equation

$$v = (v(s_1) \ldots v(s_n))' \sim N\left(0, \sum_{j=1}^{p} R_j \otimes T_j\right)$$

where $[R_j]_{ik} = \rho_j(s_i - s_k)$.

Integrating v out we have

$$p(Y|\beta, D, \rho, T) = N\left(\mu, \sum_{j=1}^{p} R_j \otimes T_j + I_n \otimes D\right)$$

We can write

$$p(v(s)) = p(v_1(s))p(v_2(s)|v_1(s)) \dots p(v_p(s)|v_1(s), \dots, v_{p-1}(s))$$

In the case $p = 2$ we have

$$v_1(s) = a_{11}w_1(s), \quad p(v_2(s)|v_1(s)) = N\left(\frac{a_{21}}{a_1 1}v_1(s), a_{22}^2\right)$$

There is a big computational advantage in working with the conditional parameterization. Instead of doing calculations with a $np \times np$ covariance matrix, we need to handle p covariance matrices of dimensions $n \times n$. In the presence of additive white noise and general covariates the conditional and multivariate formulations of the model are not equivalent.

Multivariate Predictive Process

The predictive Gaussian process extends to the multivariate case, using a multivariate parent process w at m locations, as

$$\tilde{w}(s)\text{cov}(w(s), w^*)\text{var}^{-1}(w^*)w^* = c(s)C^{-1}w^*.$$

The resulting cross-covariance function is

$$\Gamma_{\tilde{w}}(s, s') = c(s)C^{-1}c(s')$$

If the parent process is obtained via coregionalization then its cross-covariance function is

$$\Gamma_w(s, s') = A\text{Diag}(\rho_j(s, s'))A'.$$

The resulting covariance matrix for $\tilde{w}$ is

$$\mathrm{var}(\tilde{w}) = (I_n \otimes A)\tilde{\Sigma}\Sigma^{-1}\tilde{\Sigma}'(I_m \otimes A')$$

where

$$\Sigma = \mathrm{BD}(\rho_j(s_i^*, s_j^*)) \in \mathbb{R}^{mp \times mp}$$

and

$$\tilde{\Sigma} = \mathrm{Diag}(\rho(s_i, s_j^* a)) \in \mathbb{R}^{np \times mp}$$

The Kronecker product and the block diagonal structure allow for fast calculations of the $np \times np$ covariance matrix.

A modified version of the multivariate predictive process, to account for short scale variability, can be obtained in a similar manner to the univariate case.

Spatially-Varying Coregionalization

A non-stationary extension of the model of coregionalization is

$$Y(s) = A(s)w(s),$$

that corresponds to

$$T(s) = A(s)A(s)', \quad \text{with} \quad c(s, s') = \sum_j \rho_j(s, s')a_j(s)a_j(s').$$

We can make the space variability dependent on a covariate $x(s)$ by letting $T(s) = g(x(s))T$. Alternatively, we can define a Wishart process with ν degrees of freedom and scale matrix $\Gamma\Gamma'$ as

$$T(s) = \Gamma Z(s)Z(s)'\Gamma'$$

where $Z(s)$ is a $p \times \nu$ matrix whose columns are ν independent replicates of vectors with independent Gaussian process with correlations $\tau_j, j = 1, \ldots, p$.