Multi-dimensional Multivariate Gaussian
Markov Random Fields with
Application to Image Processing

K. V. MARDIA

University of Leeds, Leeds, United Kingdom
Communicated by L. D. Pitt

Recently, numerous practical applications of multivariate Gaussian Markov
random fields (GMRF) on a lattice have emerged. However, the theory is not
satisfactorily developed. We give various properties of multivariate GMRF for
multi-dimensional lattice. In particular, some multivariate MRF are given. We
discuss estimation procedures and give a numerical example from the area of image
processing.

1. INTRODUCTION

Various applications of multivariate Gaussian Markov random fields
(GMRF) on 2-dimensional lattice have appeared in image processing; see

We give here some important properties of multivariate GMRF for
multi-dimensional lattices. For the univariate MRF in general, we refer to
Besag [2] and Bartlett [1]. A multivariate GMRF for a 2-dimensional
infinite lattice has been mentioned in Kunsch’s thesis [17] but the exten-
sion is restricted to a diagonal conditional covariance matrix (see Kittler
and Fögelin [16] for a readily accessible statement of the result).

In Sections 2 and 3, we first give some properties of the multi-normal
distribution which underlies the properties of the multivariate GMRF. Section 4 gives some properties of the stationary MRF for finite lattices,
whereas Section 4 gives the properties for infinite lattices. In Section 4, we
give a sufficient condition for existence of a Gaussian MRF in some
generality which removes some doubts about their existence (see, for example, Fu and Yu [10] and Woods [25]). In Section 5, we consider some specific multivariate MRF. In Section 6 some estimation problems are discussed and a numerical example is given.

2. THE CONDITIONAL AND THE COVARIANCE STRUCTURE OF THE MULTI-NORMAL DISTRIBUTION

We first consider the conditional structure of the multi-normal distribution. Suppose that \( x_1, \ldots, x_n \) are \( p \)-dimensional conditional normal variates with

\[
E(x_i | R_i) = \mu_i + \sum_{j \neq i} \beta_{ij} (x_j - \mu_j), \quad i = 1, \ldots, n
\]

and

\[
\text{var}(x_i | R_i) = \Gamma_i, \quad i = 1, \ldots, n,
\]

where \( R_i \) stands for the rest of the variables, namely \( \{x_j, j \neq i\} \). We assume throughout that the matrices \( \Gamma_i \) are positive definite. For convenience, we take

\[
\beta_{ii} = -1.
\]

Let \( x = (x_1', \ldots, x_n') \), which is a vector of dimension \( np \). Given the \( n \) conditional distributions \( x_i | R_i \), we can ask if the (joint) distribution of \( x \) is normal. Further, what conditions should be imposed on the \( \beta_{ij} \) so that the conditional distributions are self-consistent? These questions are answered in the following theorem.

THEOREM 2.1. Given the \( n \) conditional multi-normal distributions, \( x \) is \( N_{np}(\mu, \Sigma) \), where

\[
\mu' = (\mu_1', \ldots, \mu_n'), \quad \Sigma = \{\text{Block}(-\Gamma_i^{-1} \beta_{ij})\}^{-1}
\]

provided

\[
(i) \quad \beta_{ij} \Gamma_i = \Gamma_j \beta_{ji}
\]

and

\[
(ii) \quad \{\text{Block}(-\Gamma_i^{-1} \beta_{ij})\} \text{ or } \text{Block}(-\beta_{ij}) \text{ is p.d.}
\]
Further, the p.d.f. of \( x \) is

\[
(2\pi)^{-n/2} \left\{ \prod_{i=1}^{n} |\Gamma_i| \right\}^{-1/2} \left| \text{Block}(-\beta) \right| \times \exp \left\{ \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} (x_i - \mu_i)' \Gamma_i^{-1} \beta_j (x_j - \mu_j) \right\}.
\]

(2.6)

**Proof.** We first note that the Brook expansion [6] extends immediately to the multivariate case, i.e., if \( x = (x_1', \ldots, x_n')' \) and \( y = (y_1', \ldots, y_n')' \) are identically distributed with the joint p.d.f. \( f(\cdot) \), we have

\[
f(x)/f(y) = \prod_{i=1}^{n} f(x_i|x_1, \ldots, x_{i-1}, y_{i+1}, \ldots, y_n)/f(y_i|x_1, \ldots, x_{i-1}, y_{i+1}, \ldots, y_n).
\]

(2.7)

Without any loss of generality take \( \mu_i = 0 \). After some algebra, we find from (2.1) and (2.2) that (2.7) simplifies for \( y = 0 \) to

\[
-2 \log \frac{f(x)}{f(0)} = \sum_{i=1}^{n} x_i \Gamma_i^{-1} x_i - 2 \sum_{i=2}^{n} \sum_{j=1}^{i-1} x_i \Gamma_i^{-1} \beta_j x_j.
\]

(2.8)

We can write the “forward version” of the Brook expansion (2.7) as

\[
f(x)/f(y) = \prod_{i=1}^{n} f(x_i|y_1, \ldots, y_{i-1}, x_{i+1}, \ldots, y_n)/f(y_i|y_1, \ldots, y_{i-1}, x_{i+1}, \ldots, y_n),
\]

which leads in the same manner as (2.8) to

\[
-2 \log \left[ \frac{f(x)}{f(0)} \right] = \sum_{i=1}^{n} x_i \Gamma_i^{-1} x_i - 2 \sum_{i=1}^{n-1} \sum_{j=i}^{n} x_i \Gamma_i^{-1} \beta_j x_j.
\]

(2.9)

Since (2.8) and (2.9) must be identical, the coefficients \( \beta_j \) must satisfy (2.6). Further, from (2.9), we have

\[
-2 \log f(x) = \text{Const.} + \sum_{i=1}^{n} x_i \Gamma_i^{-1} x_i - \sum_{i \neq j}^{n} x_i \Gamma_i^{-1} \beta_j x_j.
\]

Thus \( x \) is \( N_{np}(0, \Sigma) \), where \( \Sigma \) is as defined at (2.3), provided that \( \Sigma \) is p.d., i.e., if (ii) holds. Note that the normalizing constant simplifies since

\[
\text{Block}(-\Gamma_i^{-1} \beta_j) = \{ \text{Block diag}(\Gamma_1, \ldots, \Gamma_n) \} \{ \text{Block}(-\beta) \}.
\]

**Corollary 1.** If

\[
\text{var}(x_i|R_i) = \Gamma, \quad \Gamma > 0,
\]

...
then the conditions (2.4) and (2.5) are

(i) $\beta_y \Gamma = \Gamma \beta_y'$,

(ii) Block($-\beta_y$) is p.d.

Further, in (2.6),

$$\prod_{i=1}^{n} |\Gamma_i|^{-1/2} = |\Gamma|^{-(1/2)n}. \quad (2.12)$$

**Corollary 2 (Factorization Case).** If

$$\beta_y = \beta_y I \quad \text{and} \quad \beta_y = -1, \ \beta_y = \beta_y,$$

then

$$\text{Cov}(x) = -\Gamma^{-1} \otimes \beta,$$

which is p.d. if $-\beta$ is p.d.

The conditions on $\beta_y$ in Corollary 2 are identical to those for the univariate case given by Besag [2].

We now look at the converse of Theorem 2.1.

**Theorem 2.2.** Let $x = (x_1', \ldots, x_n')'$ be $N_{np}(\mu, \Sigma)$, where

$$\mu = (\mu_1', \ldots, \mu_n')', \quad \Sigma = (\Sigma^\mu)^{-1},$$

with $\Sigma$ a block matrix with each element of order $p \times p$. If only the $\Sigma^\mu$ are given, we then have

$$E(x_i | R_i) = \mu_i + \sum_{j \neq i} \beta_{ij}(x_j - \mu_j), \quad (2.13)$$

where

$$\beta_{ij} = -(\Sigma^\mu)^{-1} \Sigma^{ij}, \quad i \neq j. \quad (2.14)$$

Further

$$\text{var}(x_i | R_i) = (\Sigma^\mu)^{-1}. \quad (2.15)$$

**Proof.** Let $(y_1, y_2)$ be $N[(v_1, v_2), (\Lambda^\mu)^{-1}]$. Then it can be shown after some manipulation of standard results (see, for example, Mardia et al. [19, p. 63]) that

$$y_1 | y_2 \sim N[v_1 - (\Lambda^{11})^{-1} \Lambda^{12}(y_2 - v_2), (\Lambda^{11})^{-1}]. \quad (2.16)$$
We can take \( i = 1 \) in Theorem 2.2 without any loss of generality. On putting
\[
\begin{align*}
y_1 &= x_1, \\
y_2' &= (x_2', \ldots, x_n'),
\end{align*}
\]
we then have
\[
\Lambda^{11} = \Sigma^{11}, \quad \Lambda^{12} = (\Sigma^{12}, \ldots, \Sigma^{1n}), \quad \nu_1 = \mu_1, \quad \nu_2' = (\mu_2', \ldots, \mu_n').
\]
Substituting these in (2.16), we get Theorem 2.2 after some algebra. Note that we can also prove Theorem 2.1 by noting from (2.8) that \( x \) is normal and then use (2.16) to identify the parameters. For example, from (2.3) and (2.15), we have \( \Gamma_i^{-1} = \Sigma''_i \).

We now investigate the relationship of \( \beta_{ij} \) and the conditional independence of \( x_i \) and \( x_j \) given all the other variables.

**Theorem 2.3.** For p.d. \( \Sigma''_i, \Sigma''_j \), we have
\[
x_i \perp x_j | \text{rest if and only if } \Sigma''_i = 0.
\]

**Proof.** Write the log of the joint density of \( x \), and note that the term involving both \( x_i \) and \( x_j \) is simply
\[
-\frac{1}{2} x_i' \Sigma''_{ij} x_j.
\]
Hence \( x_i \) is independent of \( x_j \) given the rest when \( \Sigma''_i = 0 \).

However, it is more revealing to consider the covariance of \( x_1 \) and \( x_2 \) given \( z_3 = (x'_1, \ldots, x'_n)' \), i.e., to establish
\[
\text{cov}(x_1, x_2 | z_3) = 0 \iff \Sigma^{12} = 0.
\]
In fact, for any \((x, y)\),
\[
\text{var}(x | y) = (V^{11})^{-1}, \quad \text{cov}(x | y) = -\text{var}(x | y) V^{12} \text{var}(y),
\]
where
\[
V^{-1} = (V''_{ij}). \quad i, j = 1, 2.
\]
Hence, it can be seen that
\[
\begin{align*}
\text{var}(x_1 | R_1) &= (\Sigma^{11})^{-1} \\
\text{cov}(x_1, x_2 | z_3) &= -\text{var}(x_1 | (x_2, z_3)) \Sigma^{12} \text{var}(x_2 | z_3) \\
&= -\text{var}(x_1 | z_3) \Sigma^{12} \text{var}(x_2 | x_1, z_3).
\end{align*}
\]
Equation (2.19) clearly establishes (2.18).
Corollary 2.3.1. \( x_i \perp x_j \text{ rest if and only if } \beta_{ij} = 0 \).

Proof. A proof follows on using (2.17) in (2.14).

This corollary is, indirectly, a particular case of the Hammersley–Clifford theorem for a Gaussian process.

3. Stationary Markov Random Fields for Finite Lattices

Let \( \mathbb{Z}^d \) be the \( d \)-dimensional infinite integer lattice and let \( D \) denote a finite rectangular lattice within \( \mathbb{Z}^d \). Let \( \mathbf{r} = (r_1, \ldots, r_d) \) be any point in \( D \) and assume that there are \( n \) points/sites in \( D \) so that we can write

\[
D = \{ \mathbf{r}(i), i = 1, 2, \ldots, n \}.
\]

At each point \( \mathbf{r} \in D \), suppose a \( p \)-dimensional observation \( \mathbf{x}_r \in \mathbb{R}^p \) is made. Sometimes we will write \( \mathbf{x}_r = x_{r_{1,\ldots,r_d}} \) without the brackets for the vector \( \mathbf{r} \).

Let

\[
N = \{ s(1), \ldots, s(m) \}
\]

be a set of \( m \) neighbours of the origin in \( \mathbb{Z}^d \). We assume that \( N \) has the symmetry property

\[
s \in N \Rightarrow -s \in N.
\]

For example, with \( d = 2 \) and \( m = 4 \), we may have

\[
N = \{ (-1, 0), (1, 0), (0, -1), (0, 1) \}. \tag{3.1}
\]

Suppose that \( \mathbf{r} \) denotes a typical point in \( D \). The set

\[
N_r = \{ \mathbf{r} + s(1), \ldots, \mathbf{r} + s(m) \}
\]

then represents the neighbourhood of any point \( \mathbf{r} \in D \). A stationary MRF \( \{ \mathbf{x}_r \} \) in \( p \) dimensions is defined by the property

\[
P(\mathbf{x}_r | \mathbf{R}_r) = P(\mathbf{x}_r | \mathbf{x}_r, \mathbf{t} \in N_r).
\]

A stationary Gaussian MRF is defined by

\[
E(\mathbf{x}_r | \mathbf{R}_r) = \mu + \sum_{s \in N} (\beta_s - \mu) \mathbf{x}_{r+s} \tag{3.2}
\]

\[
\text{var}(\mathbf{x}_r | \mathbf{R}_r) = \Gamma, \tag{3.3}
\]

except for the boundary points \( \mathbf{r} \subset D \) (see below for adjustments). For some comments on our terminology, see Section 4.
Identifying (3.2) with (2.1), and writing $\beta^{*}_{ij}$ for $\beta_{ij}$ in (2.1), we note that if the sites $i$ and $j$ are neighbours then from Corollary 2.3.1, we have

$$\beta^{*}_{ij} = \beta_{r(i) - r(j)}, \quad r(i) - r(j) \in N$$  \hspace{1cm} (3.4)

and, if the sites $i$ and $j$ are not neighbours, then

$$\beta^{*}_{ij} = 0.$$  \hspace{1cm} (3.5)

Hence, for the existence of a homogeneous GMRF, we obtain from (2.10) and (2.11) the necessary conditions:

(i) $\beta, \Gamma = \Gamma \beta^{*}$  \hspace{1cm} (3.6)

(ii) Block($-\beta^{*}$) is p.d.  \hspace{1cm} (3.7)

**Remark.** In MRF we group the set of variables for $x$, where $\beta_{ij} \neq 0$. The neighbourhood structure can be put in terms of “cliques,” which are collections of sites in which every site is a neighbour of other sites (see Besag [2]). There are various other ways we can group variables which may not form a MRF as such, although there is some “link” (see Darroch et al. [9]). This grouping is again a consequence of Corollary 2.3.1, although we do not discuss it in detail here.

**Boundary Adjustments.** Since $n$ is finite, (3.2) is not defined for the boundary values of $D$, e.g., for $d = 1$, $m = 2$, $N = \{(-1), (1)\}$, and $D = (1, \ldots, n)$, the expectation

$$E(x_r | R_r) = \mu + \beta_{-1}(x_{r-1} - \mu) + \beta_1(x_{r+1} - \mu)$$  \hspace{1cm} (3.8)

is not defined for $r = 1$ and $r = n$. To overcome this problem, we define

$$f(\{x_r\}) \equiv f(\{x_r\} \mid \text{boundary} = \mu).$$  \hspace{1cm} (3.9)

It can be seen that the joint normal distribution of $x$ thus defined has

$$\Sigma^{-1}_r = \{\text{cov}(x_r, x_s)\}^{-1} = (\Sigma^r),$$

where

$$\Sigma^r = -\Gamma^{-1} \beta_{-s}, \quad r - s \in N,$$

$$= 0 \quad \text{otherwise.}$$

This matrix is now a Block Toeplitz matrix. For example, for a first-order Markov process, $\Sigma^{-1}_r$ has the elements

$$\sigma'' = 1/\tau^2; \quad \sigma^{r+1} = -\beta/\tau^2, \quad r = 1, 2, \ldots, n - 1,$$

$$\sigma^{r-1} = -\beta/\tau^2, \quad r = 2, \ldots, n, \sigma^{r} = 0, \text{otherwise.}$$
An alternative way of making boundary adjustments is to consider the marginal distribution of \((x_1, \ldots, x_n)\) from \((x_0, x_1, \ldots, x_n, x_{n+1})\), then
\[
\sigma^{11} = \sigma^{nn} = 1/(1 - \lambda^2), \quad \beta = \lambda/(1 + \lambda^2),
\]
but now \(\Sigma^{-1}\) is not Toeplitz as the diagonal elements are not equal in this case.

We now consider the infinite lattice case to understand some other important properties of the stationary MRF.

4. Stationary Gaussian MRF on Infinite Lattice

We assume again that the GMRF is defined by (3.2) and (3.3) for any \(r \in \mathbb{Z}^d\). Further, the process is stationary, i.e.,
\[
E(x_r x'_{r+h}) = \Sigma_h,
\]
where we will assume \(\mu = 0\) without any loss of generality. We give some properties of the GMRF.

We first obtain the matrix generating function of the process, which is defined by
\[
G(x)(Z) = \sum_{h=\infty}^{\infty} \Sigma_h Z_1^{h_1} \cdots Z_d^{h_d}, \quad (4.1)
\]
where \(Z = (Z_1, \ldots, Z_d)'\), \(h = (h_1, \ldots, h_d)'\). For this we require the following recurrence relationships which are proved by extending the proof of Moran [20].

**Theorem 4.1.** We have
\[
\Sigma_0 = \Gamma + \sum_{s \in N} \beta_s \Sigma_s = \Gamma + \sum_{s \in N} \beta_s \Sigma_{-s} \quad (4.2)
\]
and
\[
\Sigma_h = \sum_{s \in N} \beta_s \Sigma_{h-s}, \quad h \neq 0. \quad (4.3)
\]

**Proof.** Recall that \(R_r\) denotes remaining variables excluding \(x_r\). Then for \(h \neq 0\),
\[
\Sigma_h = E_{R_r}\{E(x_r \mid R_r) x'_{r+h}\},
\]
which using (4.2) becomes (4.3). For \( h = 0 \), we use

\[
\Sigma_0 = \text{Var}(x_r) = E_{R_r} \text{Var}(x_r \mid R_r) + V_{R_r} E(x_r \mid R_r).
\]

The first term in the RHS from (4.3) is \( \Gamma \), where, since \( E(x_r) = 0 \), we have

\[
V_{R_r} E(x_r \mid R_r) = E_{R_r} \left[ E(x_r \mid R_r) E(x_r' \mid R_r) \right]
\]

\[
= E_{R_r} \left[ \sum_{s \in N} \beta_s x_{s+r} E(x_r' \mid R_r) \right]
\]

\[
= E_{R_r} \left[ \sum_{s \in N} \beta_s E(x_{s+r} x_r' \mid R_r) \right]
\]

since \( x_{s+r} \) is part of \( R_r \). Hence (4.2) follows.

**Theorem 4.2.** Under certain conditions, we have for the GMRF on the infinite lattice

\[
G_x(Z) = \{ A(Z) \}^{-1} \Gamma,
\]

where

\[
A(Z) = I - \sum_{s \in N} \beta_s Z_1^s \cdots Z_d^s.
\]

*Proof.* We have by definition

\[
G_x(Z) = \Sigma_0 + \sum_{h \neq 0} \Sigma_h Z_1^h \cdots Z_d^h.
\]

Writing \( \Sigma_0 \) and \( \Sigma_h \) (\( h \neq 0 \)) from Theorem 4.1,

\[
G_x(Z) = \Gamma + \sum_{h} \sum_{s \in N} \beta_s \Sigma_{h-s} Z_1^h \cdots Z_d^h.
\]

Changing the order of summation and using the transformation \( m = h - s \) for fixed \( s \), we have

\[
G_x(Z) = \Gamma + \sum_{s \in N} \sum_{m \in M} \beta_s \Sigma_{m+s} Z_1^{m+s} \cdots Z_d^{m+s}.
\]

Hence provided that \( \{ A(Z) \}^{-1} \) exists, the proof follows.

However, for \( G_x(Z) \) to be a matrix generating function we require
further conditions, as we will see through the spectral density matrix of the process defined by

$$F_X(\mathbf{\theta}) = \sum_{n=-\infty}^{\infty} \mathbf{\Sigma}_n e^{-i\mathbf{h}^T \mathbf{\theta}},$$

where $\mathbf{\theta} = (\theta_1, ..., \theta_d)^T$, $\theta_i \in (-\pi, \pi)$.

On writing $Z_j = e^{i\theta_j}$ in Theorem 4.2, the spectral density matrix for the GMRF is given by

$$F_X(\mathbf{\theta}) = \left( I - \sum_{s \in \mathcal{N}} \mathbf{\beta}_s e^{i\mathbf{\theta}^T \mathbf{s}} \right)^{-1} \mathbf{\Gamma}.$$  \hspace{1cm} (4.4)

Note that $F_X(\mathbf{\theta})$ is Hermitian under (3.6),

$$\mathbf{\beta}_s \mathbf{\Gamma} = \mathbf{\Gamma}^T \mathbf{\beta}_{-s}.$$  

We now give a result which is a suitable extension of the Herglotz theorem for our purpose.

**Theorem 4.3.** Let $F(\mathbf{\theta}) = (f_{ij}(\mathbf{\theta}))$ be a $p \times p$ Hermitian matrix which is a periodic function of $\mathbf{\theta} \in (-\pi, \pi)^d$. If

(i) $F(\mathbf{\theta})$ is a p.d. matrix for each $\mathbf{\theta}$,  \hspace{1cm} (4.5)

(ii) $F(\mathbf{\theta})' = F(-\mathbf{\theta})$,  \hspace{1cm} (4.6)

(iii) $\int_{\mathbf{\theta} \in (-\pi, \pi)^d} \varphi(\mathbf{\theta}) \, d\mathbf{\theta} < \infty$,

where $\varphi(\mathbf{\theta})$ is the largest eigenvalue of $F(\mathbf{\theta})$ for each $\mathbf{\theta}$, then there exists a Gaussian process with $F(\mathbf{\theta})$ as its spectral density matrix.

**Proof.** To prove the theorem, it is sufficient to establish that there exist real matrices

$$\mathbf{\Sigma}_h = \frac{1}{(2\pi)^d} \int_{(-\pi, \pi)^d} F(\mathbf{\theta}) e^{-i\mathbf{h}^T \mathbf{\theta}} \, d\mathbf{\theta}, \quad h \in \mathbb{Z}^d,$$  \hspace{1cm} (4.7)

such that the block matrix of order $np \times np$,

$$\mathbf{\Sigma}^* = (\mathbf{\Sigma}_{ij}^*), \quad i, j = 1, ..., n,$$

is p.d. for all $n$, where

$$\mathbf{\Sigma}_{ij}^* = \mathbf{\Sigma}_{r(i)-r(j)}, \quad r(i) \in \mathbb{Z}^d.$$
Details of a proof of this step are given in Mardia [18]. The referee has pointed out that a proof for its continuous analogue is given in Gihman and Skorohod [12, Theorem 5, p. 216]. However, note that the conditions (i) and (iii) imply that \( f_{v}(\theta) \) is integrable. Now, the proof can be completed from the Daniell-Kolmogorov theorem on the existence of a stationary Gaussian process with covariances \( \{\Sigma_{b}\} \).

Note that we have followed here the terminology of Bartlett [1] and Besag [2]. Our Markov random fields are L-fields in the terminology of Rozanov [23] rather than so-called L-Markov fields. Chay [7] examines some deeper relationship between L-fields and L-Markov fields. Note that difficult technical questions arise when either \( \Gamma \) is singular and/or \( F(\theta) \) is semi-p.d., which our formulation avoids; see Chay [7] and Pitt [21]; cf. Brillinger [5], Cliff and Ord [8] and Yaglom [26].

For \( p = d = 1 \), condition (i) of Theorem 4.3 becomes that the polynomial

\[
F(\theta) = 1 - \sum_{s \in \mathbb{N}} \beta_{s} \cos s\theta > 0, \quad \forall \theta.
\]

A sufficient condition for this to hold is

\[\Sigma |\beta_{s}| < 1, \quad (4.8)\]

which is easier to verify in constructing a MRF. The next theorem gives similar sufficient conditions.

Before proceeding to the next theorem, we define the norm (spectral) \( \|A\| \) of any square matrix \( A \) as

\[\|A\| = (\text{maximum eigenvalue of } A^*A)^{1/2},\]

where \( A^* \) denotes the complex conjugate of \( A \). For scalar \( x \), the norm is \( |x| \), which is the absolute value for real \( x \) and the modulus for complex \( x \).

**Theorem 4.4.** Let \( \beta_{s}, s \in \mathbb{N}, \) and \( \Gamma \) satisfy

(i) \( \Gamma \) is p.d. with bounded norm,

(ii) \( \beta_{-s} = \Gamma \beta_{s} \Gamma^{-1}, \quad (4.9)\)

and

(iii) \( \sum_{s \in \mathbb{N}} \lambda_{s} < 1, \quad (4.10)\)

where \( \lambda_{s} = \|\Gamma^{-1/2}\beta_{s} \Gamma^{-1/2}\|, \) with \( \Gamma^{1/2} \) as the symmetric square root of \( \Gamma \). Then there exists a stationary GMRF, defined by (3.2) and (3.3), with spectral density matrix (4.4).
Proof. We need only verify the conditions (i) and (iii) of Theorem 4.3 for the spectral density matrix $F_x(\theta) = F$, say, given by (4.4). We have

$$F = \Gamma^{-1/2} A^{-1} \Gamma^{-1/2},$$

(4.11)

where

$$A = I - B, \quad B = \sum_{s \in \mathcal{N}} \beta_s \cos \theta_s \sin \theta_s,$$

with

$$\beta_s = \frac{\gamma}{\sqrt{2}} p_s - 1, I \gamma = \lambda^2 A'.$$

Let $\lambda_1, \ldots, \lambda_p$ be the eigenvalues of $A$ in ascending order. Suppose that $\delta_1, \ldots, \delta_p$ are the absolute eigenvalues of $B$ in ascending order. It can be seen that

$$\lambda_1 \geq 1 - \delta_p.$$  (4.12)

We now show that

$$1 - \delta_p > 0.$$  (4.13)

We have

$$\delta_p = \|B\| = \sum_{s \in \mathcal{N}} \|\beta_s \cos \theta_s \sin \theta_s\| \leq \sum \|\beta_s \cos \theta_s \sin \theta_s\| \leq \sum \|\beta_s \| \|e^{i\theta_s}\| = \sum \lambda_s.$$  

Hence from (4.10), (4.13) follows. Thus $\alpha_1 > 0$ so that $A$ is p.d., which from (4.11) implies $F$ is p.d.

Now

$$\varphi(\theta) = \|F\| < \|\Gamma\| \|A^{-1}\| = \|\Gamma\|/\alpha_1 < \|\Gamma\|/(1 - \delta_p) < \|\Gamma\|/\left(1 - \sum \lambda_s\right),$$

which is bounded by (4.10). Hence condition (iii) of Theorem 4.3 is also satisfied.

Note that on substituting (4.4) in (4.5), we find that the auto-covariance matrix of the GMRF is given by

$$\Sigma_h = (2\pi)^{-d} \left\{ \int_{(0,\pi)^d} \left( I - \sum_{s \in \mathcal{N}} \beta_s e^{i\theta_s \cdot \phi} \right)^{-1} e^{ih \cdot \phi} \right\} \Gamma.$$  (4.14)

We now consider some particular GMRFs.

**Factorized Case.** For

$$\beta_s = \beta_s I,$$
the conditions of Theorem 4.4 reduce to the following well-known univariate conditions (see, Rozanov [22]):

\[ \beta_{-s} = \beta_s, \quad \sum_{s \in N} |\beta_s| < 1. \]

Also from (4.14), we have

\[ \Sigma_b = (V_b) \Gamma, \]

where

\[ V_b = (2\pi)^{-d} \int_{(-\pi,\pi)^d} \left( 1 - \sum_{s \in N} \beta_s \cos s' \theta \right)^{-1} \cos h' \theta \, d\theta. \]

These functions have already been computed for \( d = 2 \), for some important cases (see Besag [4]). Note that for \( p = 1 \), we have

\[ \sigma^2 = (2\pi)^{-d} \gamma^2 V_o, \]

where \( \sigma^2 \) and \( \gamma^2 \) are the total and conditional variances, respectively. Also the auto-correlations are simply \( V_b/V_o \).

**The Reversible MRF.** We call the MRF reversible if the joint distribution of \((x_{r1}, \ldots, x_{rn})\) and \((x_{hr1}, \ldots, x_{hrn})\) is the same for all \( r_1, \ldots, r_n \) and \( h \). We will not discuss its modelling implication, but for the univariate case see Kelly [14, p. 51. It can be seen that the GMRF is reversible only if

\[ \begin{align*}
&(i) \quad F_x(\theta) \text{ is real or } F_x(\theta) \text{ is symmetric or} \\
&(ii) \quad \Sigma_b \text{ is symmetric.}
\end{align*} \]

Thus we find from (4.4) that

\[ \beta_s = \beta_{-s}, \quad \beta_s \Gamma = \Gamma \beta_s. \quad (4.15) \]

Then

\[ F_x(\theta) = \left( I - \sum_{s \in N} \beta_s \cos \theta's \right)^{-1} \Gamma. \]

Also then

\[ \lambda_s = \{ \text{maximum eigenvalue of } \beta_s^2 \}^{1/2}, \]

so that \( \lambda_s \) does not depend on \( \Gamma \) but the matrix \( \Gamma \) has still to satisfy (4.15). Further, if the regression parameter matrices \( \beta_s \) are symmetric then

\[ \beta_s = \beta_{-s} = \beta_s', \quad \beta_s \Gamma = \Gamma \beta_s. \quad (4.16) \]
Hence

$$\lambda_s = \text{maximum absolute eigenvalue of } \beta_s.$$ (4.17)

Note that $\Gamma$ and $\beta$ as equi-correlation matrices satisfy (4.16) and we will call this case the "balanced case."

Other important classes are "reflection symmetric" MRFs, but we will not require them here.

**Parameters.** The number of parameters for the multivariate case is very large indeed. If $\Gamma$ is fixed, note that $\beta_s$ contains $p^2$ parameters, $\beta_s$ is defined by (4.9). For the reversible case $\beta^*_s = \beta^*_r$, and therefore there are $p(p+1)/2$ parameters.

Further, for symmetric regression parameter matrices, we have

$$\beta_s = P' \Delta P \quad \text{if} \quad \Gamma = P' \Delta P,$$

where $P$ is an orthogonal matrix. Thus the number of parameters in $\beta_s$ reduces to $p$, equal to the number of diagonal elements of $\Lambda$.

5. **Specific GMRF**

We now consider various specific cases where we will assume that the condition (4.9) of Theorem 4.4 is fulfilled but the stationarity condition

$$\sum_{s \in N} \|\Gamma^{-1/2} \beta_s \Gamma^{-1/2}\| < 1 \quad (5.1)$$

needs to be simplified. We will first deal with GMRF for any $p$ but $d = 1$ and $m = 2$. Later it will be seen that this case leads to sufficient stationarity conditions for $d = 2$.

We have

$$E(x_r | R_r) = \beta_{-1} x_{r-1} + \beta_1 x_{r+1}, \quad \text{var}(x_r | R_r) = \Gamma. \quad (5.2)$$

Note that this can be expressed in terms of the multivariate AR1. We can write a multivariate AR1 as (see Jenkins and Watts [13, p. 473])

$$x_r = \alpha x_{r-1} + \epsilon_r,$$

with spectral density matrix (SDM)

$$[I - \alpha e^{it}]^{-1} W [I - \alpha' e^{-it}]^{-1}.$$
where $W$ is a p.d. matrix. Indeed, on equating this SDM to that given by (5.2), we get

$$
\Gamma^{-1} \mathbf{b}_1 = W^{-1} \mathbf{a}, \quad \Gamma^{-1} \mathbf{b}_{-1} = \mathbf{a}'W^{-1}
$$

and

$$
\Gamma^{-1} = W^{-1} + \mathbf{a}'W^{-1}\mathbf{a}.
$$

Hence there is a one-to-one correspondence. Such a “causal” relationship is well known for $p = 1$ and $m = 2$.

Note that the MRF may not be reversible. Thus there exists a process for $p = 2$,

$$
E(x_0^{(1)} | \cdot) = \beta x_1^{(2)}, \quad E(x_0^{(2)} | \cdot) = \beta x_1^{(1)}, \quad |\beta| < \frac{1}{2},
$$

where $x' = (x_1^{(1)}, x_1^{(2)})$.

It can be shown from the Fourier transform (4.14) that for the reversible MRF with $p = 2$, and $\beta_{11} = \beta_{22} = \beta_1$, $\beta_{21} = \beta_{22} = \beta_2$,

$$
\sigma_{11}(h) = \sigma_{22}(h) = \frac{1}{4} \left\{ \sigma_1(h) + \sigma_2(h) \right\}
$$

$$
\sigma_{12}(h) = \sigma_{21}(h) = \frac{1}{4} \left\{ \sigma_1(h) - \sigma_2(h) \right\},
$$

where

$$
\sigma_i(h) = \left\{ (1 + \lambda_i^2)/(1 - \lambda_i^2) \right\} \hat{\lambda}_i[h], \quad |\hat{\lambda}_i| < 1, \quad i = 1, 2,
$$

and

$$
\beta_1 + \beta_2 = \lambda_1/(1 + \lambda_1^2) \quad \text{and} \quad \beta_1 - \beta_2 = \lambda_2/(1 + \lambda_2^2).
$$

**Stationarity Conditions for $p = 2$.** For general $\Gamma$,

$$
\Gamma = \begin{pmatrix} 1 & e \\ e & 1 \end{pmatrix}, \quad |e| < 1
$$

$$
\mathbf{b}_1 = \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix},
$$

it can be seen that we must have

$$
|f| + (f^2 - 4g^2)^{1/2} < (1 - e^2)/2,
$$

where

$$
f = a^2 + b^2 + c^2 + d^2 - 2e^2(ad + bc) + 2e(a - d)(b - c)
$$
and 

\[ g = (ad - bc)(1 - e^2). \]

Various particular cases can be obtained. Note that for the factorized case, \(|a| < 1\).

For the reversible MRF case either

(i) \( e = (c - b)/(a - d) \) if \( a \neq d \) or

(ii) \( a = d, \ b = c \) with \( e \) as a free parameter.

The case (ii) is "balanced." The condition under (i) is

\[ a^2 + d^2 + 2|bc| + |a + d| \left\{ (a - d)^2 + 4bc \right\}^{1/2} < \frac{1}{2} \]

and under (ii)

\[ |a| + |b| < \frac{1}{2}. \] (5.7)

It can be shown that this condition holds with \( \lambda_1 \) and \( \lambda_2 \) defined by (5.5).

For the balanced case with

\[ \Gamma = (\gamma_{rs}), \quad \gamma_{rs} = \gamma^2, \ r = s; = \gamma_{12}, \ r \neq s; \ r, s = 1, \ldots, p \]

and

\[ \beta_1 = (\beta_{1,r,s}) \quad \text{with} \quad \beta_{1,r,s} = \beta_1, \ r = s; = \beta_2, \ r \neq s, \]

the stationarity condition becomes

\[ |\beta_1| + (p - 1) |\beta_2| < \frac{1}{2}. \] (5.8)

Hence

\[ \| \beta_1 \| = \max \{ |\beta_1 + (p - 1) \beta_2|, |\beta_1 - \beta_2| \} \]

and

\[ \| \beta_1 \| < \frac{1}{2}. \]

A less restrictive condition is simply to select \((\beta_1, \beta_2)\) subject to the following constraints:

\begin{align*}
\text{for } \beta_2 \in (0, 1/p), & \text{ we see that admissible } \beta_1 \in (\beta_2 - \frac{1}{2}, \frac{1}{2} - (p - 1) \beta_2), \\
\text{while } \beta_2 \in (-1/p, 0), & \text{ we have admissible } \beta_1 \in (\frac{1}{2} + \beta_2, -\frac{1}{2} - (p - 1) \beta_2).
\end{align*}

We can use the above restrictions for \( d > 1 \) as we now illustrate.
Multivariate Planar GMRF. Consider \( d = 2 \) with
\[
N = \{(-1,0), (1,0), (0,-1), (0,1)\}.
\]
We have
\[
E(x_{r_1,r_2} | \cdot) = \beta_{-1,0} x_{r_1-1,r_2} + \beta_{1,0} x_{r_1+1,r_2} + \beta_{0,-1} x_{r_1,r_2-1} + \beta_{0,1} x_{r_1,r_2+1}.
\]
For the "balanced" case, \( \Gamma, \beta_{1,0}, \text{ and } \beta_{0,1} \) are equi-correlation matrices with diagonal and non-diagonal elements as, say, \((\gamma_1, \gamma_2), (\beta_1, \beta_2), \text{ and } (\beta_3, \beta_4)\). It can be seen that, as in (5.8) for \( d = 1 \), the stationarity conditions become
\[
|\beta_1| + |\beta_2| + (p-1)(|\beta_3| + |\beta_4|) < \frac{1}{2}.
\]
As in (5.7), we can also write the condition explicitly for \( p = 2 \) with \( \beta_{1,0} = \beta_{0,-1} \), i.e., the right-hand inequality in (5.6) becomes \( \frac{1}{4} \) rather than \( \frac{1}{2} \). (This model measures a departure from the "balanced case.") In this way, we can derive sufficient conditions for any general case from \( d = 1 \) and \( m = 2 \).

6. AN APPLICATION TO IMAGE PROCESSING

Suppose that we are given a training data of mixed pixels \( \{y_r\} \) for \( d = 2 \), i.e., \( y_r \) are, say, \( p + 1 \) proportions, \( 1' y_r = 1 \). Using the logistic transformation,
\[
\log \frac{y_r^{(i)}}{y_r^{(p+1)}} = x_r^{(i)}, \quad i = 1, \ldots, p,
\]
we may assume \( \{x_r\} \) to have a Gaussian MRF (see Kent and Mardia [15]). The aim is to estimate the parameters of this MRF. For simplicity, we take \( p = 2 \) and
\[
N = \{(1,0), (-1,0), (0,1)(0,-1)\}.
\]
For invariance of the model under permutations of the labels of the proportions, we require
\[
\Gamma = \tau^2 \left( \begin{array}{cc} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{array} \right).
\]
Further, to reduce the number of the parameters, we take reversible MRF of the type
\[
E(x_r | \cdot) = \sum_{s \in N} x_{r+s}.
\]
Since we require $\beta^\Gamma = (\beta^\Gamma)'$ for $\Gamma$ given by (6.1), the most general parameterization of $\beta$ is

$$
\beta = \frac{1}{3} \begin{bmatrix} 2\delta_1 - \delta_2, & 2\delta_2 - \delta_1 \\ 2\delta_2 - \delta_3, & 2\delta_3 - \delta_2 \end{bmatrix},
$$

(6.3)

where $\delta_1$, $\delta_2$, and $\delta_3$ are free parameters. Note that the balanced case $\delta_1 = \delta_3$ is not label-invariant in the sense that $\log \frac{y_1^{(1)}y_2^{(2)}}{y_2^{(1)}y_1^{(2)}} = x^{(3)}$, say, leads to

$$
E(x^{(3)}_r| \cdot) = \text{const} + \sum_{s \in N} x^{(3)}_{r+s},
$$

which does not contain other variables like $E(x^{(1)}_r| \cdot)$. However, note that for the factorized model requiring $\delta_1 = \delta_3 = \delta/2$, or

$$
\beta = \frac{1}{2}\delta I,
$$

(6.4)

the MRF is label-invariant. It is important to assess (6.4) against (6.3) so that a comparatively simpler model (6.4) can be used if there is no gain with (6.3).

We illustrate the method through Switzer's landsat data [24]. The ground truth is a $16 \times 25$ lattice of three rock types with hard classification, i.e., $Y_i = 1$ or 0. We assign the value $Y_i = .001$ for $Y_i = 0$ and $Y_i = .998$ for $Y_i = 1$ to make the data slightly fuzzier. We estimate the parameters using the pseudo-likelihood method of Besag [3]. To remove the edge effects we only consider the $14 \times 23$ internal grid. Under (6.3), we have

$$
\hat{\beta} = \begin{pmatrix} .2641 & .0027 \\ .0034 & .2628 \end{pmatrix}, \quad \hat{\tau}^2 = 5.5288.
$$

Note that $F_x(\theta)$ is p.d. only if $\sum_{s \in N} |\beta_{js}| < 1$ for all $i$, $\beta_j = (\beta_{ij})$. This does not hold here as $4\beta_{11} > 1$, but in the context of this application only local behaviour is important, rather than global behaviour. Also $\hat{\tau}^2$ is large as the data are nearly hard. Letting $L$ be the pseudo-likelihood, it is found that

$$
-2 \log \hat{L} = -2695.00.
$$

Under (6.4), we have

$$
\hat{\beta} = .2624 I, \quad \hat{\tau}^2 = 5.5320, \quad -2 \log \hat{L} = -2695.75.
$$

These values indicate that there is no loss in using the factorized model. The formal properties of these tests will be discussed elsewhere but this example indicates that the psuedo-likelihood estimators can over-shoot at the boundary.
ACKNOWLEDGMENTS

The author wishes to thank the referee, John Kent, and Alan Watkins for their helpful comments.

REFERENCES


