5

Estimation of the Separable MGMRF Parameters for Thematic Classification

Rolando D. Navarro, Jr., Joselito C. Magadia and Enrico C. Paringit
University of the Philippines, Diliman, Quezon City
Philippines

1. Introduction

Because of its ability to describe interdependence between neighboring sites, the Markov Random Field (MRF) is a very attractive model in characterizing correlated observations (Moura and Balram, 1993) and it has potential applications in areas of remote sensing, such as spatio-temporal modeling and machine vision. In this study, we model image random field conditional to the texture label as a Multivariate Gauss Markov Random Field (MGMRF); whereas; the thematic map is modeled as a discrete label MRF (Li, 1995). The observations in the Gauss Markov Random Field (GMRF) are distributed with the Gaussian distribution.

There are some MGMRF models where the interaction matrices are modeled in some simplified form, including the MGMRF with isotropic interaction matrix which we shall refer here as Hazel's GMRF (Hazel, 2000), The MGFMRF with anisotropic interaction matrix proportional to the identity matrix which we shall refer here as Rellier's GMRF (Rellier et. al., 2004), and the Gaussian Symmetric Clustering (GSC) (Hazel, 2000).

From these developments, the model for anisotropic GMRF was generalized and its parameter estimator for an arbitrary neighborhood system is characterized (Navarro et al., 2009). Using our model, the classification performance was analyzed and compared with the GMRF models in literature.

Spectral classes are explored in segmenting image random field models to be able to extract the spatial, spectral, and temporal information. A special case is addressed when the observation includes spectral and temporal information known as the spectro-temporal observation. With respect to the spectral and temporal dimensions, the separability structure is considered based on the Kronecker tensor product of the GMRF model parameters. Separable parameters contain less parameters, compared with its non-separable counterpart. In addition, the spectral and temporal dimensions on a separable model can be analyzed separately. We analyzed whether the separability of the GMRF parameters would improve the classification of the thematic map.

2. Image random field modelling and thematic classification

This section covers statistical background in characterizing random fields based on the MRF. Then, we will present estimation for the thematic map and image random field parameters.

Finally the thematic map classifier is presented based on the Iterated Conditional Modes (ICM) algorithm.

2.1 Markov random fields

A random field $\mathbf{Z} = \{\mathbf{Z_s} : \mathbf{s} \in \mathcal{S}\}$ where $\mathbf{s}$ is a site on the lattice $\mathcal{S}$ with the neighborhood system ∂ with parameter $\mathbf{\Pi}$ is a MRF if for $\mathbf{s} \in \mathcal{S}$ (Winkler, 2003).

$$p\left(\mathbf{Z_s} \middle| \mathbf{Z}_{\mathcal{S}/\mathbf{s}}; \mathbf{\Pi}\right) = p\left(\mathbf{Z_s} \middle| \mathbf{Z}_{\partial \mathbf{s}}; \mathbf{\Pi}\right) \tag{1}$$

where $\mathbf{Z}_{\partial \mathbf{s}} = \{\mathbf{Z_t} : \mathbf{t} \in \partial \mathbf{s}\}$ is the random field which consists of observations of the neighbors of $\mathbf{s}$. Similarly, $\mathbf{Z}_{\mathcal{S}/\mathbf{s}} = \{\mathbf{Z_t} : \mathbf{t} \in \mathcal{S}/\mathbf{s}\}$ is the random field, which consists of observations that exclude $\mathbf{s}$.

2.2 Thematic map modeling

Let $L = \{L_\mathbf{s}\}_{\mathbf{s}\in\mathcal{S}}$ be denoted as the thematic map, where $L_\mathbf{s} \in \{1, \cdots, M\}$ is the labeled thematic class at site $\mathbf{s}$ and M is the number of thematic classes. The thematic map is modeled as a discrete space, discrete domain MRF with parameters $\boldsymbol{\varphi} = \left\{\{a_m\}_{1\le m\le M}, \{b_\mathbf{r}\}_{\mathbf{r}\in\mathcal{N}}\right\}$ where a_m is the singleton potential coefficient for the m^{th} thematic class, $b_\mathbf{r}$ are made up by the pairwise potential coefficients, and $\mathcal{N}$ is region of support (Jeng & Woods, 1991) or the neighborhood set (Kasyap & Chellappa, 1983). Its conditional probability density function (pdf) is given by

$$p\left(L_\mathbf{s} \middle| \mathbf{L}_{\partial\mathbf{s}}, \boldsymbol{\varphi}\right) = \frac{\exp\left(\sum_{m=1}^{M} a_m \mathbf{1}_{\{\mathbf{L_s}=m\}} + \sum_{\mathbf{r}\in\mathcal{N}} b_\mathbf{r} \cdot V\left(L_\mathbf{s}, L_{\mathbf{s}-\mathbf{r}}\right)\right)}{\sum_{l=1}^{M} \exp\left(a_l + \sum_{\mathbf{r}\in\mathcal{N}} b_\mathbf{r} \cdot V\left(L_\mathbf{s} = l, L_{\mathbf{s}-\mathbf{r}}\right)\right)} \tag{2}$$

(Li, 1995), where

$$V(x,y) = \begin{cases} 1 & x=y \\ -1 & x\neq y. \end{cases}$$

2.3 Image random field modeling

The observation $\mathbf{Y_s}$ given the thematic map $\mathbf{L}$ is modeled with the conditional distribution $\mathbf{Y_s} \middle| \mathbf{L} \sim N_N\left(\boldsymbol{\mu}(L_\mathbf{s}), \boldsymbol{\Sigma}(L_\mathbf{s})\right)$. It is conditionally dependent on $L_\mathbf{s}$, the thematic class at site $\mathbf{s}$, and it is driven by an autoregressive Gaussian colored noise process $\mathbf{X_s} \middle| \mathbf{L} \sim N_N\left(\mathbf{0}_{N\times1}, \boldsymbol{\Sigma}(L_\mathbf{s})\right)$. Two noise processes $\mathbf{X_s}$ and $\mathbf{X_{s-r}}$ are statistically independent if the corresponding thematic classes $L_\mathbf{s}$ and $L_{\mathbf{s}-\mathbf{r}}$ are different for all $\mathbf{r} \in \mathcal{N}$ and $\mathbf{s} \in \mathcal{S}$. This model tends to avoid the blurring effect created between segment boundaries which, in turn, may yield poor classification performance. The resulting equation can be written as follows:

$$\mathbf{X_s} = \left(\mathbf{Y_s} - \boldsymbol{\mu}(L_\mathbf{s})\right) - \sum_{\mathbf{r}\in\mathcal{N}} \boldsymbol{\theta}_\mathbf{r}(L_\mathbf{s}) \mathbf{1}_{\{L_\mathbf{s}=L_{\mathbf{s}-\mathbf{r}}\}} \left(\mathbf{Y_{s-r}} - \boldsymbol{\mu}(L_\mathbf{s})\right). \tag{3}$$

The noise process has the following characterization:

$$E\left[\mathbf{X_s}\middle|\mathbf{L};\mathbf{\Theta}\right]=\mathbf{0}_{N\times 1} \tag{4}$$

$$\operatorname{cov}\left(\mathbf{X_s},\mathbf{X_{s-r}}\middle|\mathbf{L};\mathbf{\Theta}\right)=\begin{cases}\mathbf{\Sigma}(L_\mathbf{s}) & \mathbf{r}=\mathbf{0}_{p\times 1}\\ -\boldsymbol{\theta}_\mathbf{r}(L_\mathbf{s})\mathbf{\Sigma}(L_\mathbf{s})\mathbf{1}_{\{L_\mathbf{s}=L_{\mathbf{s-r}}\}} & \mathbf{r}\in\mathcal{N}\\ \mathbf{0}_{N\times N} & \text{otherwise}\end{cases} \tag{5}$$

$$\operatorname{cov}\left(\mathbf{X_s},\mathbf{Y_{s-r}}\middle|\mathbf{L};\mathbf{\Theta}\right)=\mathbf{\Sigma}(L_\mathbf{s})\cdot\mathbf{1}_{\{\mathbf{r}=\mathbf{0}_{p\times 1}\}}. \tag{6}$$

The conditional probability on the other hand is given as

$$p\left(\mathbf{Y_s}\middle|\mathbf{Y}_{\partial\mathbf{s}},\mathbf{L};\mathbf{\Theta}\right)=\frac{1}{(2\pi)^{N/2}\left|\mathbf{\Sigma}(L_\mathbf{s})\right|^{1/2}}\exp\left(-\frac{1}{2}\mathbf{X_s}^T\mathbf{\Sigma}^{-1}(L_\mathbf{s})\mathbf{X_s}\right). \tag{7}$$

2.4 Maximum pseudo-likelihood estimation

The maximum pseudo-likelihood estimation (MPLE) combines sites to form the pseudo-likelihood function from the conditional probabilities (Li, 1995). The pseudo-likelihood functions for the thematic map random field and image random field parameters are given as follows:

$$PL(\boldsymbol{\varphi})=\prod_{\mathbf{s}\in\mathcal{S}}p\left(L_\mathbf{s}\middle|\mathbf{L}_{\partial\mathbf{s}};\boldsymbol{\varphi}\right) \tag{8}$$

$$PL\left(\mathbf{\Theta}\middle|\mathbf{L}\right)=\prod_{m=1}^{M}\prod_{\mathbf{s}\in\mathcal{S}(m)}p\left(\mathbf{Y_s}\middle|\mathbf{Y}_{\partial\mathbf{s}},\mathbf{L};\mathbf{\Theta}\right) \tag{9}$$

where $\mathcal{S}(m)$ is the collection of sites with the m^{th} thematic class. The MPLE possesses an invariance property, that is, if $\hat{\mathbf{\Pi}}$ is the MPLE of the parameter $\mathbf{\Pi}$, then for an arbitrary function τ, $\tau(\hat{\mathbf{\Pi}})$ is the MPLE of the parameter $\tau(\mathbf{\Pi})$. The proof is similar to that of the invariance property of the MLE (Casella and Berger, 2002) since the form of the pseudo-likelihood function is analogous that of the likelihood function, depending on the parameter given the data. Moreover, the MPLE converges to the MLE almost surely as the lattice size approaches infinity (Geman and Greffigne, 1987).

2.5 Thematic classification

The thematic map can be recovered by the maximum a posteriori probability (MAP) rule. It can be implemented using a numerical optimization technique such as Simulated Annealing (SA) (Jeng & Woods, 1991). Although the global convergence employing SA is guaranteed almost surely, its convergence is very slow (Aarts & Korts, 1987; Winkler, 2006). An alternative to this is to use the ICM algorithm (Besag, 1986) given as

$$\hat{L}_\mathbf{s}=\arg\max_{1\le m\le M}p\left(\mathbf{Y}\middle|L_\mathbf{s}=m,\mathbf{L}_{\mathcal{S}/\mathbf{s}};\mathbf{\Theta}\right)p\left(L_\mathbf{s}=m\middle|\mathbf{L}_{\mathcal{S}/\mathbf{s}};\boldsymbol{\varphi}\right). \tag{10}$$

This is interpreted as the instantaneous freezing of the annealing schedule of the SA. However, since $p(\mathbf{Y}|\mathbf{L};\boldsymbol{\Theta})$ is difficult to evaluate, alternatively, it is replaced by its pseudo-likelihood (Hazel, 2000) given as

$$p(\mathbf{Y}|\mathbf{L};\boldsymbol{\Theta}) \approx \prod_{\mathbf{s}\in\mathcal{S}} p(\mathbf{Y}_\mathbf{s}|\mathbf{Y}_{\partial\mathbf{s}},\mathbf{L};\boldsymbol{\Theta}). \tag{11}$$

Hence, the classifier is reduced to

$$\hat{L}_\mathbf{s} = \arg\max_{1\le m\le M} \prod_{\mathbf{s}\in\mathcal{S}} p(\mathbf{Y}_\mathbf{s}|\mathbf{Y}_{\partial\mathbf{s}},\mathbf{L};\boldsymbol{\Theta})\cdot p(L_\mathbf{s}=m|\mathbf{L}_{\mathcal{S}/\mathbf{s}};\boldsymbol{\varphi}). \tag{12}$$

The ICM algorithm, unlike the SA, is only guaranteed to converge to the local maxima. This problem can be alleviated by initializing the thematic map from the Gaussian Spectral Clustering (GSC) model (Hazel, 2000).

2.6 Numerical implementation

The MPLE-based estimators are not in their closed form and must be evaluated numerically. The pseudocode for estimating the parameters is presented below.

Initialize $\mathbf{L}$, $\boldsymbol{\varphi}$, and $\boldsymbol{\Theta}$
Estimate $\boldsymbol{\varphi}$
Estimate $\boldsymbol{\Theta}$
    Estimate $\boldsymbol{\mu}(m)$ given $\boldsymbol{\theta}_\mathrm{r}(m)$ and $\boldsymbol{\Sigma}(m)$
    Estimate $\boldsymbol{\theta}_\mathrm{r}(m)$ given $\boldsymbol{\Sigma}(m)$ and $\boldsymbol{\mu}(m)$
    Estimate $\boldsymbol{\Sigma}(m)$ given $\boldsymbol{\mu}(m)$ and $\boldsymbol{\theta}_\mathrm{r}(m)$
Estimate L_s by the ICM Algorithm

The image random field parameters are estimated using a method with some resemblance to the Gauss-Seidel iteration method (Kreyzig, 1993). The convergence criterion for estimating these parameters using this iteration method has yet to be established. As a precautionary measure, a single iteration was performed. This method was also applied in estimating the image random field estimators in Rellier's GMRF (Rellier, et. al., 2004).

3. Spectro-temporal MGMRF modelling

The spectro-temporal observation image random field will be characterized with hybrid separable MGMRF parameters.

3.1 Image random field modeling

We let M_1 - number of lines, M_2 - number of samples, N_1 - number of spectral bands, and N_2 - number of temporal slots. The image random field is characterized as follows:

Lattice $\mathcal{S} = \{(s_1, s_2):\ 1\le s_1 \le M_1,\ 1\le s_2\le M_2\}$

Thematic Class $L_\mathrm{s} = L_{(s_1,s_2)}$

The thematic class $L_{\mathbf{s}}$ at a given site $\mathbf{s} \in \mathcal{S}$ is modeled to be fixed over time.

Observation $$\mathbf{Y}_{\mathbf{s}} = \mathbf{Y}_{(s_1,s_2)} = \left(W_{(s_1,s_2,1,1)} \quad \cdots \quad W_{(s_1,s_2,k,l)} \quad \cdots \quad W_{(s_1,s_2,N_1,N_2)} \right)^T$$

The observation $\mathbf{Y}_{\mathbf{s}}$ is a multispectral and mono-temporal vector of reflectance of the given spatial location (s_1, s_2) measured at the k^{th} spectral band with wavelength $\lambda = \lambda_k$, for $1 \le k \le N_1$, and at the l^{th} temporal slot with time $T = T_l$ for $1 \le l \le N_2$. More specifically, the $\left(k + (l-1)N_1\right)^{th}$ element of $\mathbf{Y}_{\mathbf{s}}$ denoted as $Y_{\mathbf{s},(k,l)}$ is given as $Y_{\mathbf{s},(k,l)} = W_{(s_1,s_2,k,l)}$.

Let us consider the matrix $\mathbf{Y}_{\mathbf{s}}^{\#}$ defined by rearranging the elements of the spectro-temporal observation $\mathbf{Y}_{\mathbf{s}}$ with the reshape operator $\mathbf{Y}_{\mathbf{s}}^{\#} = reshape\left(\mathbf{Y}_{\mathbf{s}}, N_1, N_2\right)$. The reshape function given as $\mathbf{B} = reshape\left(\mathbf{A}, N_1, N_2\right)$ transforms the vector $\mathbf{A} = \left\{a_k\right\} \in \mathbb{R}^{N_1 N_2}$ into the $N_1 \times N_2$ matrix $\mathbf{B} = \left\{b_{ij}\right\} \in \mathbb{R}^{N_1 \times N_2}$ by the mapping $b_{ij} = a_{k=i+(j-1)N_1}$ for all $1 \le i \le N_1$ and $1 \le j \le N_2$, i.e.

$$\mathbf{Y}_{\mathbf{s}}^{\#} = \begin{bmatrix} Y_{\mathbf{s},(1,1)} & Y_{\mathbf{s},(1,2)} & \cdots & Y_{\mathbf{s},(1,N_2)} \\ Y_{\mathbf{s},(2,1)} & Y_{\mathbf{s},(2,2)} & \cdots & Y_{\mathbf{s},(2,N_2)} \\ \cdots & \cdots & \cdots & \cdots \\ Y_{\mathbf{s},(N_1,1)} & Y_{\mathbf{s},(N_1,2)} & \cdots & Y_{\mathbf{s},(N_1,N_2)} \end{bmatrix}. \tag{13}$$

The matrix $\mathbf{Y}_{\mathbf{s}}^{\#}$ is characterized by allocating the reflectance across the bands for a given time by column and the reflectance across time for a given band by row.

3.2 Separable structure of the covariance matrix

There is a growing interest in modeling the covariance structure with more than one attribute. For example, in spatio-temporal modeling, the covariance structure of "spatial" and "temporal" attributes is jointly considered (Kyriakidis and Journel, 1999; Huizenga, et. al., 2002). On the other hand, in the area of longitudinal studies the covariance structure between "factors" and "temporal" attributes are jointly considered (Naik and Rao, 2001). Both studies mentioned above considered covariance matrices with a separable structure between these attributes.

In the realm of remote sensing, few studies have been conducted combining the covariance structure involving spectro-temporal attributes. Campbell and Kiiveri demonstrated canonical variates calculations are reduced to simultaneous between-groups and within-group analyses of a linear combination of spectral bands over time, and the analyses of a linear combination of the time over the spectral bands (Campbell and Kiiveri, 1988).

In light of recent literature, we propose to model the GMRF models as applied to remote sensing image processing where the covariance structure of the "spectral" and "temporal" attributes is characterized jointly. The separable covariance structure associated with the matrix Gaussian distribution has been considered.

3.2.1 Non-separable covariance structure

The matrix observation driven by a colored noise and its vectorized distribution, is assumed to be a realization from the process whose conditional form is given by

$\mathbf{X}_{\mathbf{s}} | \mathbf{L} \sim N_N(\mathbf{0}_N, \mathbf{\Sigma}(L_{\mathbf{s}}))$. The covariance matrix $\mathbf{\Sigma}(L_{\mathbf{s}})$ does not have any special structure, except it has to be a positive definite symmetric matrix. This covariance matrix structure referred to as an unpatterned covariance matrix (Dutilleul, 1999). The statistical characterization is similar to the MGMRF discussed in Section 2.3.

3.2.2 Matrix gaussian distribution

Let $\mathbf{X}^{\#}$ be a random matrix distributed as $\mathbf{X}^{\#} \sim N_{m,n}(\mathbf{M}^{\#}, \mathbf{\Xi}^{(1)}, \mathbf{\Xi}^{(2)})$ where $\mathbf{M}^{\#} \in \mathbb{R}^{m\times n}$ is the expectation matrix, $\mathbf{\Xi}^{(1)} \in \mathbb{R}^{m\times m}$ is the covariance matrix across the rows, and $\mathbf{\Xi}^{(2)} \in \mathbb{R}^{n\times n}$ is the covariance matrix across the columns. Hence, the pdf of $\mathbf{X}^{\#}$ is given as

$$p(\mathbf{X}^{\#}) = \frac{1}{(2\pi)^{mn/2} \left|\mathbf{\Xi}^{(1)}\right|^{n/2} \left|\mathbf{\Xi}^{(2)}\right|^{m/2}} \exp\left[-\frac{1}{2} tr\left(\left(\mathbf{\Xi}^{(1)}\right)^{-1} \left(\mathbf{X}^{\#} - \mathbf{M}^{\#}\right)\left(\mathbf{\Xi}^{(2)}\right)^{-1} \left(\mathbf{X}^{\#} - \mathbf{M}^{\#}\right)^T\right)\right] \quad (14)$$

(Arnold, 1981). Also, if we stack the matrix $\mathbf{X}^{\#}$ into the random vector $\mathbf{X} \equiv vec(\mathbf{X}^{\#})$, then $\mathbf{X} \sim N_{mn}(\mathbf{M}, \mathbf{\Xi})$ where $\mathbf{M} = vec(\mathbf{M}^{\#}) \in \mathbb{R}^{mn}$ is the expectation matrix and $\mathbf{\Xi} = \mathbf{\Xi}^{(2)} \otimes \mathbf{\Xi}^{(1)} \in \mathbb{R}^{mn\times mn}$ is the covariance matrix (Arnold, 1981), and its pdf is given as

$$p(\mathbf{X}) = \frac{1}{(2\pi)^{mn/2} |\mathbf{\Xi}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{X} - \mathbf{M})^T \mathbf{\Xi}^{-1} (\mathbf{X} - \mathbf{M})\right]. \quad (15)$$

We model the associated noise process $\mathbf{X}_{\mathbf{s}}^{\#}$ as a matrix Gaussian distribution, i.e. $\mathbf{X}_{\mathbf{s}}^{\#} | \mathbf{L} \sim N_{N_1,N_2}(\mathbf{0}_{N_1\times N_2}, \mathbf{\Sigma}^{(1)}(L_{\mathbf{s}}), \mathbf{\Sigma}^{(2)}(L_{\mathbf{s}}))$ where $\mathbf{\Sigma}^{(1)}(L_{\mathbf{s}}) \in \mathbb{R}^{N_1\times N_1}$ is the covariance matrix across the bands, and $\mathbf{\Sigma}^{(2)}(L_{\mathbf{s}}) \in \mathbb{R}^{N_2\times N_2}$ is the covariance matrix across time. Stacking the matrix $\mathbf{X}_{\mathbf{s}}^{\#}$ into a random vector $\mathbf{X}_{\mathbf{s}} \equiv vec(\mathbf{X}_{\mathbf{s}}^{\#}) \in \mathbb{R}^{N_1 N_2}$ corresponds to the vectorized colored noise with conditional distribution $\mathbf{X}_{\mathbf{s}} | \mathbf{L} \sim N_{N_1 N_2}(\mathbf{0}_{N_1 N_2}, \mathbf{\Sigma}^{(2)}(L_{\mathbf{s}}) \otimes \mathbf{\Sigma}^{(1)}(L_{\mathbf{s}}))$.

3.2.3 Separable covariance structure

The spectro-temporal, separable covariance matrix model (Lu and Zimmerman, 2005; Fuentes, 2006) has the form

$$\mathbf{\Sigma}(m) = \mathbf{\Sigma}^{(2)}(m) \otimes \mathbf{\Sigma}^{(1)}(m) \quad (16)$$

for $1 \le m \le M$ where $\mathbf{\Sigma}^{(1)}(m) = \left\{\sigma_{ij}^{(1)}(m)\right\} \in \mathbb{R}^{N_1\times N_1}$ is the covariance matrix across bands and $\mathbf{\Sigma}^{(2)}(m) = \left\{\sigma_{kl}^{(2)}(m)\right\} \in \mathbb{R}^{N_2\times N_2}$ is the covariance matrix across time. Now, since

$$\mathbf{X}_{\mathbf{s}} | \mathbf{L} \sim N_{N_1 N_2}\left(\mathbf{0}_{N_1\times N_2} \mathbf{\Sigma}^{(2)}(L_{\mathbf{s}}) \otimes \mathbf{\Sigma}^{(1)}(L_{\mathbf{s}})\right) \quad (17)$$

$$\mathbf{Y}_{\mathbf{s}} | \mathbf{L} \sim N_{N_1 N_2}\left(\boldsymbol{\mu}(L_{\mathbf{s}}), \mathbf{\Sigma}^{(2)}(L_{\mathbf{s}}) \otimes \mathbf{\Sigma}^{(1)}(L_{\mathbf{s}})\right) \quad (18)$$

then, the covariance is given as (Arnold, 1981):

$$\mathrm{cov}\left(X_{\mathbf{s},(k,l)}, X_{\mathbf{s},(k,l)} \middle| \mathbf{L};\mathbf{\Theta}\right) = \sigma_{kk}^{(1)}\left(L_{\mathbf{s}}\right)\sigma_{ll}^{(2)}\left(L_{\mathbf{s}}\right) \tag{19}$$

$$\mathrm{cov}\left(Y_{\mathbf{s},(k,l)}, Y_{\mathbf{s},(k,l)} \middle| \mathbf{L};\mathbf{\Theta}\right) = \sigma_{kk}^{(1)}\left(L_{\mathbf{s}}\right)\sigma_{ll}^{(2)}\left(L_{\mathbf{s}}\right) \ . \tag{20}$$

This corresponds to the product of the variance associated with the reflectance at the k^{th} spectral band $\sigma_{kk}^{(1)}\left(L_{\mathbf{s}}\right)$ and the variance associated with the reflectance at the l^{th} temporal slot $\sigma_{ll}^{(2)}\left(L_{\mathbf{s}}\right)$. Likewise, the cross-covariance is given as (Arnold, 1981):

$$\mathrm{cov}\left(X_{\mathbf{s},(i,j)}, X_{\mathbf{s},(k,l)} \middle| \mathbf{L};\mathbf{\Theta}\right) = \sigma_{ik}^{(1)}\left(L_{\mathbf{s}}\right)\sigma_{jl}^{(2)}\left(L_{\mathbf{s}}\right) \tag{21}$$

$$\mathrm{cov}\left(Y_{\mathbf{s},(i,j)}, Y_{\mathbf{s},(k,l)} \middle| \mathbf{L};\mathbf{\Theta}\right) = \sigma_{ik}^{(1)}\left(L_{\mathbf{s}}\right)\sigma_{jl}^{(2)}\left(L_{\mathbf{s}}\right) . \tag{22}$$

This corresponds to the product of the covariance associated with the reflectance at the i^{th} and the k^{th} spectral band $\sigma_{ik}^{(1)}\left(L_{\mathbf{s}}\right)$ and the covariance associated with the reflectance at the j^{th} and the l^{th} temporal slot $\sigma_{jl}^{(2)}\left(L_{\mathbf{s}}\right)$.

The number of parameters in the unpatterned covariance matrix is $N(N+1)/2 = N_1N_2(N_1N_2+1)/2$. On the other hand, the number of parameters for a separable covariance matrix is $\left[N_1(N_1+1)+N_2(N_2+1)\right]/2$, which has fewer parameters compared to its non-separable counterpart.

3.2.4 Separable of interaction matrix structure

We can also model the interaction matrix coefficients with a separable structure for all $\mathbf{r} \in \mathcal{N}$ and $1 \le m \le M$ of the form

$$\mathbf{\theta}_{\mathbf{r}}(m) = \mathbf{\theta}_{\mathbf{r}}^{(2)}(m) \otimes \mathbf{\theta}_{\mathbf{r}}^{(1)}(m) \tag{23}$$

where $\mathbf{\theta}_{\mathbf{r}}^{(1)}(m) \in \mathbb{R}^{N_1 \times N_1}$ is the interaction matrix across the bands and $\mathbf{\theta}_{\mathbf{r}}^{(2)}(m) \in \mathbb{R}^{N_2 \times N_2}$ is the interaction matrix across time. In the next section, the interaction matrix coefficient $\mathbf{\theta}_{\mathbf{r}}(m)$ can be made separable for $\mathbf{r} \in \mathcal{N}$ and $1 \le m \le M$ provided that $\mathbf{\Sigma}(m)$ is separable. Furthermore, if $\mathbf{\Sigma}(m)$ is separable, then the following is the resulting statistical characterization of $\mathbf{X}_{\mathbf{s}}$:

$$E\left[\mathbf{X}_{\mathbf{s}} \middle| \mathbf{L};\mathbf{\Theta}\right] = \mathbf{0}_{N\times 1} \tag{24}$$

$$\mathrm{cov}\left(\mathbf{X}_{\mathbf{s}}, \mathbf{X}_{\mathbf{s}-\mathbf{r}} \middle| \mathbf{L};\mathbf{\Theta}\right) = \begin{cases} \mathbf{\Sigma}^{(2)}\left(L_{\mathbf{s}}\right) \otimes \mathbf{\Sigma}^{(1)}\left(L_{\mathbf{s}}\right) & \mathbf{r} = \mathbf{0}_{p\times 1} \\ -\left(-\mathbf{\theta}_{\mathbf{r}}^{(2)}\left(L_{\mathbf{s}}\right)\mathbf{\Sigma}^{(2)}\left(L_{\mathbf{s}}\right)\cdot \mathbf{1}_{\{L_{\mathbf{s}}=L_{\mathbf{s}-\mathbf{r}}\}}\right) \otimes \left(-\mathbf{\theta}_{\mathbf{r}}^{(1)}\left(L_{\mathbf{s}}\right)\mathbf{\Sigma}^{(1)}\left(L_{\mathbf{s}}\right)\cdot \mathbf{1}_{\{L_{\mathbf{s}}=L_{\mathbf{s}-\mathbf{r}}\}}\right) & \mathbf{r} \in \mathcal{N} \\ \mathbf{0}_{N_2} \otimes \mathbf{0}_{N_1} & \text{otherwise} \end{cases} \tag{25}$$

$$\operatorname{cov}\left(\mathbf{X}_{\mathbf{s}},\mathbf{Y}_{\mathbf{s}-\mathbf{r}}\middle|\mathbf{L};\boldsymbol{\Theta}\right)=\boldsymbol{\Sigma}^{(2)}\left(L_{\mathbf{s}}\right)\cdot\mathbf{1}_{\{L_{\mathbf{s}}=L_{\mathbf{s}-\mathbf{r}}\}}\otimes\boldsymbol{\Sigma}^{(1)}\left(L_{\mathbf{s}}\right)\cdot\mathbf{1}_{\{L_{\mathbf{s}}=L_{\mathbf{s}-\mathbf{r}}\}}. \tag{26}$$

The covariance matrix, from the above equation, $\operatorname{cov}\left(\mathbf{X}_{\mathbf{s}},\mathbf{X}_{\mathbf{s}-\mathbf{r}}\middle|\mathbf{L};\boldsymbol{\Theta}\right)$ has a separable structure between the spectral domain and temporal dimensions. It has a form analogous to that of what is shown in (4) through (6), which is intuitively appealing.

The number of parameters in the unpatterned interaction matrix coefficient is $N^2=N_1^2N_2^2$. On the other hand, the number of parameters for the separable interaction matrix coefficient is $N_1^2+N_2^2$, which has fewer parameters compared to its non-separable counterpart.

3.2.5 Separable mean structure

Likewise, we can also model the mean with a separable structure of the form

$$\boldsymbol{\mu}(m)=\boldsymbol{\mu}^{(2)}(m)\otimes\boldsymbol{\mu}^{(1)}(m) \tag{27}$$

for $1\le m\le M$ where $\boldsymbol{\mu}^{(1)}(m)\in\mathbb{R}^{N_1\times 1}$ is the mean across the bands and $\boldsymbol{\mu}^{(2)}(m)\in\mathbb{R}^{N_2\times 1}$ is the mean across time. The number of parameters in the unpatterned mean vector is $N=N_1N_2$. On the other hand, the number of parameters for the separable mean vector is N_1+N_2 which has fewer number of parameters compared to its non-separable counterpart.

3.2.6 Hybrid separable structure

Finally, we can model the GMRF parameters as having a hybrid separability structure, that is, some of its parameters are separable while the rest are not. Hence, there are eight combinations to consider. As shown in Section 5.2, it is impossible to model a separable interaction matrix with a non-separabable matrix. This leave us six cases to consider in this study.

4. Estimation of thematic map parameters

The MPLE of $\boldsymbol{\varphi}$ is obtained by taking the derivative of $\log PL(\boldsymbol{\varphi})$ with respect to $\{a_m\}_{1\le m\le M}$ and $\{b_{\mathbf{r}}\}_{\mathbf{r}\in\mathcal{N}}$, then equating to zero (Li, 1995). Accordingly, the estimators are obtained numerically by solving the following set of simultaneous nonlinear equations:

$$\sum_{\mathbf{s}\in\mathcal{S}}\frac{\exp\left(a_m+\sum_{\mathbf{r}\in\mathcal{N}}b_{\mathbf{r}}\cdot V\left(L_{\mathbf{s}}=m,L_{\mathbf{s}-\mathbf{r}}\right)\right)}{\sum_{l=1}^{M}\exp\left(a_l+\sum_{\mathbf{r}\in\mathcal{N}}b_{\mathbf{r}}\cdot V\left(L_{\mathbf{s}}=l,L_{\mathbf{s}-\mathbf{r}}\right)\right)}=\sum_{\mathbf{s}\in\mathcal{S}}\mathbf{1}_{\{L_{\mathbf{s}}=m\}}\quad\forall a_m,\ 1\le m\le M \tag{28}$$

$$\sum_{\mathbf{s}\in\mathcal{S}}\frac{\sum_{l=1}^{M}\exp\left(a_l+\sum_{\mathbf{t}\in\mathcal{N}}b_{\mathbf{t}}\cdot V\left(L_{\mathbf{s}}=l,L_{\mathbf{s}-\mathbf{t}}\right)\right)\cdot V\left(L_{\mathbf{s}}=l,L_{\mathbf{s}-\mathbf{r}}\right)}{\sum_{l=1}^{M}\exp\left(a_l+\sum_{\mathbf{t}\in\mathcal{N}}b_{\mathbf{t}}\cdot V\left(L_{\mathbf{s}}=l,L_{\mathbf{s}-\mathbf{t}}\right)\right)}=\sum_{\mathbf{s}\in\mathcal{S}}V\left(L_{\mathbf{s}},L_{\mathbf{s}-\mathbf{r}}\right)\quad\forall b_{\mathbf{r}},\ \mathbf{r}\in\mathcal{N}. \tag{29}$$

5. Important MGMRF specifications

This section provides important characterizations enable us to derive the estimators of the GMRF parameters in the next section. We present a simple, yet powerful, method to derive the MPL estimators of the mean and the interaction matrix. Finally, new problems arise in estimating the multivariate observation GMRFs, which were not encountered in the univariate case, are discussed.

5.1 MPL-based method technique of deriving mean and the interaction matrix estimators

In this section, a method of deriving the MPL estimators for the mean and the vectorized interaction coefficients are presented regardless of separability. The MPL estimator of the interaction matrix coefficients can be derived by taking the matrix derivative of the log of the pseudo-likelihood function with respect to the interaction matrix coefficient or with respect to its vectorized version from the equivalence relation (Neudecker, 1969)

$$\frac{\partial f}{\partial \mathbf{X}} = \mathbf{P} \Leftrightarrow \frac{\partial f}{\partial vec(\mathbf{X})} = vec(\mathbf{P}) \tag{30}$$

where $f(\mathbf{X}) \in \mathbb{R}$, and $\mathbf{X},\ \mathbf{P} \in \mathbb{R}^{m\times n}$. The latter expression is preferred, since it is easier to evaluate. The following proposition provides a simple way of deriving the MPL estimators, where the estimator is either the mean or the vectorized interaction matrix coefficient (Navarro, et. al., 2009).

Proposition 1 Let $\mathbf{\Phi}(m) \in \mathbb{R}^{q\times 1}$, $1 \le m \le M$ be a vector of parameters which is either the mean or the vectorized interaction matrix coefficient. Suppose that $\mathbf{X}_s$ can be expressed in the form

$$\mathbf{X}_\mathbf{s} = \mathbf{P}_\mathbf{s} - \mathbf{Q}_\mathbf{s}\mathbf{\Phi}(L_\mathbf{s}) \tag{31}$$

where $\mathbf{P}_s = \mathbf{P}_s(\mathbf{\Theta}|\mathbf{L}) \in \mathbb{R}^{N\times 1}, \mathbf{Q}_s = \mathbf{Q}_s(\mathbf{\Theta}|\mathbf{L}) \in \mathbb{R}^{N\times q}$ is independent of $\mathbf{\Phi}(L_s)$, and the covariance matrix $\mathbf{\Sigma}(m)$, $1 \le m \le M$ is known, then the MPL estimator for $\mathbf{\Phi}(m)$, $1 \le m \le M$ is obtained by solving the equation

$$\sum_{\mathbf{s}\in\mathcal{S}(m)} \mathbf{Q}_\mathbf{s}^T \mathbf{\Sigma}^{-1}(m)\mathbf{X}_\mathbf{s} = \mathbf{0}_{q\times 1}\ . \tag{32}$$

Proof From (7) and (9), the log pseudo-likelihood of the image random field conditional to the thematic map is given as

$$\log PL(\mathbf{\Theta}|\mathbf{L}) = -\frac{1}{2}\sum_{m=1}^{M}\sum_{\mathbf{s}\in\mathcal{S}(m)}\left[N\log 2\pi + \log\left|\mathbf{\Sigma}(L_\mathbf{s})\right| + \mathbf{X}_\mathbf{s}^T\mathbf{\Sigma}^{-1}(L_\mathbf{s})\mathbf{X}_\mathbf{s}\right]. \tag{33}$$

Taking the gradient of the log pseudo-likelihood function in (33) with respect to $\mathbf{\Phi}(m)$ for $1 \le m \le M$, and equating to $\mathbf{0}_{q\times 1}$ yields

$$\mathbf{0}_{q\times 1}=\frac{\partial}{\partial \mathbf{\Phi}(m)}\log PL(\mathbf{\Theta}|\mathbf{L})=-\frac{1}{2}\sum_{l=1}^{M}\sum_{\mathbf{s}\in\mathcal{S}(l)}\frac{\partial}{\partial \mathbf{\Phi}(m)}\mathbf{X}_{\mathbf{s}}^{T}\mathbf{\Sigma}^{-1}(L_{\mathbf{s}})\mathbf{X}_{\mathbf{s}}. \tag{34}$$

Since

$$\begin{aligned}\mathbf{X}_{\mathbf{s}}^{T}\mathbf{\Sigma}^{-1}(L_{\mathbf{s}})\mathbf{X}_{\mathbf{s}}&=\left(\mathbf{P}_{\mathbf{s}}-\mathbf{Q}_{\mathbf{s}}\mathbf{\Phi}(L_{\mathbf{s}})\right)^{T}\mathbf{\Sigma}^{-1}(L_{\mathbf{s}})\left(\mathbf{P}_{\mathbf{s}}-\mathbf{Q}_{\mathbf{s}}\mathbf{\Phi}(L_{\mathbf{s}})\right)\\&=\mathbf{P}_{\mathbf{s}}^{T}\mathbf{\Sigma}^{-1}(L_{\mathbf{s}})\mathbf{P}_{\mathbf{s}}-2\mathbf{P}_{\mathbf{s}}^{T}\mathbf{\Sigma}^{-1}(L_{\mathbf{s}})\mathbf{Q}_{\mathbf{s}}\mathbf{\Phi}(L_{\mathbf{s}})+\mathbf{\Phi}^{T}(L_{\mathbf{s}})\mathbf{Q}_{\mathbf{s}}^{T}\mathbf{\Sigma}^{-1}(L_{\mathbf{s}})\mathbf{Q}_{\mathbf{s}}\mathbf{\Phi}(L_{\mathbf{s}})\end{aligned} \tag{35}$$

then taking the gradient in (34) with respect to $\mathbf{\Phi}$ yields

$$\begin{aligned}\frac{\partial}{\partial\mathbf{\Phi}}\mathbf{X}_{\mathbf{s}}^{T}\mathbf{\Sigma}^{-1}(L_{\mathbf{s}})\mathbf{X}_{\mathbf{s}}&=2\mathbf{Q}_{\mathbf{s}}^{T}\mathbf{\Sigma}^{-1}(L_{\mathbf{s}})\mathbf{P}_{\mathbf{s}}\mathbf{1}_{\{L_{\mathbf{s}}=m\}}-2\mathbf{Q}_{\mathbf{s}}^{T}\mathbf{\Sigma}^{-1}(L_{\mathbf{s}})\mathbf{Q}_{\mathbf{s}}\mathbf{\Phi}(L_{\mathbf{s}})\mathbf{1}_{\{L_{\mathbf{s}}=m\}}\\&=2\mathbf{Q}_{\mathbf{s}}^{T}\mathbf{\Sigma}^{-1}(L_{\mathbf{s}})\left(\mathbf{P}_{\mathbf{s}}-\mathbf{Q}_{\mathbf{s}}\mathbf{\Phi}(L_{\mathbf{s}})\right)\mathbf{1}_{\{L_{\mathbf{s}}=m\}}\\&=2\mathbf{Q}_{\mathbf{s}}^{T}\mathbf{\Sigma}^{-1}(L_{\mathbf{s}})\mathbf{X}_{\mathbf{s}}\mathbf{1}_{\{L_{\mathbf{s}}=m\}}\ .\end{aligned} \tag{36}$$

Finally, substituting the result of (36) into (34) gives us the identity

$$\mathbf{0}_{q\times 1}=\sum_{l=1}^{M}\sum_{\mathbf{s}\in\mathcal{S}(l)}\mathbf{Q}_{\mathbf{s}}^{T}\mathbf{\Sigma}^{-1}(L_{\mathbf{s}})\mathbf{X}_{\mathbf{s}}\mathbf{1}_{\{L_{\mathbf{s}}=m\}}=\sum_{\mathbf{s}\in\mathcal{S}(m)}\mathbf{Q}_{\mathbf{s}}^{T}\mathbf{\Sigma}^{-1}(m)\mathbf{X}_{\mathbf{s}}. \tag{37}$$

5.2 Interaction matrix identities

From the covariance identity

$$\operatorname{cov}\left(\mathbf{X}_{\mathbf{s}},\mathbf{X}_{\mathbf{s-r}}|\mathbf{L};\mathbf{\Theta}\right)=\operatorname{cov}^{T}\left(\mathbf{X}_{\mathbf{s-r}},\mathbf{X}_{\mathbf{s}}|\mathbf{L};\mathbf{\Theta}\right) \tag{38}$$

(Ravishanker and Dey, 2002), from (5), we obtain the following relationship:

$$\boldsymbol{\theta}_{-\mathbf{r}}(L_{\mathbf{s}})=\mathbf{\Sigma}(L_{\mathbf{s}})\boldsymbol{\theta}_{\mathbf{r}}^{T}(L_{\mathbf{s}})\mathbf{\Sigma}^{-1}(L_{\mathbf{s}}). \tag{39}$$

One consequence of this result is that $\mathbf{X}_s$ can be written as follows:

$$\mathbf{X}_{\mathbf{s}}=\left(\mathbf{Y}_{\mathbf{s}}-\boldsymbol{\mu}(L_{\mathbf{s}})\right)-\sum_{\mathbf{r}\in\mathcal{N}_S}\left[\boldsymbol{\theta}_{\mathbf{r}}(L_{\mathbf{s}})\mathbf{1}_{\{L_{\mathbf{s}}=L_{\mathbf{s-r}}\}}\left(\mathbf{Y}_{\mathbf{s-r}}-\boldsymbol{\mu}(L_{\mathbf{s}})\right)+\ \mathbf{\Sigma}(L_{\mathbf{s}})\boldsymbol{\theta}_{\mathbf{r}}^{T}(L_{\mathbf{s}})\mathbf{\Sigma}^{-1}(L_{\mathbf{s}})\mathbf{1}_{\{L_{\mathbf{s}}=L_{\mathbf{s+r}}\}}\left(\mathbf{Y}_{\mathbf{s+r}}-\boldsymbol{\mu}(L_{\mathbf{s}})\right)\right] \tag{40}$$

where $\mathcal{N}_s$, a subset of $\mathcal{N}$ which represents the symmetric neighborhood set (Kashyap and Chellappa, 1983), is defined as follows: $\mathbf{r}\in\mathcal{N}_s \Rightarrow -\mathbf{r}\notin\mathcal{N}_s$ and $\mathcal{N}=\left\{\mathbf{r}\in\mathcal{N}_s\cup-\mathbf{r}\in\mathcal{N}_s\right\}$.

Another consequence of (39) are the specifications of the interaction matrices in the separable case. If the interaction matrices are modeled as separable, then by (39), we obtain

$$\boldsymbol{\theta}_{-\mathbf{r}}(m)=\boldsymbol{\theta}_{-\mathbf{r}}^{(2)}(m)\otimes\boldsymbol{\theta}_{-\mathbf{r}}^{(1)}(m)=\mathbf{\Sigma}(m)\left(\boldsymbol{\theta}_{\mathbf{r}}^{(2)}(m)\otimes\boldsymbol{\theta}_{\mathbf{r}}^{(1)}(m)\right)^{T}\mathbf{\Sigma}^{-1}(m)=\mathbf{\Sigma}(m)\boldsymbol{\theta}_{\mathbf{r}}^{T}(m)\mathbf{\Sigma}^{-1}(m) \tag{41}$$

for $1\le m\le M$. The RHS of (40) can be made separable if $\mathbf{\Sigma}(m)$ is also separable. Hence,

$$\begin{aligned}\boldsymbol{\theta}_{-\mathbf{r}}^{(2)}(m)\otimes\boldsymbol{\theta}_{-\mathbf{r}}^{(1)}(m)&=\left(\boldsymbol{\Sigma}^{(2)}(m)\otimes\boldsymbol{\Sigma}^{(1)}(m)\right)\left(\boldsymbol{\theta}_{\mathbf{r}}^{(2)}(m)\otimes\boldsymbol{\theta}_{\mathbf{r}}^{(1)}(m)\right)^{T}\left(\boldsymbol{\Sigma}^{(2)}(m)\otimes\boldsymbol{\Sigma}^{(1)}(m)\right)^{-1}\\&=\left(\boldsymbol{\Sigma}^{(2)}(m)\otimes\boldsymbol{\Sigma}^{(1)}(m)\right)\left(\boldsymbol{\theta}_{\mathbf{r}}^{(2)T}(m)\otimes\boldsymbol{\theta}_{\mathbf{r}}^{(1)T}(m)\right)\left(\left(\boldsymbol{\Sigma}^{(2)}(m)\right)^{-1}\otimes\left(\boldsymbol{\Sigma}^{(1)}(m)\right)^{-1}\right)\\&=\boldsymbol{\Sigma}^{(2)}(m)\boldsymbol{\theta}_{\mathbf{r}}^{(2)T}(m)\left(\boldsymbol{\Sigma}^{(2)}(m)\right)^{-1}\otimes\boldsymbol{\Sigma}^{(1)}(m)\boldsymbol{\theta}_{\mathbf{r}}^{(1)T}(m)\left(\boldsymbol{\Sigma}^{(1)}(m)\right)^{-1}.\end{aligned}\tag{42}$$

The identification of $\boldsymbol{\theta}_{-\mathbf{r}}(m)$ is completely specified from (39) if we take

$$\boldsymbol{\theta}_{-\mathbf{r}}^{(1)}(m)=\boldsymbol{\Sigma}^{(1)}(m)\boldsymbol{\theta}_{\mathbf{r}}^{(1)T}(m)\left(\boldsymbol{\Sigma}^{(1)}(m)\right)^{-1}\tag{43}$$

$$\boldsymbol{\theta}_{-\mathbf{r}}^{(2)}(m)=\boldsymbol{\Sigma}^{(2)}(m)\boldsymbol{\theta}_{\mathbf{r}}^{(2)T}(m)\left(\boldsymbol{\Sigma}^{(2)}(m)\right)^{-1},\tag{44}$$

which is analogous to the relation in (39).

By considering the hybrid separability cases which involve a separable interaction matrix and a non-separable covariance matrix, the expression $\boldsymbol{\Sigma}(m)\boldsymbol{\theta}_{\mathbf{r}}^{T}(m)\boldsymbol{\Sigma}^{-1}(m)$ is not separable, in general. This implies that $\boldsymbol{\theta}_{-\mathbf{r}}(m)$ cannot be expressed in the form $\boldsymbol{\theta}_{-\mathbf{r}}(m)=\boldsymbol{\theta}_{-\mathbf{r}}^{(2)}(m)\otimes\boldsymbol{\theta}_{-\mathbf{r}}^{(1)}(m)$ for $\mathbf{r}\in\mathcal{N}_S$, $1\le m\le M$ and thus these cases are not possible.

6. GMRF parameter estimation

This section proposes an estimation procedure for the GMRF parameters for both separable and non-separable cases based on the MPL.

6.1 Mean parameter estimation

Proposition 2 Assume that the interaction matrix coefficients $\boldsymbol{\theta}_{\mathbf{r}}(m)$ for $\mathbf{r}\in\mathcal{N}$, $1\le m\le M$ and the covariance matrices $\boldsymbol{\Sigma}(m)$ for $1\le m\le M$ are known. Then the mean parameters are estimated as follows:

a. Non-Separable Case:

$$\hat{\boldsymbol{\mu}}(m)=\left[\sum_{\mathbf{s}\in\mathcal{S}(m)}\left(\mathbf{I}_N-\sum_{\mathbf{r}\in\mathcal{N}}\boldsymbol{\theta}_{\mathbf{r}}(m)\mathbf{1}_{\{L_{\mathbf{s}-\mathbf{r}}=m\}}\right)^{T}\boldsymbol{\Sigma}^{-1}(m)\left(\mathbf{I}_N-\sum_{\mathbf{r}\in\mathcal{N}}\boldsymbol{\theta}_{\mathbf{r}}(m)\mathbf{1}_{\{L_{\mathbf{s}-\mathbf{r}}=m\}}\right)\right]^{-1}\cdot\left[\sum_{\mathbf{s}\in\mathcal{S}(m)}\left(\mathbf{I}_N-\sum_{\mathbf{r}\in\mathcal{N}}\boldsymbol{\theta}_{\mathbf{r}}(m)\mathbf{1}_{\{L_{\mathbf{s}-\mathbf{r}}=m\}}\right)^{T}\boldsymbol{\Sigma}^{-1}(m)\left(\mathbf{Y}_{\mathbf{s}}-\sum_{\mathbf{r}\in\mathcal{N}}\boldsymbol{\theta}_{\mathbf{r}}(m)\mathbf{1}_{\{L_{\mathbf{s}-\mathbf{r}}=m\}}\mathbf{Y}_{\mathbf{s}-\mathbf{r}}\right)\right]\tag{45}$$

for $1\le m\le M$.

b. Separable Case:

In addition, if we assume the following for $1\le m\le M$:

- $\boldsymbol{\mu}^{(1)}(m)$ is estimated, given that $\boldsymbol{\mu}^{(2)}(m)$ is known
- $\boldsymbol{\mu}^{(2)}(m)$ is estimated, given that $\boldsymbol{\mu}^{(1)}(m)$ is known.

Thus

$$\hat{\boldsymbol{\mu}}^{(1)}(m)$$
$$= \left[\sum_{\mathbf{s}\in\mathcal{S}(m)} \left(\boldsymbol{\mu}^{(2)}(m)\otimes \mathbf{I}_{N_1}\right)^T \left(\mathbf{I}_N - \sum_{\mathbf{r}\in\mathcal{N}} \boldsymbol{\theta}_\mathbf{r}(m)\mathbf{1}_{\{L_{\mathbf{s}-\mathbf{r}}=m\}}\right)^T \boldsymbol{\Sigma}^{-1}(m)\left(\mathbf{I}_N - \sum_{\mathbf{r}\in\mathcal{N}} \boldsymbol{\theta}_\mathbf{r}(m)\mathbf{1}_{\{L_{\mathbf{s}-\mathbf{r}}=m\}}\right)\left(\boldsymbol{\mu}^{(2)}(m)\otimes \mathbf{I}_{N_1}\right)\right]^{-1} \cdot \tag{46}$$
$$\left[\sum_{\mathbf{s}\in\mathcal{S}(m)} \left(\boldsymbol{\mu}^{(2)}(m)\otimes \mathbf{I}_{N_1}\right)^T \left(\mathbf{I}_N - \sum_{\mathbf{r}\in\mathcal{N}} \boldsymbol{\theta}_\mathbf{r}(m)\mathbf{1}_{\{L_{\mathbf{s}-\mathbf{r}}=m\}}\right)^T \boldsymbol{\Sigma}^{-1}(m)\left(\mathbf{Y}_\mathbf{s} - \sum_{\mathbf{r}\in\mathcal{N}} \boldsymbol{\theta}_\mathbf{r}(m)\mathbf{1}_{\{L_{\mathbf{s}-\mathbf{r}}=m\}}\mathbf{Y}_{\mathbf{s}-\mathbf{r}}\right)\right]$$

$$\hat{\boldsymbol{\mu}}^{(2)}(m)$$
$$= \left[\sum_{\mathbf{s}\in\mathcal{S}(m)} \left(\mathbf{I}_{N_2}\otimes \boldsymbol{\mu}^{(1)}(m)\right)^T \left(\mathbf{I}_N - \sum_{\mathbf{r}\in\mathcal{N}} \boldsymbol{\theta}_\mathbf{r}(m)\mathbf{1}_{\{L_{\mathbf{s}-\mathbf{r}}=m\}}\right)^T \boldsymbol{\Sigma}^{-1}(m)\left(\mathbf{I}_N - \sum_{\mathbf{r}\in\mathcal{N}} \boldsymbol{\theta}_\mathbf{r}(m)\mathbf{1}_{\{L_{\mathbf{s}-\mathbf{r}}=m\}}\right)\left(\mathbf{I}_{N_2}\otimes \boldsymbol{\mu}^{(1)}(m)\right)\right]^{-1} \cdot \tag{47}$$
$$\left[\sum_{\mathbf{s}\in\mathcal{S}(m)} \left(\mathbf{I}_{N_2}\otimes \boldsymbol{\mu}^{(1)}(m)\right)^T \left(\mathbf{I}_N - \sum_{\mathbf{r}\in\mathcal{N}} \boldsymbol{\theta}_\mathbf{r}(m)\mathbf{1}_{\{L_{\mathbf{s}-\mathbf{r}}=m\}}\right)^T \boldsymbol{\Sigma}^{-1}(m)\left(\mathbf{Y}_\mathbf{s} - \sum_{\mathbf{r}\in\mathcal{N}} \boldsymbol{\theta}_\mathbf{r}(m)\mathbf{1}_{\{L_{\mathbf{s}-\mathbf{r}}=m\}}\mathbf{Y}_{\mathbf{s}-\mathbf{r}}\right)\right]$$

for $1 \le m \le M$.

Proof

a. The proof for the non-separable case is derived by applying Proposition 1 (Navarro, et. al., 2009).
b. From (3), $\mathbf{X}_\mathbf{s}$ can be written as follows:

$$\mathbf{X}_\mathbf{s} = \left(\mathbf{Y}_\mathbf{s} - \sum_{\mathbf{r}\in\mathcal{N}} \boldsymbol{\theta}_\mathbf{r}(L_\mathbf{s})\mathbf{1}_{\{L_{\mathbf{s}-\mathbf{r}}=L_\mathbf{s}\}}\mathbf{Y}_{\mathbf{s}-\mathbf{r}}\right) - \left(\mathbf{I}_N - \sum_{\mathbf{r}\in\mathcal{N}} \boldsymbol{\theta}_\mathbf{r}(L_\mathbf{s})\mathbf{1}_{\{L_{\mathbf{s}-\mathbf{r}}=L_\mathbf{s}\}}\right)\boldsymbol{\mu}(L_\mathbf{s}). \tag{48}$$

For the separable case, the mean can be written as follows:

$$\begin{aligned}\boldsymbol{\mu}(m) &= \boldsymbol{\mu}^{(2)}(m)\otimes\boldsymbol{\mu}^{(1)}(m)\\ &= \left(\boldsymbol{\mu}^{(2)}(m)\otimes\mathbf{I}_{N_1}\right)\left(1\otimes\boldsymbol{\mu}^{(1)}(m)\right) = \left(\boldsymbol{\mu}^{(2)}(m)\otimes\mathbf{I}_{N_1}\right)\boldsymbol{\mu}^{(1)}(m)\\ &= \left(\mathbf{I}_{N_2}\otimes\boldsymbol{\mu}^{(1)}(m)\right)\left(\boldsymbol{\mu}^{(2)}(m)\otimes 1\right) = \left(\mathbf{I}_{N_2}\otimes\boldsymbol{\mu}^{(1)}(m)\right)\boldsymbol{\mu}^{(2)}(m).\end{aligned} \tag{49}$$

Plugging the results of (49) into (48) yields

$$\begin{aligned}\mathbf{X}_\mathbf{s} &= \left(\mathbf{Y}_\mathbf{s} - \sum_{\mathbf{r}\in\mathcal{N}} \boldsymbol{\theta}_\mathbf{r}(L_\mathbf{s})\mathbf{1}_{\{L_{\mathbf{s}-\mathbf{r}}=L_\mathbf{s}\}}\mathbf{Y}_{\mathbf{s}-\mathbf{r}}\right) - \left(\mathbf{I}_N - \sum_{\mathbf{r}\in\mathcal{N}} \boldsymbol{\theta}_\mathbf{r}(L_\mathbf{s})\mathbf{1}_{\{L_{\mathbf{s}-\mathbf{r}}=L_\mathbf{s}\}}\right)\left(\boldsymbol{\mu}^{(2)}(L_\mathbf{s})\otimes\mathbf{I}_{N_1}\right)\boldsymbol{\mu}^{(1)}(L_\mathbf{s})\\ &= \left(\mathbf{Y}_\mathbf{s} - \sum_{\mathbf{r}\in\mathcal{N}} \boldsymbol{\theta}_\mathbf{r}(L_\mathbf{s})\mathbf{1}_{\{L_{\mathbf{s}-\mathbf{r}}=L_\mathbf{s}\}}\mathbf{Y}_{\mathbf{s}-\mathbf{r}}\right) - \left(\mathbf{I}_N - \sum_{\mathbf{r}\in\mathcal{N}} \boldsymbol{\theta}_\mathbf{r}(L_\mathbf{s})\mathbf{1}_{\{L_{\mathbf{s}-\mathbf{r}}=L_\mathbf{s}\}}\right)\left(\mathbf{I}_{N_2}\otimes\boldsymbol{\mu}^{(1)}(L_\mathbf{s})\right)\boldsymbol{\mu}^{(2)}(L_\mathbf{s}).\end{aligned} \tag{50}$$

$$\left(1^{\circ}\right) \qquad \mathbf{\Phi}(m) = \boldsymbol{\mu}^{(1)}(m),\ 1 \le m \le M$$

For this case, we recognize the following from (50):

$$\mathbf{Q}_{\mathbf{s}} = \left(\mathbf{I}_N - \sum_{\mathbf{r}\in\mathcal{N}} \boldsymbol{\theta}_{\mathbf{r}}(L_{\mathbf{s}})\mathbf{1}_{\{L_{\mathbf{s-r}}=L_{\mathbf{s}}\}}\right)\left(\boldsymbol{\mu}^{(2)}(L_{\mathbf{s}}) \otimes \mathbf{I}_{N_1}\right). \tag{51}$$

By applying Preposition 1 and rearranging terms, we obtain (46).

$$\left(2^{\circ}\right) \qquad \mathbf{\Phi}(m) = \boldsymbol{\mu}^{(1)}(m),\ 1 \le m \le M$$

For this case from (50), we recognize

$$\mathbf{Q}_{\mathbf{s}} = \left(\mathbf{I}_N - \sum_{\mathbf{r}\in\mathcal{N}} \boldsymbol{\theta}_{\mathbf{r}}(L_{\mathbf{s}})\mathbf{1}_{\{L_{\mathbf{s-r}}=L_{\mathbf{s}}\}}\right)\left(\mathbf{I}_{N_2} \otimes \boldsymbol{\mu}^{(1)}(L_{\mathbf{s}})\right). \tag{52}$$

by applying Preposition 1 and rearranging terms, we obtain (47).

6.2 Interaction matrix parameter estimation

Proposition 3 Assume that the mean vectors $\boldsymbol{\mu}(m)$ for $1 \le m \le M$ and the covariance matrices $\boldsymbol{\Sigma}(m)$ for $1 \le m \le M$ are known, then interaction matrix parameters are estimated by solving the simultaneous linear equations given as follows:

a. Non-Separable Case:

$$\mathbf{H}(m)\mathbf{\Psi}(m) = \mathbf{\Gamma}(m) \tag{53}$$

where

$$\mathbf{H}(m) = row\left\{col\left(\sum_{\mathbf{s}\in\mathcal{S}(m)} \mathbf{A}_{\mathbf{s},\mathbf{t}}(m)\boldsymbol{\Sigma}^{-1}(m)\mathbf{A}_{\mathbf{s},\mathbf{r}}^{T}(m),\ \mathbf{r}\in\mathcal{N}_S\right),\ \mathbf{t}\in\mathcal{N}_S\right\} \tag{54}$$

$$\mathbf{\Gamma}(m) = row\left(\sum_{\mathbf{s}\in\mathcal{S}(m)} \mathbf{A}_{\mathbf{s},\mathbf{t}}(m)\boldsymbol{\Sigma}^{-1}(m)\left(\mathbf{Y}_{\mathbf{s}} - \boldsymbol{\mu}(m)\right),\ \mathbf{t}\in\mathcal{N}_S\right) \tag{55}$$

$$\mathbf{\Psi}(m) = row\left(vec\left(\hat{\boldsymbol{\theta}}_{\mathbf{r}}(m)\right),\ \mathbf{r}\in\mathcal{N}_S\right) \tag{56}$$

and

$$\mathbf{A}_{\mathbf{s},\mathbf{r}}(m) = \left(\left(\mathbf{Y}_{\mathbf{s-r}} - \boldsymbol{\mu}(m)\right)\mathbf{1}_{\{L_{\mathbf{s-r}}=m\}} \otimes \mathbf{I}_N\right) + \mathbf{K}_{N,N}\left(\boldsymbol{\Sigma}^{-1}(m)\otimes\boldsymbol{\Sigma}(m)\right)\left(\left(\mathbf{Y}_{\mathbf{s+r}} - \boldsymbol{\mu}(m)\right)\mathbf{1}_{\{L_{\mathbf{s+r}}=m\}} \otimes \mathbf{I}_N\right). \tag{57}$$

From the invariance property of the MPL, the complete set of non-separable interaction matrix estimators is estimated as follows:

$$\hat{\boldsymbol{\theta}}_{\mathbf{r}}(m) = reshape\left(vec\left(\hat{\boldsymbol{\theta}}_{\mathbf{r}}(m)\right), N, N\right) \tag{58}$$

$$\hat{\boldsymbol{\theta}}_{-\mathbf{r}}(m) = \boldsymbol{\Sigma}(m)\hat{\boldsymbol{\theta}}_{\mathbf{r}}^{T}(m)\left(\boldsymbol{\Sigma}(m)\right)^{-1} \tag{59}$$

for $\mathbf{r} \in \mathcal{N}_s$, $1 \le m \le M$.

b. Separable Case:

In addition, if we assume the following for $\mathbf{r} \in \mathcal{N}_s$ and $1 \le m \le M$:

- $\boldsymbol{\theta}_{\mathbf{r}}^{(1)}(m)$ is estimated, given that $\boldsymbol{\theta}_{\mathbf{r}}^{(2)}(m)$ is known
- $\boldsymbol{\theta}_{\mathbf{r}}^{(2)}(m)$ is estimated, given that $\boldsymbol{\theta}_{\mathbf{r}}^{(1)}(m)$ is known

then

$$\mathbf{H}^{(k)}(m)\boldsymbol{\Psi}^{(k)}(m) = \boldsymbol{\Gamma}^{(k)}(m) \tag{60}$$

where

$$\mathbf{H}^{(k)}(m) = row\left\{col\left(\sum_{\mathbf{s}\in\mathcal{S}(m)} \mathbf{A}_{\mathbf{s},\mathbf{t}}^{(k)}(m)\boldsymbol{\Sigma}^{-1}(m)\mathbf{A}_{\mathbf{s},\mathbf{r}}^{(k)T}(m),\ \mathbf{r}\in\mathcal{N}_S\right),\ \mathbf{t}\in\mathcal{N}_S\right\} \tag{61}$$

$$\boldsymbol{\Gamma}^{(k)}(m) = row\left(\sum_{\mathbf{s}\in\mathcal{S}(m)} \mathbf{A}_{\mathbf{s},\mathbf{t}}^{(k)}(m)\boldsymbol{\Sigma}^{-1}(m)\left(\mathbf{Y}_{\mathbf{s}} - \boldsymbol{\mu}(m)\right),\ \mathbf{t}\in\mathcal{N}_S\right) \tag{62}$$

$$\boldsymbol{\Psi}^{(k)}(m) = row\left(vec\left(\hat{\boldsymbol{\theta}}_{\mathbf{r}}^{(k)}(m)\right),\ \mathbf{r}\in\mathcal{N}_S\right) \tag{63}$$

for $1 \le k \le 2$ and

$$\mathbf{A}_{\mathbf{s},\mathbf{r}}^{(1)}(m) = \left(vec\left(\boldsymbol{\theta}_{\mathbf{r}}^{(2)}(m)\right)\otimes\mathbf{I}_{N_1}\right)^{T}\left(\mathbf{I}_{N_2}\otimes\mathbf{K}_{N_1,N_2}\otimes\mathbf{I}_{N_1}\right)^{T}\mathbf{A}_{\mathbf{s},\mathbf{r}}(m) \tag{64}$$

$$\mathbf{A}_{\mathbf{s},\mathbf{r}}^{(2)}(m) = \left(\mathbf{I}_{N_2}\otimes vec\left(\boldsymbol{\theta}_{\mathbf{r}}^{(1)}(m)\right)\right)^{T}\left(\mathbf{I}_{N_2}\otimes\mathbf{K}_{N_1,N_2}\otimes\mathbf{I}_{N_1}\right)^{T}\mathbf{A}_{\mathbf{s},\mathbf{r}}(m). \tag{65}$$

From the invariance property of the MPL, the complete set of separable interaction matrix estimators is estimated as follows for $\mathbf{r}\in\mathcal{N}_s$, $1\le m\le M$, $1\le k\le 2$:

$$\hat{\boldsymbol{\theta}}_{\mathbf{r}}^{(k)}(m) = reshape\left(vec\left(\hat{\boldsymbol{\theta}}_{\mathbf{r}}^{(k)}(m)\right), N_k, N_k\right) \tag{66}$$

$$\hat{\boldsymbol{\theta}}_{\mathbf{r}}^{(k)}(m) = \boldsymbol{\Sigma}^{(k)}(m)\hat{\boldsymbol{\theta}}_{\mathbf{r}}^{(k)T}(m)\left(\boldsymbol{\Sigma}^{(k)}(m)\right)^{-1} \tag{67}$$

and also

$$\hat{\boldsymbol{\theta}}_{\mathbf{r}}(m) = \hat{\boldsymbol{\theta}}_{\mathbf{r}}^{(2)}(m) \otimes \hat{\boldsymbol{\theta}}_{\mathbf{r}}^{(1)}(m) \tag{68}$$

for $\mathbf{r} \in \mathcal{N}_s$ and $1 \le m \le M$.

Proof

a. The proof for the non-separable case is derived by applying Proposition 1 (Navarro, et. al., 2009).

b. From (3), $\mathbf{X_s}$ can be written as

$$\mathbf{X_s} = vec(\mathbf{X_s}) = (\mathbf{Y_s} - \boldsymbol{\mu}(L_\mathbf{s})) - \sum_{\mathbf{r}\in\mathcal{N}_S} \mathbf{A}_{\mathbf{s},\mathbf{r}}^T(L_\mathbf{s}) vec(\boldsymbol{\theta}_\mathbf{r}(L_\mathbf{s})). \tag{69}$$

The above expression can also be written using the following matrix identities (Magnus and Neudecker, 1999)

$$vec(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A}) vec(\mathbf{B}) \tag{70}$$

where $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{B} \in \mathbb{R}^{n\times p}$, and $\mathbf{C} \in \mathbb{R}^{p\times q}$.

$$(\mathbf{A} \otimes \mathbf{B})^T = \mathbf{A}^T \otimes \mathbf{B}^T \tag{71}$$

$$vec(\mathbf{A}^T) = \mathbf{K}_{m,n} vec(\mathbf{A}) \tag{72}$$

where $\mathbf{A} \in \mathbb{R}^{m\times n}$. In addition, from the identity (Magnus and Neudecker, 1999)

$$vec(\mathbf{A} \otimes \mathbf{B}) = (\mathbf{I}_n \otimes \mathbf{K}_{q,m} \otimes \mathbf{I}_p) \cdot (vec(\mathbf{A}) \otimes vec(\mathbf{B})) \tag{73}$$

where $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{B} \in \mathbb{R}^{p\times q}$, it follows that

$$\begin{aligned} vec(\boldsymbol{\theta}_\mathbf{r}(m)) &= vec\left(\boldsymbol{\theta}_\mathbf{r}^{(2)}(m) \otimes \boldsymbol{\theta}_\mathbf{r}^{(1)}(m)\right) \\ &= \left(\mathbf{I}_{N_2} \otimes \mathbf{K}_{N_1,N_2} \otimes \mathbf{I}_{N_1}\right) \cdot \left(vec\left(\boldsymbol{\theta}_\mathbf{r}^{(2)}(m)\right) \otimes vec\left(\boldsymbol{\theta}_\mathbf{r}^{(1)}(m)\right)\right). \end{aligned} \tag{74}$$

Furthermore, since

$$\begin{aligned} &\left(vec\left(\boldsymbol{\theta}_\mathbf{r}^{(2)}(m)\right) \otimes vec\left(\boldsymbol{\theta}_\mathbf{r}^{(1)}(m)\right)\right) \\ &= \left(vec\left(\boldsymbol{\theta}_\mathbf{r}^{(2)}(m)\right) \otimes \mathbf{I}_{N_1}\right)\left(1 \otimes vec\left(\boldsymbol{\theta}_\mathbf{r}^{(1)}(m)\right)\right) = \left(vec\left(\boldsymbol{\theta}_\mathbf{r}^{(2)}(m)\right) \otimes \mathbf{I}_{N_1}\right) vec\left(\boldsymbol{\theta}_\mathbf{r}^{(1)}(m)\right) \\ &= \left(\mathbf{I}_{N_2} \otimes vec\left(\boldsymbol{\theta}_\mathbf{r}^{(1)}(m)\right)\right)\left(vec\left(\boldsymbol{\theta}_\mathbf{r}^{(2)}(m)\right) \otimes 1\right) = \left(\mathbf{I}_{N_2} \otimes vec\left(\boldsymbol{\theta}_\mathbf{r}^{(1)}(m)\right)\right) vec\left(\boldsymbol{\theta}_\mathbf{r}^{(2)}(m)\right) \end{aligned} \tag{75}$$

then,

$$\begin{aligned} vec(\boldsymbol{\theta}_\mathbf{r}(m)) &= \left(\mathbf{I}_{N_2} \otimes \mathbf{K}_{N_1,N_2} \otimes \mathbf{I}_{N_1}\right) \cdot \left(vec\left(\boldsymbol{\theta}_\mathbf{r}^{(2)}(m)\right) \otimes \mathbf{I}_{N_1}\right) vec\left(\boldsymbol{\theta}_\mathbf{r}^{(1)}(m)\right) \\ &= \left(\mathbf{I}_{N_2} \otimes \mathbf{K}_{N_1,N_2} \otimes \mathbf{I}_{N_1}\right) \cdot \left(\mathbf{I}_{N_2} \otimes vec\left(\boldsymbol{\theta}_\mathbf{r}^{(1)}(m)\right)\right) vec\left(\boldsymbol{\theta}_\mathbf{r}^{(2)}(m)\right). \end{aligned} \tag{76}$$

Plugging the results of (76) into (69) yields

$$\begin{aligned}\mathbf{X_s} &= \left(\mathbf{Y_s} - \boldsymbol{\mu}(L_\mathbf{s})\right) - \sum_{\mathbf{r}\in\mathcal{N}_S} \mathbf{A}_{\mathbf{s,r}}^{(1)T}(L_\mathbf{s})\, vec\left(\boldsymbol{\theta}_\mathbf{r}^{(1)}(L_\mathbf{s})\right) \\ &= \left(\mathbf{Y_s} - \boldsymbol{\mu}(L_\mathbf{s})\right) - \sum_{\mathbf{r}\in\mathcal{N}_S} \mathbf{A}_{\mathbf{s,r}}^{(2)T}(L_\mathbf{s})\, vec\left(\boldsymbol{\theta}_\mathbf{r}^{(2)}(L_\mathbf{s})\right)\end{aligned} \tag{77}$$

$$\left(1^\circ\right) \qquad \boldsymbol{\Phi}(m) = vec\left(\boldsymbol{\theta}_\mathbf{t}^{(1)}(m)\right),\ \mathbf{t}\in\mathcal{N}_s,\ 1\le m\le M$$

For this case, we recognize from (77),

$$\mathbf{Q_s} = \mathbf{A}_{\mathbf{s,t}}^{(1)T}(m). \tag{78}$$

By applying Preposition 1 and rearranging terms, we obtain the following expression

$$\sum_{\mathbf{r}\in\mathcal{N}_S}\sum_{\mathbf{s}\in\mathcal{S}(m)} \mathbf{A}_{\mathbf{s,t}}^{(1)}(m)\boldsymbol{\Sigma}^{-1}(m)\mathbf{A}_{\mathbf{s,r}}^{(1)T}(m)\, vec\left(\boldsymbol{\theta}_\mathbf{r}^{(1)}(m)\right) = \sum_{\mathbf{s}\in\mathcal{S}(m)} \mathbf{A}_{\mathbf{s,t}}^{(1)}(m)\boldsymbol{\Sigma}^{-1}(m)\left(\mathbf{Y_s} - \boldsymbol{\mu}(m)\right). \tag{79}$$

By aggregating the equation in (79) for $\mathbf{t}\in\mathcal{N}_S$, the interaction matrix coefficients are estimated by solving the simultaneous linear equations in (60) for $k=1$.

$$\left(2^\circ\right) \qquad \boldsymbol{\Phi}(m) = vec\left(\boldsymbol{\theta}_\mathbf{t}^{(2)}(m)\right),\ \mathbf{t}\in\mathcal{N}_s,\ 1\le m\le M$$

For this case, we recognize from (77)

$$\mathbf{Q_s} = \mathbf{A}_{\mathbf{s,t}}^{(2)T}(m). \tag{80}$$

By applying Preposition 1 and rearranging terms, we obtain the following expression

$$\sum_{\mathbf{r}\in\mathcal{N}_S}\sum_{\mathbf{s}\in\mathcal{S}(m)} \mathbf{A}_{\mathbf{s,t}}^{(2)}(m)\boldsymbol{\Sigma}^{-1}(m)\mathbf{A}_{\mathbf{s,r}}^{(2)T}(m)\, vec\left(\boldsymbol{\theta}_\mathbf{r}^{(2)}(m)\right) = \sum_{\mathbf{s}\in\mathcal{S}(m)} \mathbf{A}_{\mathbf{s,t}}^{(2)}(m)\boldsymbol{\Sigma}^{-1}(m)\left(\mathbf{Y_s} - \boldsymbol{\mu}(m)\right). \tag{81}$$

By aggregating the equations in (79) for $\mathbf{t}\in\mathcal{N}_S$, the interaction matrix coefficients are estimated by solving the simultaneous linear equations in (60) for $k=2$.

6.3 Covariance matrix parameter estimation

Since $\mathbf{X_s}$ is dependent on a covariance matrix in finding the MPL estimator of $\boldsymbol{\Sigma}(m)$, for all $1\le m\le M$ is cumbersome to derive. As an alternative, we estimate the covariance matrix as the sample covariance matrix given that the mean vectors $\boldsymbol{\mu}(m)$ for $1\le m\le M$ and the interaction matrix coefficients $\boldsymbol{\theta}_\mathbf{r}(m)$, for $\mathbf{r}\in\mathcal{N}$, $1\le m\le M$ are known, then the covariance matrix parameters are estimated as follows:

a. Non-Separable Case:

$$\hat{\boldsymbol{\Sigma}}(m) = \frac{1}{r(m)}\sum_{\mathbf{s}\in\mathcal{S}(m)} \mathbf{X_s}\mathbf{X_s}^T \tag{82}$$

b. Separable Case:

In addition, if we assume the following for $1 \le m \le M$:

- $\mathbf{\Sigma}^{(1)}(m)$ is estimated, given that $\mathbf{\Sigma}^{(2)}(m)$ is known
- $\mathbf{\Sigma}^{(2)}(m)$ is estimated, given that $\mathbf{\Sigma}^{(1)}(m)$ is known

then

$$\hat{\mathbf{\Sigma}}^{(1)}(m) = \frac{1}{r(m)N_2} \sum_{\mathbf{s}\in\mathcal{S}(m)} \mathbf{X}_{\mathbf{s}}^{\#} \left(\hat{\mathbf{\Sigma}}^{(2)}(m)\right)^{-1} \mathbf{X}_{\mathbf{s}}^{\#T} \tag{83}$$

$$\hat{\mathbf{\Sigma}}^{(2)}(m) = \frac{1}{r(m)N_1} \sum_{\mathbf{s}\in\mathcal{S}(m)} \mathbf{X}_{\mathbf{s}}^{\#T} \left(\mathbf{\Sigma}^{(1)}(m)\right)^{-1} \mathbf{X}_{\mathbf{s}}^{\#}. \tag{84}$$

The above estimators are not in their closed form. The estimators can be solved iteratively using the flip-flop algorithm (Dutilleul, 1999).

7. Data preparation

The multispectral and multitemporal satellite image under consideration is the 'Butuan' image acquired from the LANDSAT TM. The image shows the scenery of Butuan City and its surroundings in Northeastern Mindanao, Philippines. It consists of six spectral bands and four temporal slots with a dynamic range of 8 bits. The images were captured chronologically on the following dates: August 1, 1992, August 7, 2000, May 22, 2001, and December 3, 2002. The images were radiometrically corrected, geometrically co-registered with each other, and have been resized to 600 x 800 pixels. The image in Fig. 1 is a gray-scaled RGB realization captured on May 22, 2001.

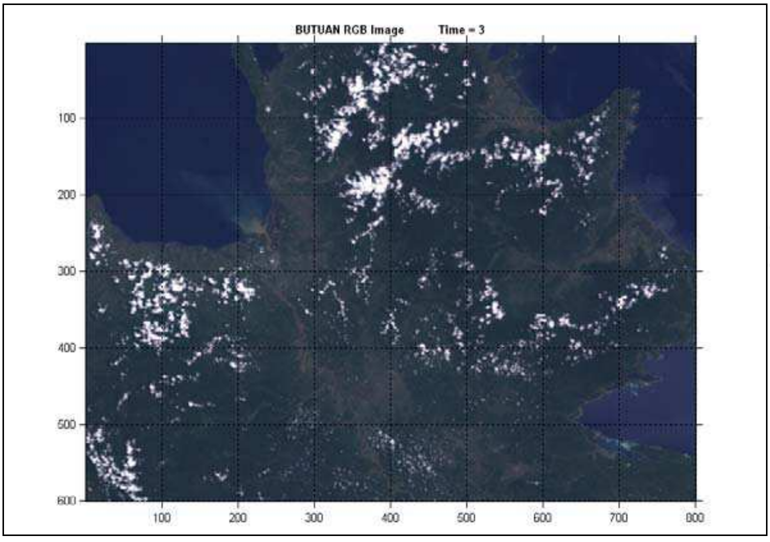


Fig. 1. RGB image of 'Butuan' captured on May 22, 2001.

The thematic classes were established by employing the k-means algorithm (Richards and Jia, 2006). The thematic classes were identified and their mean reflectance vector form the training data are shown in Table 1.

M	Thematic Class	Landsat TM Band Number					
		1	*2*	*3*	*4*	*5*	*7*
1	Thick Vegetation	62	48	33	91	69	29
2	Sparse Vegetation	70	58	43	99	83	37
3	Built Up Areas	77	63	54	75	78	41
4	Body of Water	72	41	29	12	13	11
5	Thin Clouds	104	84	76	88	85	53
6	Thick Clouds	197	190	190	144	167	115

Table 1. Average reflectances from the training data.

Training and verification sites were obtained from a random sample of 1200 sites. The first-order neighborhood system in the MRF modeling of the thematic map and the image were used.

8. Discussion

8.1 Non-separable case

The classification performance of our model with non-separable MGMRF parameters, as compared to the GSC, Hazel's, and Rellier's models are presented in Table 2.

Model	Accuracy
GSC	55.3%
Hazel's GMRF	45.6%
Rellier's GMRF	83.1%
Our Model	84.3%

Table 2. Classification Accuracy of Different MGMRF models.

The GSC model has a low accuracy compared to the remaining MGMRF models. It substantiates that Markov dependence would yield a better accuracy to the thematic map classification than to the site independence model.

It is noticeable that Hazel's GMRF presents a relatively poor classification accuracy which is attributed to the bilateral symmetry imposed into the interaction matrices, that is,

$$\boldsymbol{\theta}_{\mathbf{r}}\left(L_{\mathbf{s}}\right)=\boldsymbol{\theta}_{-\mathbf{r}}\left(L_{\mathbf{s}}\right) \tag{85}$$

(Hazel, 2000) which in general, does not hold the multivariate case. This relation, however, holds in the univariate case (Kashyap and Chellappa, 1983) as well as the Rellier's GMRF.

On the other hand, anisotropic models, such as Rellier's GMRF, and our model exhibited a substantially better classification performance as compared to the GSC. Since the covariance matrix estimators used a sub-optimal alternative, some slight performance degradation has resulted.

8.2 Hybrid separable case

Denote S_{μ}, S_{θ}, and S_{Σ} to be the separable indicators for the mean, interaction matrix, and covariance matrix, respectively.

8.2.1 Hybrid separable GSC model

Since the GSC model is a degenerate form of our MGMRF with zero interaction matrices, the separability structure of the mean and covariance matrices are examined. The results are presented in Table 3 showed that no improvement in the classification performance, regardless of separability of the parameters.

S_{Σ}	S_{μ}	Accuracy
0	0	55.3%
0	1	54.2%
0	0	54.3%
1	1	54.1%

Table 3. Classification Accuracy of Hybrid Separable GSC models.

8.2.2 Hybrid separable anisotropic GMRF model

The hybrid separable anisotropic MGMRF shows the separability of the covariance matrix has a slight improvement in performance over a non-separable spectro-temporal observation. As discussed in Section 5.2, the hybrid separable model with separable interaction matrix, together with a non-separable matrix, were excluded in the model performance as these modes are not possible. The classification accuracy is presented in Table 4.

S_{Σ}	S_{θ}	S_{μ}	Accuracy
0	0	0	84.3%
0	0	1	84.6%
0	1	0	
0	1	1	

S_Σ	S_θ	S_μ	Accuracy
1	0	0	84.5%
1	0	1	86.6%
1	1	0	83.8%
1	1	1	86.2%

Table 4. Classification Accuracy of Hybrid Separable Anisotropic MGMRF models.

8.3 Thematic maps

Some of the thematic map labels are presented in Figs. 2 to 4, based on the May 22, 2001 satellite image. For clarity of visual presentation, thematic map labels were based on the gray-scaled average RGB reflectance of the training data.

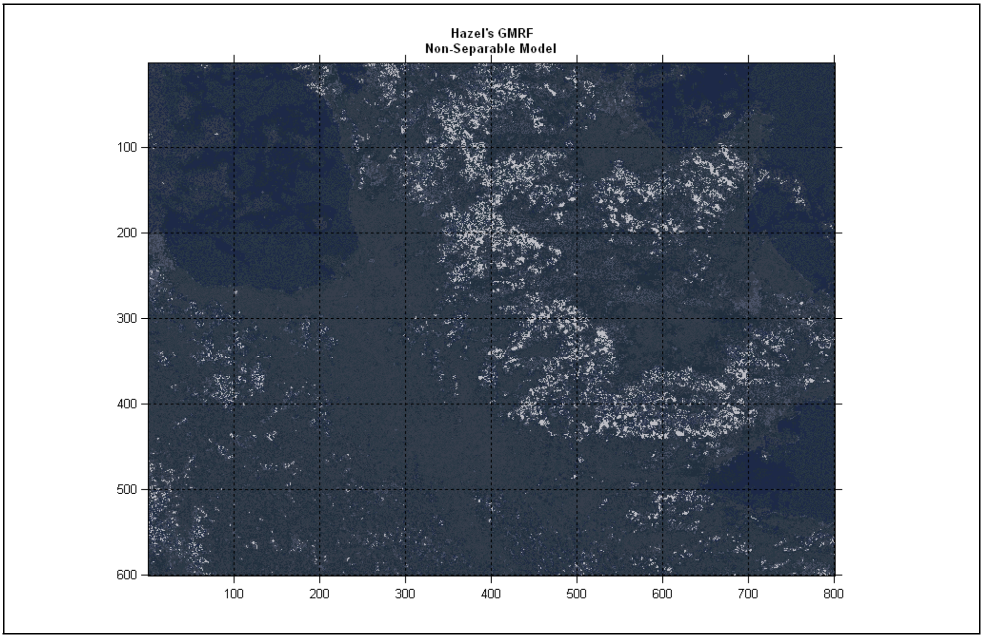


Fig. 2. Thematic Map – Hazel's MGMRF

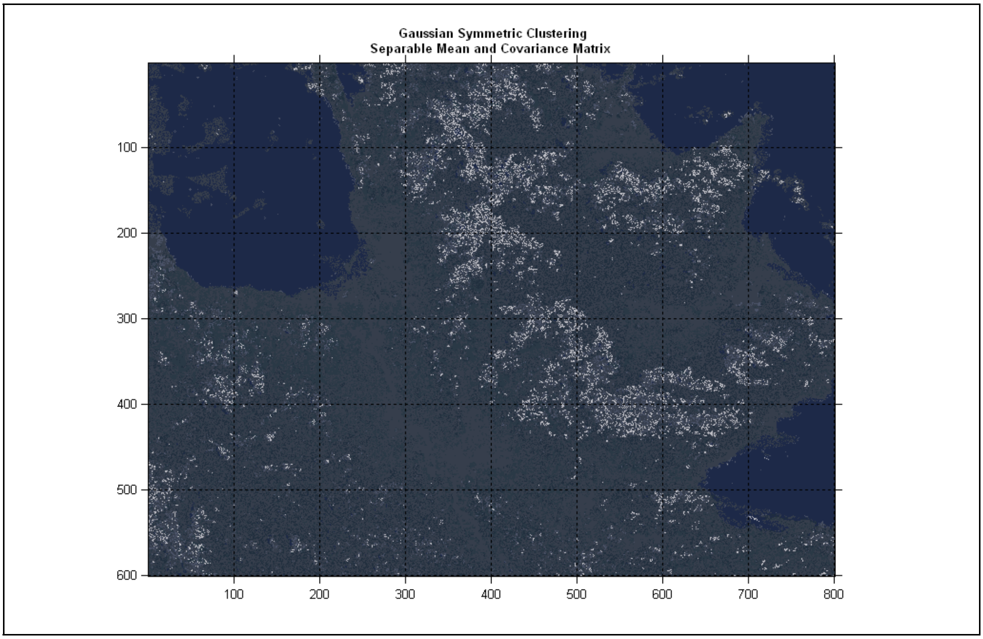


Fig. 3. Thematic Map – GSC with separable mean and covariance matrix

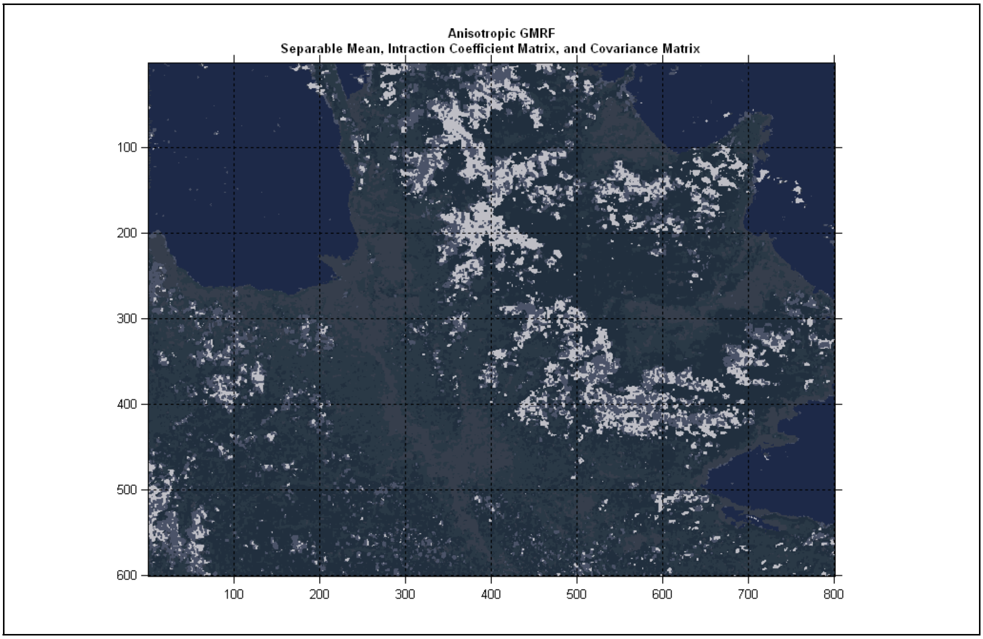


Fig. 4. Thematic Map – Anisotropic MGMRF separable mean, interaction matrices, and covaraince matrices

9. Summary, conclusions, and recommendations

This study presents a parameter estimation procedure based on the MPL for an anisotropic MGMRF with hybrid-separable parameters. Although the MGMRF is a natural extension of its univariate counterpart, the interaction matrix relationship is, in general, dependent on the covariance matrix. In an effort to make the estimation and classification procedure more tractable to compute, some sub-optimal approximations were incorporated. This resulted in a slight degradation in the classification performance. The classification performance based on our model performed well when compared to the GSC model and Hazel’s MGMRF. Nonetheless, its performance is comparable to the Rellier’s MGMRF. Moreover, for spectro-temporal observations, the separability of the interaction matrix as well as the covariance matrix improved the classification performance. Computational capabilities are foreseen to further advance in the near future following the improvement of numerical estimation and classification procedures.

This study presents a parameter estimation procedure based on the MPL for anisotropic MGMRF with hybrid-separable parameters. Although the MGMRF is a natural extension of its univariate counterpart, the interaction matrix relationship is, in general, dependent on the covariance matrix. In an effort to make the estimation and classification procedure more tractable to compute, some sub-optimal approximations were incorporated in the process. This resulted in a slight degradation in the classification performance. The classification performance, based on our model, has performed well, as compared to the GSC model and Hazel’s MGMRF. Furthermore, its performance is comparable to Rellier's MGMRF. In terms of spectro-temporal observations, the separability of the covariance matrix has improved the classification performance. This study can be improved even more with numerical estimation and classification procedure as computational capabilities. This is foreseen to further advance in the near future.

10. Acknowledgment

We acknowledge the invaluable support of extended by the Statistical Training and Research Center of the Philippine Statistical System.

11. References

Aarts, E. and Korts, J. (1987). *Simulated Annealing and Boltzmann Machines*, Wiley, ISBN 978-047-1921-46-2, New York

Arnold, S. F. (1981). *The Theory of Linear Models and Multivariate Analysis*, Wiley, ISBN 978-047-1050-65-0, New York

Besag, J. (1986). On the Statistical Analysis of Dirty Pictures (with discussions). *Journal of Royal Statistical Society B*. Vol. 48, No. 3, pp. 259-302, ISSN 0035-9246

Campbell, N. A. and Kiiveri, H. T. (1988). Spectral-Temporal Indices for Discrimination. *Applied Statistics*, Vol. 37, No. 1, pp. 51-62, ISSN 0035-9254

Casella, G. & Berger, R. L. (2002). *Statistical Inference 2nd ed.*, Wadswoth Group, ISBN 978-053-4243-12-8, Pacific Grove, CA

Dutilleul, P. (1999). The MLE Algorithm for the Matrix Normal Distribution. *Journal of Statistical Computation and Simulation*, Vol. 64, No. 2 , ISSN 0094-9655

Fuentes, M. (2006). Testing for Separability of Spatial-Temporal Covariance Functions. *Journal of Statistical Planning and Inference*, Vol. 136, pp. 447-466, ISSN 0378-3758

Geman, S. & Graffigne, C. (1987). Markov Random Field Models and Their Applications to Computer Vision, Proceedings of the International Congress of Mathematicians, ISBN 978-082-1801-10-9,Berkeley, CA, August, 1986

Hazel, G. G. (2000). Multivariate Gaussian MRF for Multispectral Scene Segmentation and Anomaly Detection. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 38, No. 3, (May 2000), pp. 1199–1211, ISSN 0196-2892

Huizenga, H., Munck, J., Waldorp, R., Grasman, R. (2002). Spatiotemporal EEG/MEG Source Analysis Based on a Parametric Noise Covariance Model. *IEEE Transactions on Biomedical Engineering*, Vol. 49, No. 6, (June 2002), pp. 533-539, ISSN 0018-9294

Jeng, F. and Woods, J. (1991). Compound Gauss-Markov Random Fields for Image Estimation. *IEEE Transactions on Signal Processing*, Vol. 39, No. 3, (March 1991), pp. 683–697, ISSN 1053-587X

Kashyap, R. and Chellappa, R. (1983). Estimation and Choice of Neighbors in Spatial-Interaction Models of Images. *IEEE Transactions on Information Theory*, Vol. 29, No. 1, (January 1983), pp. 60-72, ISSN 0018-9448

Kreyzig, E. (2005). *Advanced Engineering Mathematics, 8th. ed.*, Wiley, ISBN 978-047-1488-85-9, New York

Kyriakidis, P. C. & Journel, A. G. (1999). Geostatistical Space-Time Models: A Review. *Mathematical Geology*, Vol. 31, No. 6, (August 1999), pp. 651-684, ISSN 0882-8121

Li, S. Z. (1995) *Markov Random Field Modeling in Computer Vision*, Springer-Verlag, ISBN 978-4431701453, New York

Lu, N. & Zimmerman, D. (2005). The Likelihood Ratio Test for a Separable Covariance Matrix. *Statistics and Probability Letters*, Vol. 73, No. 4, (July 2005), pp. 449-457, ISSN 0167-7152

Magnus, J. R. & H. Neudecker (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics 2nd ed.*, Wiley, ISBN 978-047-1986-33-1, Chichester

Moura, J. M. F. & Balram N. (1993). Chapter 15: Statistical Algorithms for Noncausal Markov Random Fields, In: *Handbook of Statistics Volume 10*, Bose, N. K. & Rao, C. R., pp. 623-691, North Holland, ISBN 978-044-4892-05-8, Amsterdam

Naik, D. N. & Rao, S. S. (2001). Analysis of Multivariate Repeated Measures Data with a Kronecker Product Structured Covariance Matrix. *Journal of Applied Statistics*, Vol. 28, No. 1, (January 2001), pp. 91-105, ISSN 0013-1644;

Navarro, R. D. Jr., Magadia, J. C., & Paringit, E. C. (2009). Estimating the Gauss-Markov Random Field Parameters for Remote Sensing Image Textures, Proceedings of TENCON 2009 - 2009 IEEE Region 10 Conference, ISBN 978-142-4445-46-2, Singapore, November, 2009

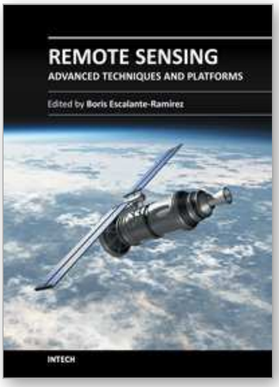
Neudecker, H. (1969). Some Theorems on Matrix Differentiation with Special Reference to Kronecker Matrix Products. Journal of American Statistical Association, Vol. 64, No. 327, (September 1969), pp. 953-963, ISSN 0162-1459

Ravishanker, N. & Dey, D. K. (2002). *A First Course in Linear Model Theory*, CRC Press LLC, ISBN 978-158-4882-47-3, Boca Raton, FL

Richards, J. A. & Jia, X. (1999). *Remote Sensing Image Analysis: An Introduction, 4th ed.*, Springer-Verlag. ISBN 978-354-0251-28-6, Berlin

Rellier, G., Descombes, X., Falzon, F., & Zerubia, J. (2004). Texture Feature Analysis Using a Gauss–Markov Model in Hyperspectral Image Segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 42, No. 7, (July 2004), pp. 1543-1551, ISSN 0196-2892

Winkler, G. (2006). *Image Analysis, Random Fields and Dynamic Monte Carlo Methods: A Mathematical Introduction 2nd ed.*, Springer-Verlag. ISBN 978-354-0442-13-4, Berlin



Remote Sensing - Advanced Techniques and Platforms

Edited by Dr. Boris Escalante

ISBN 978-953-51-0652-4
Hard cover, 462 pages
Publisher InTech
Published online 13, June, 2012
Published in print edition June, 2012

This dual conception of remote sensing brought us to the idea of preparing two different books; in addition to the first book which displays recent advances in remote sensing applications, this book is devoted to new techniques for data processing, sensors and platforms. We do not intend this book to cover all aspects of remote sensing techniques and platforms, since it would be an impossible task for a single volume. Instead, we have collected a number of high-quality, original and representative contributions in those areas.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Rolando D. Navarro, Jr., Joselito C. Magadia and Enrico C. Paringit (2012). Estimation of the Separable MGMRF Parameters for Thematic Classification, Remote Sensing - Advanced Techniques and Platforms, Dr. Boris Escalante (Ed.), ISBN: 978-953-51-0652-4, InTech, Available from: http://www.intechopen.com/books/remote-sensing-advanced-techniques-and-platforms/estimation-of-the-separable-multivariate-gauss-markov-random-field-parameters-for-thematic-class

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821