12

Statistical Inference on Markov Random Fields: Parameter Estimation, Asymptotic Evaluation and Contextual Classification of NMR Multispectral Images

Alexandre L. M. Levada[1], Nelson D. A. Mascarenhas[2] and Alberto Tannús[1]
[1]*University of São Paulo, Physics Institute of São Carlos*
[2]*Federal University of São Carlos, Computer Department*
Brazil

1. Introduction

Undoubtedly, Markov Random Fields (*MRF*) define a powerful mathematical tool for contextual modelling of spatial data. With advances in probability and statistics (Hammersley & Clifford, 1971), as the development of Markov Chain Monte Carlo (*MCMC*) simulation techniques (Metropolis et al., 1953; Geman & Geman, 1984; Swendsen & Wang, 1987; Wolff, 1989) and relaxation algorithms for combinatorial optimization (Besag, 1986; Marroquin et al., 1987; Yu & Berthod, 1995), *MRF's* became a central topic in fields including image processing, computer vision and pattern recognition. In this chapter, we are concerned with the multispectral image contextual classification problem. A Bayesian approach is used to combine two MRF models: a Gaussian Markov Random Field (*GMRF*) for the observations (likelihood) and a *Potts* model for the *a priori* knowledge. Hence, the problem is stated according to a Maximum *a Posteriori* (*MAP*) framework.

One of the main difficulties in contextual classification using a *MAP-MRF* approach relies on the *MRF* parameter estimation stage. Traditional methods, as Maximum Likelihood (*ML*), cannot be applied due to the existence of a partition function in the joint Gibbs distribution, which is computationally intractable. A solution proposed by Besag to surmount this problem is to use the local conditional density functions (*LCDF*) to perform maximum pseudo-likelihood (*MPL*) estimation (Besag, 1974). The main motivation for employing this approach is that *MPL* estimation is a computationally feasible method. Besides, from a statistical perspective, *MPL* estimators have a series of desirable and interesting properties, such as consistency and asymptotic normality (Jensen & Künsh, 1994). However, a serious limitation of contextual classification has been the use of extremely restricted neighbourhood systems. Actually, traditional methods often consider only first-order neighbourhood systems.

The main motivation for this chapter is to discuss the incorporation of higher-order neighbourhood systems in *MRF* models, since among several drawbacks existing in classification problems, the lack of an accurate contextual modelling is definitely a major

one, especially when we are dealing with real image data, often degraded by noise. And, with the introduction of higher-order systems, novel, robust and suitable parameter estimation methods are also required. This chapter presents a novel framework for Bayesian image contextual classification through the definition of statistical inference and parameter estimation techniques in higher-order systems. Pseudo-likelihood equations for both *Potts* and *GMRF* models are presented and analysed using asymptotic evaluations and *MCMC* simulation algorithms. Two combinatorial optimization algorithms for *MAP-MRF* contextual classification are described and compared: *Iterated Conditional Modes* (*ICM*) and *Maximizer of the Posterior Marginals* (*MPM*). Experiments on real Nuclear Magnetic Resonance (*NMR*) images illustrate the proposed methodology.
The remaining of this chapter is organized as follows: Section 2 introduces the reader to the combined *GMRF* + *Potts MRF* model for contextual classification and discusses how to perform parameter estimation using the maximum pseudo-likelihood approach. In Section 3 we present asymptotic evaluations to assess the accuracy of *MPL* estimation by means of approximations for the *MPL* estimators' variances. Section 4 introduces the combinatorial optimization algorithms for contextual classification. Section 5 discusses metrics for quantitative performance evaluation, more precisely Cohen's Kappa coefficient. Section 6 shows results of the proposed classification methodology on real multispectral magnetic resonance images. Finally, Section 7 presents the conclusions and final remarks.

2. *MAP-MRF* Contextual Classification

Let $\mathbf{x}_w^{(p)}$ be the label field representing the classification map at the p-th iteration, $\boldsymbol{y}$ the observed multispectral image, $\boldsymbol{\theta}$ the vector of *GMRF* hyperparameters, $\boldsymbol{\Phi}$ the vector of *GMRF* spectral parameters for each class ($\boldsymbol{\mu}_m$, $\boldsymbol{\Sigma}_m$) and β the *Potts MRF* model hyperparameter. Considering a multispectral *GMRF* model for the observations and a *Potts* model for the *a priori* knowledge, according to the Bayes' rule, the current label of pixel (i, j) can be iteratively updated by choosing the label that maximizes the functional (Yamazaki & Gingras, 1996):

$$Q\left(x_{ij} = m \mid \boldsymbol{x}_w^{(p)}, \boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{\Phi}, \beta\right) = -\frac{1}{2}\ln\left|\hat{\boldsymbol{\Sigma}}_m\right| - \frac{1}{2}\left\{\boldsymbol{y}_{ij} - \hat{\boldsymbol{\mu}}_m\left[\hat{\boldsymbol{\theta}}^T \vec{y}_{\eta_{ij}} - 2\left(\sum_{ct}\hat{\theta}^{ct}\right)\hat{\boldsymbol{\mu}}_m\right]\right\}^T \hat{\boldsymbol{\Sigma}}_m^{-1} \times \left\{\boldsymbol{y}_{ij} - \hat{\boldsymbol{\mu}}_m\left[\hat{\boldsymbol{\theta}}^T \boldsymbol{y}_{\eta_{ij}} - 2\left(\sum_{ct}\hat{\theta}^{ct}\right)\hat{\boldsymbol{\mu}}_m\right]\right\} + \hat{\beta}U_{ij}(m) \tag{1}$$

where $\hat{\theta}^{ct}$ is a diagonal matrix whose elements are the horizontal, vertical and diagonals hyperparameters (4 x 4), $ct = 1, \ldots, K$, where K is the number of bands, $\hat{\boldsymbol{\theta}}^T$ is a matrix build by stacking the $\hat{\theta}^{ct}$ diagonal matrices from each image band (4 x 4K), that is, $\hat{\boldsymbol{\theta}}^T = \left[\hat{\theta}^{ct1}, \hat{\theta}^{ct2}, \ldots, \hat{\theta}^{ctK}\right]$ and $\boldsymbol{y}_{\eta_{ij}}$ is a vector whose elements are defined as the sum of the

two neighbouring elements on each direction (horizontal, vertical and diagonals) for all the image bands ($4K$ x 1).

2.1 *MPL* Estimation for *GMRF* model parameters

As the proposed model for contextual classification of multispectral images assumes independency between different image bands, it is quite reasonable to perform *MPL* estimation in each image band separately. Assuming this hypothesis and considering a second-order neighbourhood system, the pseudo-likelihood equation for the *GMRF* hyperparameters becomes (Won & Gray, 2004):

$$\log PL\left(\theta,\mu,\sigma^2\right)=\sum_{(i,j)\in W}\left\{-\frac{1}{2}\log\left(2\pi\sigma^2\right)-\frac{1}{2\sigma^2}\left[y_{ij}-\theta^T\boldsymbol{\Psi}_{ij}-\mu\left(1-2\theta\boldsymbol{I}\right)\right]^2\right\} \tag{2}$$

where W represents an image band, $\theta^T=\left[\theta_1,\theta_2,\theta_3,\theta_4\right]$ is the hyperparameters vector and $\boldsymbol{\Psi}_{ij}=\left[\left(y_{i+1j}+y_{i-1j}\right),\left(y_{ij+1}+y_{ij-1}\right),\left(y_{i+1j-1}+y_{i-1j+1}\right),\left(y_{i-1j-1}+y_{i+1j+1}\right)\right]$. Fortunately, the *MPL* estimator of θ admits a closed solution, given by (Won & Gray, 2004):

$$\hat{\theta}=\left\{\left[\sum_{(i,j)\in W}\left(y_{ij}-\hat{\mu}\right)\tilde{\boldsymbol{\Psi}}_{ij}^T\right]\left[\sum_{(i,j)\in W}\tilde{\boldsymbol{\Psi}}_{ij}\tilde{\boldsymbol{\Psi}}_{ij}^T\right]^{-1}\right\} \tag{3}$$

where $\hat{\mu}$ is the sample mean of the image pixels, $\tilde{\boldsymbol{\Psi}}_{ij}=\boldsymbol{\Psi}_{ij}-\frac{1}{N}\sum_{(k,l)\in\Omega}\boldsymbol{\Psi}_{ij}$ and N is the number of image pixels.

2.2 *MPL* Estimation for *Potts MRF* model parameter

One of the most widely used prior models in Bayesian image modelling is the Potts *MRF* pair-wise interaction (*PWI*) model. Two fundamental characteristics of the Potts model considered here are: it is both isotropic and stationary. According to (Hammersley & Clifford, 1971), the Potts *MRF* model can be equivalently defined in two manners: by a joint Gibbs distribution (global model) or by a set of local conditional density functions (*LCDF's*). For a general *s-th* order neighbourhood system η^s, we define the former by the following expression:

$$p\left(x_{ij}=m|\eta^s,\beta\right)=\frac{\exp\left\{\beta U_{ij}(m)\right\}}{\sum_{l=1}^{M}\exp\left\{\beta U_{ij}(l)\right\}} \tag{4}$$

where $U_{ij}(l)$ is the number of neighbours of the *i-th* element having label equal to l, β is the spatial dependency parameter (known as inverse temperature), $m, l \in G = \{1,2,\ldots,M\}$, with M denoting the number of classes. So, the pseudo-likelihood equation for the Potts model is given by:

$$PL(\beta) = \prod_{(i,j)\in W} \frac{\exp\{\beta U_{ij}(m)\}}{\sum_{l=1}^{M} \exp\{\beta U_{ij}(l)\}} \tag{5}$$

Taking the logarithms, differentiating on the parameter and setting the result to zero, leads to the following expression, which is the basis for the derivation of the proposed equations:

$$\frac{\partial}{\partial\beta} \log PL(\beta) = \sum_{(i,j)\in W} U_{ij}(m) - \sum_{(i,j)\in W} \left[\frac{\sum_{l=1}^{M} U_{ij}(l) \exp\{\beta U_{ij}(l)\}}{\sum_{l=1}^{M} \exp\{\beta U_{ij}(l)\}} \right] = 0 \tag{6}$$

where m denotes the observed value for the *ij-th* element of the field.
Looking at equation (6) it is possible to see that its first term is independent of the parameter. Thus, it is possible to expand the second term of (6) in all possible spatial configuration patterns that provide different contributions to the pseudo-likelihood equation regarding a pre-defined neighbourhood system. For example, in first order systems, the enumeration of these configuration patterns is straightforward, since there are only five cases, from zero agreement (situation of four different labels) to total agreement (situation of four identical labels), as shown in Figure 1.

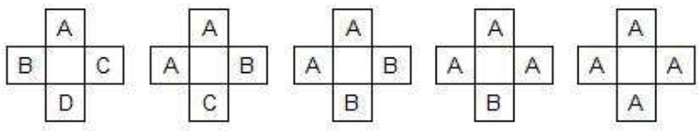


Fig. 1. Contextual configuration patterns for Potts *MRF* model in first-order neighbourhood systems

These configuration patterns can represented by vectors, as presented in relations (7), indicating the number of occurrences of each element around the central element. In the Potts model location information is irrelevant since it is an isotropic model:

$$v_0 = [1,1,1,1]; \quad v_1 = [2,1,1,0]; \quad v_2 = [2,2,0,0]; \quad v_3 = [3,1,0,0]; \quad v_4 = [4,0,0,0]; \tag{7}$$

Let N be the number of elements in the neighbourhood system η^s. For each $L = 1,\ldots,N$, let:

$$A_N(L) = \left\{ \left(a_1, \ldots, a_L\right) / \, a_i \in \left\{1, 2, \ldots, N\right\}, a_1 \leq \cdots \leq a_L, \sum_{i=1}^{L} a_i = N \right\} \tag{8}$$

and $n_N(L)$ the number of elements of the set $A_N(L)$. Then, the number of configuration patterns for the neighbourhood system η^s is $\lambda = n_N(1) + n_N(2) + \cdots + n_N(L)$.
Thus, the problem of finding the possible contextual configuration patterns can be solved automatically. The solution vectors can be found by exhaustive searching, by isolating one variable and searching on the subspace spanned by the remainder variables. Table 1 presents the number of configuration patterns for several neighbourhood systems.

Neighbourhood System	*Number of configuration patterns (λ)*
First order	5
Second order	22
Third order	77
Fourth order	637
Fifth order	1575

Table 1. Number of possible contextual configuration patterns for five neighbourhood systems

Now, given the complete set of contextual configuration patterns for a neighbourhood system, it is possible to expand the second term of equation (6). We can regard its numerator as an inner product of two vectors $\boldsymbol{U_{ij}}$ and $\boldsymbol{\omega_{ij}}$, where $\boldsymbol{U_{ij}}$ represents the contextual configuration vector for the current pixel (i.e., $\boldsymbol{U_{ij}} = [5,2,1,0,0,0,0,0]$ in case of a second-order system) and $\boldsymbol{\omega_{ij}}$ is a vector such that $\omega_{ij}[n] = \exp\left\{\beta U_{ij}[n]\right\}$. Similarly, the denominator is the inner product of $\boldsymbol{\omega_{ij}}$ with the identity column vector $\boldsymbol{r} = [1,1,\ldots,1]$. So, the second term of equation (6) can be expanded as a summation of λ terms, each one associated with a possible configuration pattern. However, as it involves a summation for all elements of the MRF, we define constants $K_i, i = 1, \ldots, \lambda$, representing the number of number of occurrences of each possible configuration patters along the entire image.
Basically, the idea is that the set of all K_i coefficients defines a contextual histogram, that is, instead of indicating the distribution of individual pixel gray levels, this set shows the distribution of spatial patterns defined in terms of the adopted neighbourhood system. For instance, in image analysis applications, smooth images, with many homogeneous regions, tend to present more concentration of configuration patterns with similar labels. On the other hand, heterogeneous regions tend to present concentration of configuration patterns with higher variation in the labels. Figures 2, 3 and 4 show an example with the Lena image.
It is worthwhile noting that in a physical interpretation we are using the proposed equations to estimate a quantity called *inverse temperature* in a system of particles arranged on a 2-D lattice using only pair-wise interactions. The proposed pseudo-likelihood equation for second-order neighbourhood systems is given by equation (9) (Levada et al., 2008a).

Similarly, for third-order systems, the equation is obtained by expanding (6) on the *77* configuration patterns derived by solving (8) for *N=12*. The proposed pseudo-likelihood equation for third-order systems is given by equation (10) (Levada et al., 2008b).

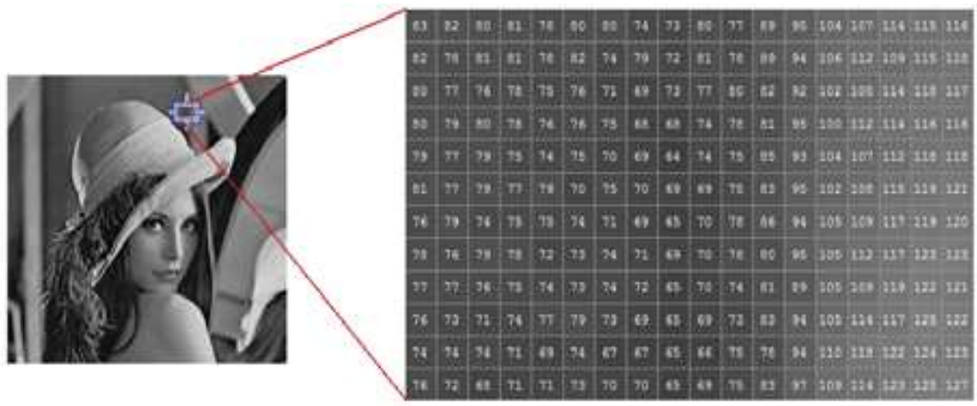


Fig. 2. Smooth regions present more homogeneous contextual configuration patterns

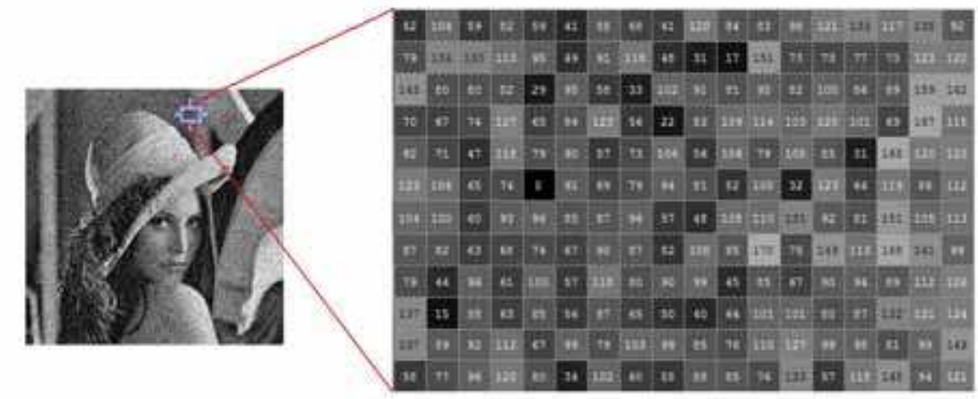


Fig. 3. Noisy regions present more heterogeneous contextual configuration patterns

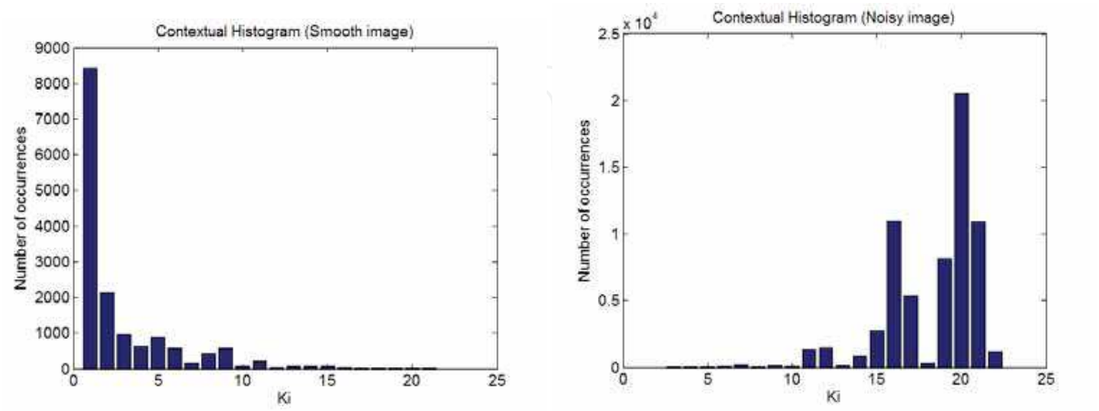


Fig. 4. Comparison between the distribution of contextual configuration patterns for both smooth and noisy Lena images (k_0 stands for total agreement and k_{22} for zero agreement).

$$\frac{\partial}{\partial\beta}\log PL(\beta)=\sum_{(i,j)\in W}U_{ij}(m)-\frac{8e^{8\hat{\beta}}}{e^{8\hat{\beta}}+M-1}K_1-\frac{7e^{7\hat{\beta}}+e^{\hat{\beta}}}{e^{7\hat{\beta}}+e^{\hat{\beta}}+M-2}K_2-\frac{6e^{6\hat{\beta}}+2e^{2\hat{\beta}}}{e^{6\hat{\beta}}+e^{2\hat{\beta}}+M-2}K_3$$
$$-\frac{6e^{6\hat{\beta}}+2e^{\hat{\beta}}}{e^{6\hat{\beta}}+2e^{\hat{\beta}}+M-3}K_4-\frac{5e^{5\hat{\beta}}+3e^{3\hat{\beta}}}{e^{5\hat{\beta}}+e^{3\hat{\beta}}+M-2}K_5-\frac{5e^{5\hat{\beta}}+2e^{2\hat{\beta}}+e^{\hat{\beta}}}{e^{5\hat{\beta}}+e^{2\hat{\beta}}+e^{\hat{\beta}}+M-3}K_6-\frac{5e^{5\hat{\beta}}+3e^{\hat{\beta}}}{e^{5\hat{\beta}}+e^{3\hat{\beta}}+M-4}K_7$$
$$-\frac{8e^{4\hat{\beta}}}{2e^{4\hat{\beta}}+M-2}K_8-\frac{4e^{4\hat{\beta}}+3e^{3\hat{\beta}}+e^{\hat{\beta}}}{e^{4\hat{\beta}}+e^{3\hat{\beta}}+e^{\hat{\beta}}+M-3}K_9-\frac{4e^{4\hat{\beta}}+4e^{2\hat{\beta}}}{e^{4\hat{\beta}}+2e^{2\hat{\beta}}+M-3}K_{10}-\frac{4e^{4\hat{\beta}}+2e^{2\hat{\beta}}+2e^{\hat{\beta}}}{e^{4\hat{\beta}}+e^{2\hat{\beta}}+2e^{\hat{\beta}}+M-4}K_{11}$$
$$-\frac{4e^{4\hat{\beta}}+4e^{\hat{\beta}}}{e^{4\hat{\beta}}+4e^{\hat{\beta}}+M-5}K_{12}-\frac{6e^{3\hat{\beta}}+2e^{2\hat{\beta}}}{2e^{3\hat{\beta}}+e^{2\hat{\beta}}+M-3}K_{13}-\frac{6e^{3\hat{\beta}}+2e^{\hat{\beta}}}{2e^{3\hat{\beta}}+2e^{\hat{\beta}}+M-4}K_{14}-\frac{3e^{3\hat{\beta}}+4e^{2\hat{\beta}}+e^{\hat{\beta}}}{e^{3\hat{\beta}}+2e^{2\hat{\beta}}+e^{\hat{\beta}}+M-4}K_{15}$$
$$-\frac{3e^{3\hat{\beta}}+2e^{2\hat{\beta}}+3e^{\hat{\beta}}}{e^{3\hat{\beta}}+e^{2\hat{\beta}}+3e^{\hat{\beta}}+M-5}K_{16}-\frac{3e^{3\hat{\beta}}+5e^{\hat{\beta}}}{e^{5\hat{\beta}}+5e^{\hat{\beta}}+M-6}K_{17}-\frac{8e^{2\hat{\beta}}}{4e^{2\hat{\beta}}+M-4}K_{18}-\frac{6e^{2\hat{\beta}}+2e^{\hat{\beta}}}{3e^{2\hat{\beta}}+2e^{\hat{\beta}}+M-5}K_{19}$$
$$-\frac{4e^{2\hat{\beta}}+4e^{\hat{\beta}}}{2e^{2\hat{\beta}}+4e^{\hat{\beta}}+M-6}K_{20}-\frac{2e^{2\hat{\beta}}+6e^{\hat{\beta}}}{e^{2\hat{\beta}}+6e^{\hat{\beta}}+M-7}K_{21}-\frac{8e^{\hat{\beta}}}{8e^{\hat{\beta}}+M-8}K_{22}=0 \tag{9}$$

The derived transcendental equations do not have closed solution, so in order to solve them a root-finding algorithm is required. In all experiments along this chapter, the *MPL* estimator is obtained by Brent's method (Brent, 1973), a numerical method that does not require the computation (not even the existence) of derivatives or analytical gradients. In this case, the computation of derivatives of the objective function would be prohibitive, given the large extension of the expressions. Basically, the advantages of this method can be summarized by: it uses a combination of bisection, secant and inverse quadratic interpolation methods, leading to a very robust approach and also it has super-linear convergence rate.

3. Statistical Inference and Asymptotic Evaluation on Markov Random Fields

Unbiasedness is not granted by either maximum likelihood (*ML*) or maximum pseudo-likelihood (*MPL*) estimation. Actually, according to statistical inference theory, there is no method that guarantees the existence of unbiased estimators for a fixed *N*-size sample. Often, in the exponential family, *ML* estimators coincide with *UMVU* (*Uniform Minimum Variance Unbiased*) estimators because they are functions of complete sufficient statistics (if a *ML* estimator is unique then it is a function of sufficient statistics). Besides, there are several characteristics that make *ML* estimation a reference method (Lehman, 1983; Bickel, 1991; Casella, 2002). Making the sample size grow infinitely ($N\to\infty$), ML estimator becomes asymptotically unbiased and efficient. Unfortunately, there is no result showing that the same occurs in *MPL* estimation. In this section, we show how to approximate the asymptotic variances of Potts and *GMRF* model parameters in terms of expressions for the observed Fisher information using both first and second derivatives.

$$\frac{\partial}{\partial\beta}\log PL\left(\beta\right)=\sum_{(i,j)\in W}U_{ij}(m)-\frac{12e^{12\hat{\beta}}}{e^{12\hat{\beta}}+M-1}K_1-\frac{11e^{11\hat{\beta}}+e^{\hat{\beta}}}{e^{11\hat{\beta}}+e^{\hat{\beta}}+M-2}K_2-\frac{10e^{10\hat{\beta}}+2e^{2\hat{\beta}}}{e^{10\hat{\beta}}+e^{2\hat{\beta}}+M-2}K_3-\frac{9e^{9\hat{\beta}}+3e^{3\hat{\beta}}}{e^{9\hat{\beta}}+e^{3\hat{\beta}}+M-2}K_4$$
$$-\frac{8e^{8\hat{\beta}}+4e^{4\hat{\beta}}}{e^{8\hat{\beta}}+e^{4\hat{\beta}}+M-2}K_5-\frac{7e^{7\hat{\beta}}+5e^{5\hat{\beta}}}{e^{7\hat{\beta}}+e^{5\hat{\beta}}+M-2}K_6-\frac{12e^{6\hat{\beta}}}{2e^{6\hat{\beta}}+M-2}K_7-\frac{12e^{4\hat{\beta}}}{3e^{4\hat{\beta}}+M-3}K_8-\frac{5e^{5\hat{\beta}}+4e^{4\hat{\beta}}+3e^{3\hat{\beta}}}{e^{5\hat{\beta}}+e^{4\hat{\beta}}+e^{3\hat{\beta}}+M-3}K_9$$
$$-\frac{10e^{5\hat{\beta}}+2e^{2e^{\hat{\beta}}}}{2e^{5\hat{\beta}}+e^{2\hat{\beta}}+M-3}K_{10}-\frac{6e^{6\hat{\beta}}+6e^{3\hat{\beta}}}{e^{6\hat{\beta}}+2e^{3\hat{\beta}}+M-3}K_{11}-\frac{6e^{6\hat{\beta}}+4e^{4\hat{\beta}}+2e^{2\hat{\beta}}}{e^{6\hat{\beta}}+e^{4\hat{\beta}}+e^{2\hat{\beta}}+M-3}K_{12}-\frac{6e^{6\hat{\beta}}+5e^{5\hat{\beta}}+e^{\hat{\beta}}}{e^{6\hat{\beta}}+e^{5\hat{\beta}}+e^{\hat{\beta}}+M-3}K_{13}-$$
$$\frac{7e^{7\hat{\beta}}+3e^{3\hat{\beta}}+2e^{2\hat{\beta}}}{e^{7\hat{\beta}}+e^{3\hat{\beta}}+e^{2\hat{\beta}}+M-3}K_{14}-\frac{7e^{7\hat{\beta}}+4e^{4\hat{\beta}}+e^{\hat{\beta}}}{e^{7\hat{\beta}}+e^{4\hat{\beta}}+e^{\hat{\beta}}+M-3}K_{15}-\frac{8e^{8\hat{\beta}}+4e^{2\hat{\beta}}}{e^{8\hat{\beta}}+2e^{2\hat{\beta}}+M-3}K_{16}-\frac{8e^{8\hat{\beta}}+3e^{3\hat{\beta}}+e^{\hat{\beta}}}{e^{8\hat{\beta}}+e^{3\hat{\beta}}+e^{\hat{\beta}}+M-3}K_{17}-$$
$$\frac{9e^{9\hat{\beta}}+2e^{2\hat{\beta}}+e^{\hat{\beta}}}{e^{9\hat{\beta}}+e^{2\hat{\beta}}+e^{\hat{\beta}}+M-3}K_{18}-\frac{10e^{10\hat{\beta}}+2e^{\hat{\beta}}}{e^{10\hat{\beta}}+2e^{\hat{\beta}}+M-3}K_{19}-\frac{12e^{3\hat{\beta}}}{4e^{3\hat{\beta}}+M-4}K_{20}-\frac{4e^{4\hat{\beta}}+6e^{3\hat{\beta}}+2e^{2\hat{\beta}}}{e^{4\hat{\beta}}+2e^{3\hat{\beta}}+e^{2\hat{\beta}}+M-4}K_{21}-\frac{8e^{4\hat{\beta}}+4e^{2\hat{\beta}}}{2e^{4\hat{\beta}}+2e^{2\hat{\beta}}+M-4}K_{22}$$
$$-\frac{8e^{4\hat{\beta}}+3e^{3\hat{\beta}}+e^{\hat{\beta}}}{2e^{4\hat{\beta}}+e^{3\hat{\beta}}+e^{\hat{\beta}}+M-4}K_{23}-\frac{5e^{5\hat{\beta}}+3e^{3\hat{\beta}}+4e^{2\hat{\beta}}}{e^{5\hat{\beta}}+e^{3\hat{\beta}}+2e^{2\hat{\beta}}+M-4}K_{24}-\frac{5e^{5\hat{\beta}}+6e^{3\hat{\beta}}+e^{\hat{\beta}}}{e^{5\hat{\beta}}+2e^{3\hat{\beta}}+e^{\hat{\beta}}+M-4}K_{25}-\frac{5e^{5\hat{\beta}}+4e^{4\hat{\beta}}+2e^{2\hat{\beta}}+e^{\hat{\beta}}}{e^{5\hat{\beta}}+e^{4\hat{\beta}}+e^{2\hat{\beta}}+e^{\hat{\beta}}+M-4}K_{26}$$
$$-\frac{10e^{5\hat{\beta}}+2e^{\hat{\beta}}}{2e^{5\hat{\beta}}+2e^{\hat{\beta}}+M-4}K_{27}-\frac{6e^{6\hat{\beta}}+6e^{2\hat{\beta}}}{e^{6\hat{\beta}}+3e^{2\hat{\beta}}+M-4}K_{28}-\frac{6e^{6\hat{\beta}}+3e^{3\hat{\beta}}+2e^{2\hat{\beta}}+e^{\hat{\beta}}}{e^{6\hat{\beta}}+e^{3\hat{\beta}}+e^{2\hat{\beta}}+e^{\hat{\beta}}+M-4}K_{29}-\frac{6e^{6\hat{\beta}}+4e^{4\hat{\beta}}+2e^{\hat{\beta}}}{e^{6\hat{\beta}}+e^{4\hat{\beta}}+2e^{\hat{\beta}}+M-4}K_{30}$$
$$-\frac{7e^{7\hat{\beta}}+4e^{2\hat{\beta}}+e^{\hat{\beta}}}{e^{7\hat{\beta}}+2e^{2\hat{\beta}}+e^{\hat{\beta}}+M-4}K_{31}-\frac{7e^{7\hat{\beta}}+3e^{3\hat{\beta}}+2e^{\hat{\beta}}}{e^{7\hat{\beta}}+e^{3\hat{\beta}}+2e^{\hat{\beta}}+M-4}K_{32}-\frac{8e^{8\hat{\beta}}+2e^{2\hat{\beta}}+2e^{\hat{\beta}}}{e^{8\hat{\beta}}+e^{2\hat{\beta}}+2e^{\hat{\beta}}+M-4}K_{33}-\frac{9e^{9\hat{\beta}}+3e^{\hat{\beta}}}{e^{9\hat{\beta}}+3e^{\hat{\beta}}+M-4}K_{34}$$
$$-\frac{6e^{3\hat{\beta}}+6e^{2\hat{\beta}}}{2e^{3\hat{\beta}}+3e^{2\hat{\beta}}+M-5}K_{35}-\frac{9e^{3\hat{\beta}}+2e^{2\hat{\beta}}+e^{\hat{\beta}}}{3e^{3\hat{\beta}}+e^{2\hat{\beta}}+e^{\hat{\beta}}+M-5}K_{36}-\frac{4e^{4\hat{\beta}}+8e^{2\hat{\beta}}}{e^{4\hat{\beta}}+4e^{2\hat{\beta}}+M-5}K_{37}-\frac{4e^{4\hat{\beta}}+3e^{3\hat{\beta}}+4e^{2\hat{\beta}}+e^{\hat{\beta}}}{e^{4\hat{\beta}}+e^{3\hat{\beta}}+2e^{2\hat{\beta}}+e^{\hat{\beta}}+M-5}K_{38}$$
$$-\frac{4e^{4\hat{\beta}}+6e^{3\hat{\beta}}+2e^{\hat{\beta}}}{e^{4\hat{\beta}}+2e^{3\hat{\beta}}+2e^{\hat{\beta}}+M-5}K_{39}-\frac{8e^{4\hat{\beta}}+2e^{2\hat{\beta}}+2e^{\hat{\beta}}}{2e^{4\hat{\beta}}+e^{2\hat{\beta}}+2e^{\hat{\beta}}+M-5}K_{40}-\frac{5e^{5\hat{\beta}}+6e^{2\hat{\beta}}+e^{\hat{\beta}}}{e^{5\hat{\beta}}+3e^{2\hat{\beta}}+e^{\hat{\beta}}+M-5}K_{41}-\frac{5e^{5\hat{\beta}}+3e^{3\hat{\beta}}+2e^{2\hat{\beta}}+2e^{\hat{\beta}}}{e^{5\hat{\beta}}+e^{3\hat{\beta}}+e^{2\hat{\beta}}+2e^{\hat{\beta}}+M-5}K_{42}$$
$$-\frac{5e^{5\hat{\beta}}+4e^{4\hat{\beta}}+3e^{\hat{\beta}}}{e^{5\hat{\beta}}+e^{4\hat{\beta}}+3e^{\hat{\beta}}+M-5}K_{43}-\frac{6e^{6\hat{\beta}}+4e^{4\hat{\beta}}+2e^{\hat{\beta}}}{e^{6\hat{\beta}}+2e^{2\hat{\beta}}+2e^{\hat{\beta}}+M-5}K_{44}-\frac{6e^{6\hat{\beta}}+3e^{3\hat{\beta}}+3e^{\hat{\beta}}}{e^{6\hat{\beta}}+e^{3\hat{\beta}}+3e^{\hat{\beta}}+M-5}K_{45}-\frac{7e^{7\hat{\beta}}+2e^{2\hat{\beta}}+3e^{\hat{\beta}}}{e^{7\hat{\beta}}+e^{2\hat{\beta}}+3e^{\hat{\beta}}+M-5}K_{46}$$
$$-\frac{8e^{8\hat{\beta}}+4e^{\hat{\beta}}}{e^{8\hat{\beta}}+4e^{\hat{\beta}}+M-5}K_{47}-\frac{12e^{2\hat{\beta}}}{6e^{2\hat{\beta}}+M-6}K_{48}-\frac{3e^{3\hat{\beta}}+8e^{2\hat{\beta}}+e^{\hat{\beta}}}{e^{3\hat{\beta}}+4e^{2\hat{\beta}}+e^{\hat{\beta}}+M-6}K_{49}-\frac{6e^{3\hat{\beta}}+4e^{2\hat{\beta}}+2e^{\hat{\beta}}}{2e^{3\hat{\beta}}+2e^{2\hat{\beta}}+2e^{\hat{\beta}}+M-6}K_{50}-\frac{9e^{3\hat{\beta}}+3e^{\hat{\beta}}}{3e^{3\hat{\beta}}+3e^{\hat{\beta}}+M-6}K_{51}$$
$$-\frac{4e^{4\hat{\beta}}+6e^{2\hat{\beta}}+2e^{\hat{\beta}}}{e^{4\hat{\beta}}+3e^{2\hat{\beta}}+2e^{\hat{\beta}}+M-6}K_{52}-\frac{4e^{4\hat{\beta}}+3e^{3\hat{\beta}}+2e^{2\hat{\beta}}-3e^{\hat{\beta}}}{e^{4\hat{\beta}}+e^{3\hat{\beta}}+e^{2\hat{\beta}}+3e^{\hat{\beta}}+M-6}K_{53}-\frac{8e^{4\hat{\beta}}+4e^{\hat{\beta}}}{2e^{4\hat{\beta}}+4e^{\hat{\beta}}+M-6}K_{54}-\frac{5e^{5\hat{\beta}}+4e^{2\hat{\beta}}+3e^{\hat{\beta}}}{e^{5\hat{\beta}}+2e^{2\hat{\beta}}+3e^{\hat{\beta}}+M-6}K_{55}$$
$$-\frac{5e^{5\hat{\beta}}+3e^{3\hat{\beta}}+4e^{\hat{\beta}}}{e^{5\hat{\beta}}+e^{3\hat{\beta}}+4e^{\hat{\beta}}+M-6}K_{56}-\frac{6e^{6\hat{\beta}}+2e^{2\hat{\beta}}+4e^{\hat{\beta}}}{e^{6\hat{\beta}}+e^{2\hat{\beta}}+4e^{\hat{\beta}}+M-6}K_{57}-\frac{7e^{7\hat{\beta}}+5e^{\hat{\beta}}}{e^{7\hat{\beta}}+5e^{\hat{\beta}}+M-6}K_{58}-\frac{10e^{2\hat{\beta}}+2e^{\hat{\beta}}}{5e^{2\hat{\beta}}+2e^{\hat{\beta}}+M-7}K_{59}$$
$$-\frac{3e^{3\hat{\beta}}+6e^{2\hat{\beta}}+3e^{\hat{\beta}}}{e^{3\hat{\beta}}+3e^{2\hat{\beta}}+3e^{\hat{\beta}}+M-7}K_{60}-\frac{6e^{3\hat{\beta}}+2e^{2\hat{\beta}}+4e^{\hat{\beta}}}{2e^{3\hat{\beta}}+e^{2\hat{\beta}}+4e^{\hat{\beta}}+M-7}K_{61}-\frac{4e^{4\hat{\beta}}+4e^{2\hat{\beta}}+4e^{\hat{\beta}}}{e^{4\hat{\beta}}+2e^{2\hat{\beta}}+4e^{\hat{\beta}}+M-7}K_{62}-\frac{4e^{4\hat{\beta}}+3e^{3\hat{\beta}}+5e^{\hat{\beta}}}{e^{4\hat{\beta}}+e^{3\hat{\beta}}+5e^{\hat{\beta}}+M-7}K_{63}\qquad(10)$$
$$\frac{5e^{5\hat{\beta}}+2e^{2\hat{\beta}}+5e^{\hat{\beta}}}{e^{5\hat{\beta}}+e^{2\hat{\beta}}+5e^{\hat{\beta}}+M-7}K_{64}-\frac{6e^{6\hat{\beta}}+6e^{\hat{\beta}}}{e^{6\hat{\beta}}+6e^{\hat{\beta}}+M-7}K_{65}-\frac{8e^{2\hat{\beta}}+4e^{\hat{\beta}}}{4e^{2\hat{\beta}}+4e^{\hat{\beta}}+M-8}K_{66}-\frac{3e^{3\hat{\beta}}+4e^{2\hat{\beta}}+5e^{\hat{\beta}}}{e^{3\hat{\beta}}+2e^{2\hat{\beta}}+5e^{\hat{\beta}}+M-8}K_{67}-\frac{6e^{3\hat{\beta}}+6e^{\hat{\beta}}}{2e^{3\hat{\beta}}+6e^{\hat{\beta}}+M-8}K_{68}$$
$$-\frac{4e^{4\hat{\beta}}+2e^{2\hat{\beta}}+6e^{\hat{\beta}}}{e^{4\hat{\beta}}+e^{2\hat{\beta}}+6e^{\hat{\beta}}+M-8}K_{69}-\frac{5e^{5\hat{\beta}}+7e^{\hat{\beta}}}{e^{5\hat{\beta}}+7e^{\hat{\beta}}+M-8}K_{70}-\frac{6e^{2\hat{\beta}}+6e^{\hat{\beta}}}{3e^{2\hat{\beta}}+6e^{\hat{\beta}}+M-9}K_{71}-\frac{3e^{3\hat{\beta}}+2e^{2\hat{\beta}}+7e^{\hat{\beta}}}{e^{3\hat{\beta}}+e^{2\hat{\beta}}+7e^{\hat{\beta}}+M-9}K_{72}-\frac{4e^{4\hat{\beta}}+8e^{\hat{\beta}}}{e^{4\hat{\beta}}+8e^{\hat{\beta}}+M-9}K_{73}$$
$$-\frac{4e^{2\hat{\beta}}+8e^{\hat{\beta}}}{2e^{2\hat{\beta}}+8e^{\hat{\beta}}+M-10}K_{74}-\frac{3e^{3\hat{\beta}}+9e^{\hat{\beta}}}{e^{3\hat{\beta}}+9e^{\hat{\beta}}+M-10}K_{75}-\frac{2e^{2\hat{\beta}}+10e^{\hat{\beta}}}{e^{2\hat{\beta}}+10e^{\hat{\beta}}+M-11}K_{76}-\frac{12e^{\hat{\beta}}}{12e^{\hat{\beta}}+M-12}K_{77}=0$$

3.1 Observed Fisher Information

Often, in practice, it is not possible to calculate the expected Fisher information $I\left(\theta\right)$. In such cases, we can adopt the observed Fisher information, $I_{obs}\left(\theta\right)$ instead. Furthermore, it has been shown (Efron & Hinkley, 1978) that the use of the observed information number is superior to the expected information number, as it appears in the Cramér–Rao lower bound. The observed Fisher information, in terms of the pseudo-likelihood function, is defined by

$$I_{obs}(\theta) = \left[\frac{\partial}{\partial\theta}\log PL(x;\theta)\right]^2 \tag{11}$$

and can be estimated by the following, justified by the Law of Large Numbers:

$$\hat{I}^1_{obs}(\theta) = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{\partial}{\partial\theta}\log p(x_i;\theta)\right]^2\Bigg|_{\theta=\hat{\theta}} \tag{12}$$

since $I(\theta) = E\left[\hat{I}^1_{obs}(\theta)\right]$, making $\hat{I}^1_{obs}(\theta) \approx I(\theta)$. Similarly, $I_{obs}(\theta)$ can be estimated using the second derivative of the pseudo-likelihood function:

$$\hat{I}^2_{obs}(\theta) = -\frac{1}{N}\sum_{i=1}^{N}\left[\frac{\partial^2}{\partial\theta^2}\log p(x_i;\theta)\right]\Bigg|_{\theta=\hat{\theta}} \tag{13}$$

3.2 On the Asymptotic Variances of *GMRF* model *MPL* estimators

The asymptotic covariance matrix for *MPL* estimators is given by (Liang and Yu, 2003):

$$C(\boldsymbol{\theta}) = H^{-1}(\boldsymbol{\theta})\,J(\boldsymbol{\theta})\,H^{-1}(\boldsymbol{\theta}) \tag{14}$$

where $J(\boldsymbol{\theta})$ and $H(\boldsymbol{\theta})$ are functions of the Jacobian (first order partial derivatives) and Hessian (second order partial derivatives) matrices respectively:

$$\begin{aligned} H(\boldsymbol{\theta}) &= E_\theta\left[\nabla^2 F(X;\boldsymbol{\theta})\right] \\ J(\boldsymbol{\theta}) &= Var_\theta\left[\nabla F(X;\boldsymbol{\theta})\right] \end{aligned} \tag{15}$$

with $F(X;\boldsymbol{\theta})$ denoting the logarithm of the pseudo-likelihood function.

Considering that the *GMRF* hyperparameters $\theta_1, \theta_2, \theta_3, \theta_4$ are uncorrelated, we have a diagonal covariance matrix, given by:

$$C(\boldsymbol{\theta}) = \begin{bmatrix} c_{11}(\boldsymbol{\theta}) & 0 & 0 & 0 \\ 0 & c_{22}(\boldsymbol{\theta}) & 0 & 0 \\ 0 & 0 & c_{33}(\boldsymbol{\theta}) & 0 \\ 0 & 0 & 0 & c_{44}(\boldsymbol{\theta}) \end{bmatrix} \tag{16}$$

with asymptotic variances given by:

$$c_{kk}(\boldsymbol{\theta}) = \frac{Var_{\theta}\left[\frac{\partial}{\partial \theta_k}\log PL(\boldsymbol{\theta})\right]}{E_{\theta}^{2}\left[\frac{\partial^2}{\partial \theta_k^2}\log PL(\boldsymbol{\theta})\right]} \tag{17}$$

for $k = 1,\ldots,4$. Rewriting the above equation using the definition of variance leads to:

$$c_{kk}(\boldsymbol{\theta}) = \frac{E_{\theta}\left[\left(\frac{\partial}{\partial \theta_k}\log PL(\boldsymbol{\theta})\right)^2\right]}{E_{\theta}^{2}\left[\frac{\partial^2}{\partial \theta_k^2}\log PL(\boldsymbol{\theta})\right]} \approx \frac{\hat{I}_{obs}^{1}(\boldsymbol{\theta})}{\left[\hat{I}_{obs}^{2}(\boldsymbol{\theta})\right]^2} \tag{18}$$

since the expected value of the log *PL* equation is zero:

$$\begin{aligned} E_{\theta}\left[\frac{\partial}{\partial \theta_k}\log PL(\boldsymbol{\theta})\right] &\approx \frac{1}{N}\sum_{i=1}^{N}\left[\frac{\partial}{\partial \theta_k}\log p\left(x_i|\eta^s,\boldsymbol{\theta}\right)\right]\Big|_{\theta=\hat{\theta}_{MPL}} \\ &= \frac{1}{N}\frac{\partial}{\partial \theta_k}\log\prod_{i=1}^{N} p\left(x_i|\eta^s,\boldsymbol{\theta}\right)\Big|_{\theta=\hat{\theta}_{MPL}} \\ &= \frac{1}{N}\frac{\partial}{\partial \theta_k}\log PL(\boldsymbol{\theta})\Big|_{\theta=\hat{\theta}_{MPL}} = 0 \end{aligned} \tag{19}$$

Thus, from the *LCDF* of the *GMRF* model and after some simple algebraic manipulations, we obtain the following expression for $\hat{I}_{obs}^{1}(\boldsymbol{\theta})$ (Levada et al., 2008c):

$$\hat{I}_{obs}^{1}(\boldsymbol{\theta}) = \frac{1}{N\sigma^4}\sum_{i=1}^{N}\left\{\left[\boldsymbol{y}_{ij} - \boldsymbol{\theta}^T\boldsymbol{\Psi}_{ij} - \mu\left(1 - 2\boldsymbol{\theta}^T\boldsymbol{I}\right)\right]\left[\Psi_{ij}^{k} - 2\mu\right]\right\}^2 \tag{20}$$

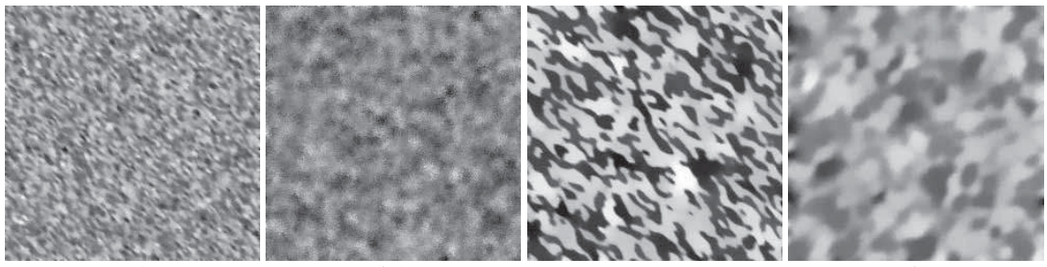
where Ψ_{ij}^{k} denotes the *k-th* element of $\boldsymbol{\Psi}_{ij}$, $k = 1,\ldots,4$. Similarly, for $\hat{I}_{obs}^{2}(\boldsymbol{\theta})$ we have:

$$\hat{I}_{obs}^{2}(\boldsymbol{\theta}) = -\frac{1}{N\sigma^2}\sum_{i=1}^{N}\left\{\Psi_{ij}^{k} - 2\mu\right\}^2 \tag{21}$$

The proposed approximation allows the calculation of the asymptotic variance of maximum pseudo-likelihood estimators of the *GMRF* model in computationally feasible way. From previous works on statistical inference it can be shown that *MPL* estimators are asymptotically normal distributed. Therefore, with the proposed method, it is possible to completely characterize the asymptotic behavior of the *MPL* estimators of the *GMRF* model, allowing interval estimation, hypothesis testing and quantitative analysis on the model

parameters in a variety of research areas, including image processing and pattern recognition (Levada et al., 2008d).

In order to demonstrate the application of the asymptotic variance estimation in stochastic image modeling, we present the results obtained in experiments using Markov Chain Monte Carlo simulation methods (Dubes & Jain, 1989; Winkler, 2006) by comparing the values of $\hat{\theta}_{MPL}$ and asymptotic variances regarding second order neighbourhood systems using synthetic images, representing several *GMRF* model outcomes. For the experiments below, we adopted the Metropolis algorithm (Metropolis et al., 1953), a single spin flip *MCMC* method, to simulate occurrences of *GRMF* model using different known parameter vectors. Simulated images are shown in Figure 5. The *MPL* estimators, obtained by (3) were compared with the real parameter vectors. In all cases, the parameters μ and σ^2 were defined as zero and five, respectively. The vector parameters used in the simulated images were $\theta = [0.25, 0.3, -0.1, 0.2]$ and $\theta = [0.2, 0.15, 0.07, 0.05]$.



a.) b.) c.) d.)

Fig. 5. Synthetic images generated by *MCMC* simulation for *GMRF* model using a second-order neighbourhood system.

A major difficulty in *GMRF* models is the selection of parameters θ for which the correlation matrix is positive definite, introducing one more problem in parameter estimation. With discrete *MRF's* almost any parameter in the parametric space lead to a mathematically valid model. However, only a portion of the parametric space generates valid *GMRF* models. The region of validity is known only for first-order systems, but not for higher-order *GMRF's* (Dubes & Jain, 1989). In fact, even parameters estimated by standard procedures may not be in the region of validity and simulations may not work properly.

Table 2 shows the *MPL* estimators, estimated asymptotic variances, 90% confidence intervals regarding the *GMRF* model parameters for the synthetic image indicated in Figure 1a. Similarly, Table 3 shows the same obtained results for the images shown in Figure 1b. The results on *MCMC* simulation images show that, in all cases, the true parameter value is contained in the obtained intervals, assessing the accuracy of the proposed methodology for asymptotic variance estimation.

K	θ_k	$\hat{\theta}_k$	$\hat{\sigma}_k$	90% CI
1	0.25	0.2217	0.0390	[0.1799 0.3077]
2	0.3	0.2758	0.0387	[0.2398 0.3667]
3	-0.1	-0.1145	0.0394	[-0.1711 -0.0479]
4	0.2	0.1743	0.0386	[0.1150 0.2416]

Table 2. *MPL* estimators, asymptotic variances and 90% confidence intervals for *GMRF* hyperparameters on simulated images (1a).

K	θ_k	$\hat{\theta}_k$	$\hat{\sigma}_k$	90% CI
1	0.2	0.1908	0.0506	[0.1079 0.2738]
2	0.15	0.1605	0.0524	[0.0746 0.2464]
3	0.07	0.0716	0.0482	[-0.0074 0.1506]
4	0.05	0.0523	0.0418	[-0.0146 0.1192]

Table 3. *MPL* estimators, asymptotic variances and 90% confidence intervals for *GMRF* hyperparameters on simulated images (1b).

3.3 On the Asymptotic Variance of *Potts MRF* model *MPL* estimator

Similarly to the *GMRF* model, we define an approximation for the asymptotic variance of Potts model *MPL* estimators through expressions for $\hat{I}^1_{obs}(\boldsymbol{\theta})$ and $\hat{I}^2_{obs}(\boldsymbol{\theta})$. From the *LCDF* of the Potts model (4) we have:

$$\hat{I}^1_{obs}(\beta) = \frac{1}{N}\sum_{i=1}^{N}\left\{\left[U_i(m) - \frac{\sum_{l=1}^{M} U_i(l)\exp\{\beta U_i(l)\}}{\sum_{l=1}^{M}\exp\{\beta U_i(l)\}}\right]^2\right\} \tag{22}$$

which, after some few algebraic manipulations, becomes (Levada et al, 2008a):

$$\hat{I}^1_{obs}(\beta) = \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{\sum_{l=1}^{M}\left[\sum_{k=1}^{M}\left(U_i(m)-U_i(l)\right)\left(U_i(m)-U_i(k)\right)\exp\left\{\beta\left(U_i(l)+U_i(k)\right)\right\}\right]}{\left[\sum_{l=1}^{M}\exp\{\beta U_i(l)\}\right]^2}\right\} \tag{23}$$

Calculating the observed Fisher information using the second derivative of the pseudo-likelihood function leads to the following:

$$\hat{I}^2_{obs}(\beta) = \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{\left[\sum_{l=1}^{M} U_i(l)^2\exp\{\beta U_i(l)\}\right]\left[\sum_{l=1}^{M}\exp\{\beta U_i(l)\}\right] - \left[\sum_{l=1}^{M} U_i(l)\exp\{\beta U_i(l)\}\right]^2}{\left[\sum_{l=1}^{M}\exp\{\beta U_i(l)\}\right]^2}\right\} \tag{24}$$

Simplifying equation (24), we have the final expression for $\hat{I}^2_{obs}(\boldsymbol{\theta})$ (Levada et al, 2008b):

$$\hat{I}^2_{obs}(\beta) = \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{\sum_{l=1}^{M-1}\left[\sum_{k=l}^{M}\left(U_i(l)-U_i(k)\right)^2\exp\left\{\beta\left(U_i(l)+U_i(k)\right)\right\}\right]}{\left[\sum_{l=1}^{M}\exp\left\{\beta U_i(l)\right\}\right]^2}\right\} \tag{25}$$

This approximation allows the calculation of the asymptotic variance of the maximum pseudo-likelihood estimator of the Potts *MRF* model. In order to demonstrate the application of the asymptotic variance in testing and evaluating the proposed pseudo-likelihood equation, some results obtained in experiments with Markov Chain Monte Carlo simulation methods are presented. The results show the values of $\hat{\beta}_{MPL}$, asymptotic variances, test statistics and *p-values* for several synthetic images generated using *MCMC* algorithms on second and third order neighbourhood systems. The objective is to validate the following hypothesis:

> *H*: the proposed pseudo-likelihood equations provide results that are statistically equivalent to the real parameter values, that is:

$$H: \beta = \hat{\beta}_{MPL} \tag{26}$$

Using the consistency property of *MPL* estimators and adopting the derived approximation for the asymptotic variance it is possible to completely characterize the asymptotic distribution of the Potts model parameter estimator and define the following test statistic:

$$Z = \frac{\beta-\hat{\beta}_{MPL}}{\hat{\sigma}^2\left(\hat{\beta}_{MPL}\right)} \approx N(0,1) \tag{27}$$

creating the decision rule: Reject *H* if $|Z| > c$. Considering a test size α (in all experiments along this chapter we set $\alpha = 0.1$), that is, the maximum probability of incorrectly rejecting *H* is α, we have $c = 1.64$. However, in order to quantify the evidence against or in favor of the hypothesis a complete analysis in terms the test size, test statistic and p-values, calculated by $P\left(|Z|>z_{obs}\right)$, is required. In case of a small p-value, we should doubt of the hypothesis being tested. In other words, to reject *H* we should have a test size α significantly higher than the p-value. This approach provides a statistically meaningful way to report the results of a hypothesis testing procedure.

For the experiments, to illustrate the example of statistical analysis in *MRF*, we adopted both single spin-flip *MCMC* methods, *Gibbs Sampler* and *Metropolis*, and a cluster-flipping *MCMC* method, the *Swendsen-Wang* algorithm, to generate several Potts model outcomes using different known parameter values. Figures 6, 7, 8 and 9 show the simulated images.

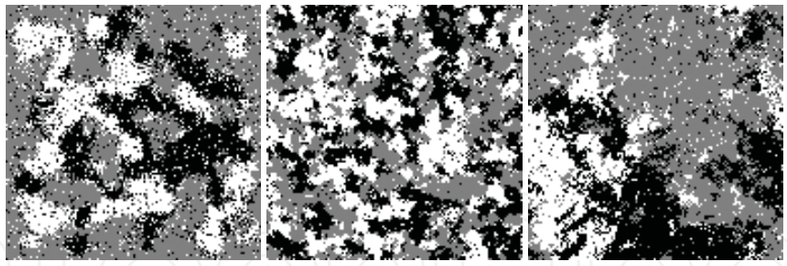


Fig. 6. Synthetic images generated by *MCMC* simulation algorithms using second-order neighbourhood systems for *M=3*: *Gibbs Sampler* ($\beta = 0.45$), *Metropolis* ($\beta = 0.5$) and *Swendsen-Wang* ($\beta = 0.4$), respectively.

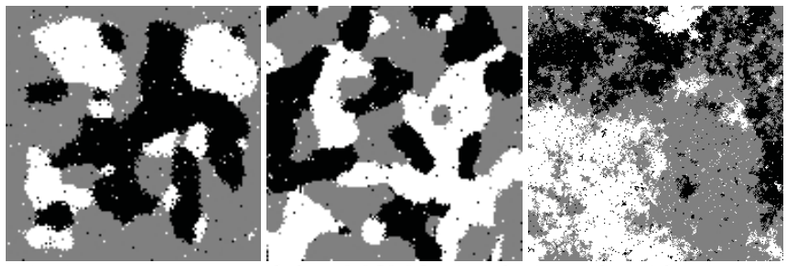


Fig. 7. Synthetic images generated by *MCMC* simulation algorithms using third-order neighbourhood systems for *M=3*: *Gibbs Sampler* ($\beta = 0.45$), *Metropolis* ($\beta = 0.5$) and *Swendsen-Wang* ($\beta = 0.4$), respectively.

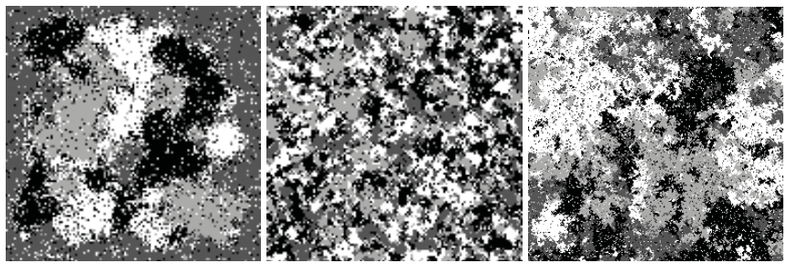


Fig. 8. Synthetic images generated by *MCMC* simulation algorithms using second-order neighbourhood systems for *M=4*: *Gibbs Sampler* ($\beta = 0.45$), *Metropolis* ($\beta = 0.5$) and *Swendsen-Wang* ($\beta = 0.4$), respectively.

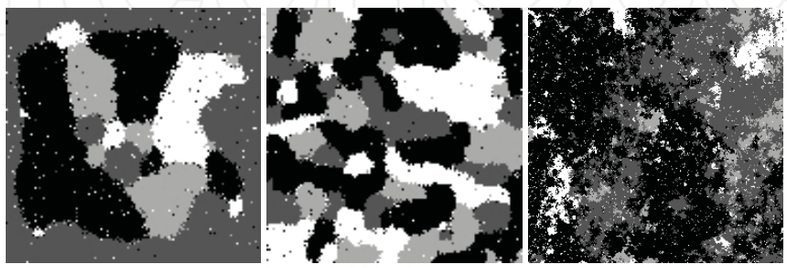


Fig. 9. Synthetic images generated by *MCMC* simulation algorithms using second-order neighbourhood systems for *M=4*: *Gibbs Sampler* ($\beta = 0.45$), *Metropolis* ($\beta = 0.5$) and *Swendsen-Wang* ($\beta = 0.4$), respectively.

The *MPL* estimators, obtained by the derived pseudo-likelihood equations were compared with the real parameter values. This information, together with the test statistics and the *p-values*, obtained from the approximation to the asymptotic variance provide a mathematical procedure to validate and assess the accuracy of the pseudo-likelihood equations. Tables 4 and 5 show the obtained results.

	Swendsen-Wang		*Gibbs Sampler*		*Metropolis*	
M	3	4	3	4	3	4
β	0.4	0.4	0.45	0.45	0.5	0.5
$\hat{\beta}_{MPL}$	0.4460	0.4878	0.3849	0.4064	0.4814	0.4884
$\lvert\beta-\hat{\beta}_{MPL}\rvert$	0.0460	0.0878	0.0651	0.0436	0.0186	0.0111
$\hat{I}^1_{obs}(\boldsymbol{\theta})$	0.4694	0.6825	0.8450	1.3106	0.3908	0.8258
$\hat{I}^2_{obs}(\boldsymbol{\theta})$	3.0080	3.3181	3.8248	4.5387	2.2935	2.6436
$\hat{\sigma}^2_n(\hat{\beta}_{MPL})$	0.0519	0.0620	0.0578	0.0636	0.0743	0.1182
Z_n	0.2458	0.3571	0.2707	0.1729	0.0682	0.0322
p-values	0.8104	0.7264	0.7872	0.8650	0.9520	0.9760

Table 4. *MPL* estimators, observed Fisher information, asymptotic variances, test statistics and p-values for synthetic *MCMC* images using second-order systems.

	Swendsen-Wang		*Gibbs Sampler*		*Metropolis*	
M	3	4	3	4	3	4
β	0.4	0.4	0.45	0.45	0.5	0.5
$\hat{\beta}_{MPL}$	0.3602	0.3772	0.4185	0.4309	0.4896	0.4988
$\lvert\beta-\hat{\beta}_{MPL}\rvert$	0.0398	0.0228	0.0315	0.0191	0.0104	0.0012
$\hat{I}^1_{obs}(\boldsymbol{\theta})$	0.2738	0.5372	0.1104	0.1433	0.0981	0.1269
$\hat{I}^2_{obs}(\boldsymbol{\theta})$	3.5691	4.6800	1.8703	2.3416	1.4165	1.4547
$\hat{\sigma}^2_n(\hat{\beta}_{MPL})$	0.0215	0.0245	0.0316	0.0261	0.0489	0.0600
Z_n	0.2510	0.1456	0.1772	0.1182	0.0470	0.0049
p-values	0.8036	0.8886	0.8572	0.9044	0.9602	0.9940

Table 5. *MPL* estimators, observed Fisher information, asymptotic variances, test statistics and p-values for synthetic *MCMC* images using third-order systems.

The obtained results show that the asymptotic variance is reduced in third-order systems, increasing the *p-values*, suggesting that the use of higher-order systems improves Potts *MRF* model *MPL* estimation. Considering the observed data, we conclude that there are no significant differences between β and $\hat{\beta}_{MPL}$. Therefore, based on statistical evidences, we should accept the hypothesis *H*, assessing the accuracy of the *MPL* estimation method.

4. Contextual Classification and Bayesian Inference

As presented in the previous sections of this chapter, multispectral image contextual classification through *MRF* models is stated as a Bayesian inference problem, since we are interested in the solution that maximizes the *a posteriori* distribution. Several combinatorial optimization algorithms can be used to approximate the *MAP* estimator, although is has been shown that the only optimal method is *Simulated Annealing* (*SA*). However, due to its high computational cost, *SA* may not be the best choice for real applications. In this chapter we discuss two suboptimal methods used in contextual classification: *ICM* and *MPM*. The main difference between these methods is that while the first one is the optimum Bayesian estimator in case of a uniform cost function, the later is optimum regarding a Hamming distance cost function (Won & Gray, 2004).

4.1 *Iterated Conditional Modes*

The *ICM* algorithm was originately proposed by Besag as a computationally feasible iterative and determinístic algorithm for approximating the *MAP* estimator in complex problems. The basic idea consists in, for each pixel, update its current value with the label that maximizes the *a posteriori* probability. By noting that $P\left(\boldsymbol{x}|\boldsymbol{y}\right) = P\left(x_{ij}|\boldsymbol{x}_{ij}^{-},\boldsymbol{y}\right)P\left(\boldsymbol{x}_{ij}^{-}|\boldsymbol{y}\right)$, where $\boldsymbol{x}_{ij}^{-}$ denotes the entire random field without the current element x_{ij}, the subsequent maximization of $P\left(x_{ij}|\boldsymbol{x}_{ij}^{-},\boldsymbol{y}\right)$ is always moving towards the a maximum of the *a posteriori* probability. Thus, *ICM* rapidly converges to a local maximum since its results are strongly dependent on the initialization. The *ICM* algorithm, as described in (Dubes & Jain, 1989) is given in the following.

Algoritm 1. *Iterated Conditinal Modes (ICM)*

1. Chose a *MRF* model for the label field $\boldsymbol{X}$.
2. Initialize $\hat{\boldsymbol{x}}$ by choosing the label $\hat{x}_{ij}$ that maximizes $p\left(y_{ij}|x_{ij}\right)$, that is, the result of maximum likelihood classification.
3. For $i = 1,\ldots,M$ and $j = 1,\ldots,N$
   a. Update the label $\hat{x}_{ij}$ with the value that maximizes

$$p\left(x_{ij}|\boldsymbol{x}_{ij}^{-},\boldsymbol{y}\right) \propto p\left(y_{ij}|x_{ij}\right)p\left(x_{ij}|\eta_{ij}^{s}\right)$$

4. Repeat (3) N_{iter} times

So, in our case, we always update the current label with the new value that maximizes the product between the *LCDF's* of *GMRF* and Potts models. Note, however, that step (2) is being generalized, since instead of maximum likelihood classification, we can initialize *ICM* using several pattern classifiers.

4.2 Maximizer of the Posterior Marginals

As the name says, the *MPM* estimator is obtained by maximizing the posterior marginal probabilities $P\left(x_{ij}|\boldsymbol{y}\right)$. Thus, the fundamental point here is the calculation of these distributions. The *MPM* algorithm, as proposed in (Marroquin et al., 1987), uses *MCMC* methods to approximate these distributions. Basically, the *MPM* algorithm simulate a Markov chain over the states that represent all possible configurations of the random field. The idea is that as each pixel is repeatedly visited, the resulting Markov chain generates a sample of the posterior distribution $P\left(\boldsymbol{x}|\boldsymbol{y}\right)$, regardless of the initial conditions. As a result of that, a sequence of configurations $x(0)\rightarrow x(1)\rightarrow\cdots x(n)\rightarrow\cdots$, corresponding to a Makov chain that reaches its equilibrium state, is generated. Once this state is reached, we can regard all configurations from this point as a sample of $P\left(\boldsymbol{x}|\boldsymbol{y}\right)$. Besides, in the equilibrium state the expected value of a function of a random variable can be estimated through the ergodic principle. Using this result, the posterior marginal distribution for the *MPM* algorihtm can be approximated by (Dubes & Jain, 1989):

$$P\left(x_{ij}=g|\boldsymbol{y}\right)\approx\frac{1}{(n-k)}\sum_{p=k+1}^{n}\delta\left(x_{ij}^{p}-g\right) \tag{28}$$

where k is the number of steps needed to the sequence to stabilize and n is a sufficiently large number of iterations so that the estimation is accurate at a certain reasonable computational cost. One problem with this approach is exactly how to choose these values (also known as *magic numbers*). Usually, they are both chosen empiracally. The pseudo-code for *MPM*, as described in (Dubes & Jain, 1989), is shown in Algorithm 2.

5. Metrics for Performance Evaluation of Image Classification

In order to evaluate the performance of the contextual classification in an objective way, the use of quantitative measures is required. Often, the most widely used criteria for evaluation of classification tasks are the correct classification rate and/or the estimated classification error (*holdout, resubstitution*). However, these measures do not allow robust statistical analysis, neither inference about the obtained results. To surmount this problem, statisticians usually adopt Cohen's *Kappa* coefficient, a measure to assess the accuracy in classification problems.

5.1 Cohen's *Kappa* Coefficient

This coefficient was originally proposed by Cohen (Cohen, 1960), as a measure of agreement between rankings and opinions of different specialists. In pattern recognition, this coefficient determines a degree of agreement between the *"ground truth"* and the output of a given classifier. The better the classification performance, the higher is the *Kappa* value. In case of perfect agreement, *Kappa* is equal to one. The main motivation for the use of *Kappa* is that it has good statistical properties, such as asymptotic normality, and also the fact that it is easily computed from the confusion matrix.

Algorithm 2. *Maximizer of the Posterior Marginals* (*MPM*)

1. Chose a *MRF* model for the label field $\boldsymbol{X}$.
2. Initialize $\hat{\boldsymbol{x}}$ by choosing the label $\hat{x}_{ij}$ that maximizes $p\left(y_{ij}|x_{ij}\right)$, that is, the result of maximum likelihood classification.
3. For $i = 1,\ldots,M$ and $j = 1,\ldots,N$
   a. Choose a random label g and define $z_{ij} = g$. Let $z_{kl} = x_{kl}$ for all $\left(k,l\right) \neq \left(i,j\right)$
   b. Let $p = \min\left\{1, \dfrac{P\left(X=z|Y=y\right)}{P\left(X=x|Y=y\right)}\right\}$
   c. Replace x by z with probability p.
4. Repeat (3) N times, saving the realizations from $x^{(l+1)}$ to $x^{(n)}$
5. Calculate $P\left(x_{ij}=g|\boldsymbol{y}\right)$ for all pixels according equation (28).
6. For $i = 1,\ldots,M$ and $j = 1,\ldots,N$
   a. Choose the label $\hat{x}_{ij}$ that maximizes the posterior marginal among all possible labels
   $P\left(x_{ij}=\hat{x}_{ij}|\boldsymbol{y}\right) > P\left(x_{ij}=g|\boldsymbol{y}\right)$

The expression for the *Kappa* coefficient from the confusion matrix is given by (Congalton, 1991):

$$\hat{K} = \frac{N\sum_{i=1}^{c} c_{ii} - \sum_{i=1}^{c} x_{i+}x_{+i}}{N^2 - \sum_{i=1}^{c} x_{i+}x_{+i}} \tag{29}$$

where c_{ii} is an element of the confusion matrix, x_{+i} is the sum of the elements of column *i*, x_{i+} is the sum of the elements of the row *i*, c is the number of classes and *N* is the number of training samples. The asymptotic variance of this estimator is given by:

$$\hat{\sigma}_K^2 = \frac{1}{N}\left[\frac{\theta_1\cdot\left(1-\theta_1\right)}{\left(1-\theta_2\right)^2}+\frac{2\cdot\left(1-\theta_1\right)\cdot\left(2\cdot\theta_1\cdot\theta_2-\theta_3\right)}{\left(1-\theta_2\right)^3}+\frac{\left(1-\theta_1\right)^2\cdot\left(\theta_4-4\cdot\theta_2^2\right)}{\left(1-\theta_2\right)^4}\right] \tag{30}$$

where

$$\theta_1 = \frac{1}{N}\sum_{i=1}^{c} x_{ii} \qquad \theta_2 = \frac{1}{N^2}\sum_{i=1}^{c} x_{i+}x_{+i} \tag{31}$$

$$\theta_3 = \frac{1}{N^2}\sum_{i=1}^{c} x_{ii}\cdot\left(x_{i+}+x_{+i}\right) \qquad \theta_4 = \frac{1}{N^3}\sum_{i=1}^{r}\sum_{j=1}^{r} x_{ij}\cdot\left(x_{j+}+x_{+i}\right)^2 \tag{32}$$

6. Experiments and Results in Nuclear Magnetic Resonance Images

In order to test and evaluate the contextual classification methods previously described in this chapter, we show some experiments in *NMR* images of *marmocets* brains, a monkey species often used in medical experiments. These images were acquired by the *CInAPCe* project, an abbreviation for the Portuguese expression "*Inter-Institutional Cooperation to Support Brain Research*", a Brazilian research project that has as main purpose the establishment of a scientific network seeking the development of neuroscience research through multidisciplinary approaches. In this sense, pattern recognition can contribute to the development of new methods and tools for processing and analyzing magnetic resonance imaging and its integration with other methodologies in the investigation of epilepsies. Figure 9 shows the bands *PD*, *T1* and *T2* of a *marmocet NMR* multispectral brain image.

Fig. 9. *PD*, *T1* and *T2 NMR* image bands of the multispectral *marmocet* brain image.

The contextual classification of the multispectral image was performed by applying both *ICM* and *MPM* using second and third order neighbourhood systems on several initializations provided by seven different pattern classifiers: Quadratic Bayesian Classifier (*QDC*) and Linear Bayesian Classifier (*LDC*) under Gaussian hypothesis, Parzen-Windows Classifier (*PARZENC*), K-Nearest Neighbour Classifier (*KNNC*), Logistic Classifier (*LOGLC*),

Nearest Mean Classifier (*NMC*) and Decision Tree Classifier (*TREEC*). All the experiments were implemented in *MATLAB*, using the pattern recognition toolbox *PRTOOLS v.4.1*[1], developed at Delft University, to provide the initializations through each one of the above classifiers.

The experiments were conducted to clarify some hypotheses and conjectures regarding contextual classification. We want to verify the following hypothesis:

A. Contextual classification is capable of significantly improving the performance of ordinary classification techniques (punctual methods).
B. Different initializations can lead to statistically different contextual classification results (for the same iterative algorithm).
C. The use of higher-order systems is capable of significantly improving the performance of contextual classification.
D. Different contextual classification algorithms are capable of producing statistically different results (for the same initialization).

We used 100 training samples for each class: white matter, gray matter and background. The classification errors and confusion matrix are estimated by the *10-Fold cross-validation* method. Convergence was considered by achieving one of two conditions: less than 1% of the pixels are updated in the current iteration, or the maximum of 5 iterations is reached. Tables 6, 7, 8, 9 and 10 show the obtained results.

Classifiers	$\hat{k}$	$\hat{\sigma}^2_{kappa}$
PARZENC	0.7816	0.00031061
KNNC	0.7550	0.00034100
LOGLC	0.7583	0.00033502
LDC	0.7716	0.00032177
QDC	0.7866	0.00030515
NMC	0.7850	0.00030629
TREEC	0.6500	0.00044737

Table 6. *Kappa* coefficients and variances for punctual classification results.

Classifiers	$\hat{k}$	$\hat{\sigma}^2_{kappa}$
PARZENC	0.9783	3.5584e-005
KNNC	0.9733	4.3614e-005
LOGLC	0.9900	1.6553e-005
LDC	0.9916	1.3811e-005
QDC	0.9700	4.8952e-005
NMC	0.9966	5.5431e-006
TREEC	1.0000	0.0000

Table 7. *Kappa* coefficients and variances for *MPM* classification on second order systems.

Classifiers	$\hat{k}$	$\hat{\sigma}^2_{kappa}$
PARZENC	0.9966	5.5431e-006

[1] *Available online at http://www.prtools.org*

KNNC	1.0000	0.0000
LOGLC	0.9950	8.3055e-006
LDC	0.9966	5.5431e-006
QDC	0.9966	5.5431e-006
NMC	0.9933	1.1062e-005
TREEC	0.9966	5.5431e-006

Table 8. *Kappa* coefficients and variances for *MPM* classification on third order systems.

Classifiers	$\hat{k}$	$\hat{\sigma}^2_{kappa}$
PARZENC	0.9700	4.8975e-005
KNNC	0.9550	7.2614e-005
LOGLC	0.9600	6.4864e-005
LDC	0.9616	6.2206e-005
QDC	0.9516	7.7788e-005
NMC	0.9583	6.7446e-005
TREEC	1.0000	0.0000

Table 9. *Kappa* coefficients and variances for *ICM* classification on second order systems.

Classifiers	$\hat{k}$	$\hat{\sigma}^2_{kappa}$
PARZENC	0.9983	2.7747e-006
KNNC	0.9866	2.202e-005
LOGLC	0.9983	2.7747e-006
LDC	0.9950	8.3053e-006
QDC	0.9850	2.4743e-005
NMC	0.9950	8.3053e-006
TREEC	1.0000	0.0000

Table 10. *Kappa* coefficients and variances for *ICM* classification on third order systems.

6.1 Statistical Analysis

To test the hypothesis and validate the proposed methodology for contextual classification, both local (confidence intervals) and global (*T* test) analysis are performed. Let $\overline{k}_1$ and $\overline{k}_2$ be the mean *Kappa* coefficients before and after the application of a given technique. Defining $\overline{k} = \overline{k}_1 - \overline{k}_2$, it is desirable to test the following hypothesis:

$$\begin{aligned} H_0 &: \overline{k} = 0 \\ H_1 &: \overline{k} \neq 0 \end{aligned} \tag{33}$$

The test statistic *T*, defined as follows, has a *t-student* distribution with *n-1* degrees of freedom (*n=7* in this case):

$$T = \frac{\overline{k}}{\sigma_{\overline{k}} / \sqrt{n}} \tag{34}$$

where $\sigma_{\bar{k}}$ denotes the standard deviation of the punctual differences. Thus, considering a α test size (i.e., $\alpha = 0.05$ or $\alpha = 0.01$), H_0 must be rejected if T is less than a critical value t_c. This information, together with the *p-values*, allows a robust statistical analysis, as well as inferences about the problem in study. Note that the decision based on the T statistic is quite intuitive, since the greater the difference between the two means, more chance that we are dealing with distinct groups (captured by the numerator of T). On the other hand, the greater the variability of the results, more difficult is to detect differences on the means (captured by the denominator of T).

To verify the hypothesis A, a T test was performed using the data presented in Tables 6 and 9 (punctual classification x *ICM* on second order systems). Considering a test size $\alpha = 0.05$, we have a critical value of $t_c = -1.943$. The obtained results indicate strong evidences against H_0, since $\bar{k} = -0.2098$, $T = -8.7741$, leading to a *p-value* smaller than 0.0005. Therefore, we should reject H_0, assessing that combinatorial optimization algorithms can significantly improve the classification performance. Figure 10 shows a comparison of the visual results obtained by the *LOGLC* classifier and the *LOGLC+ICM* classification.

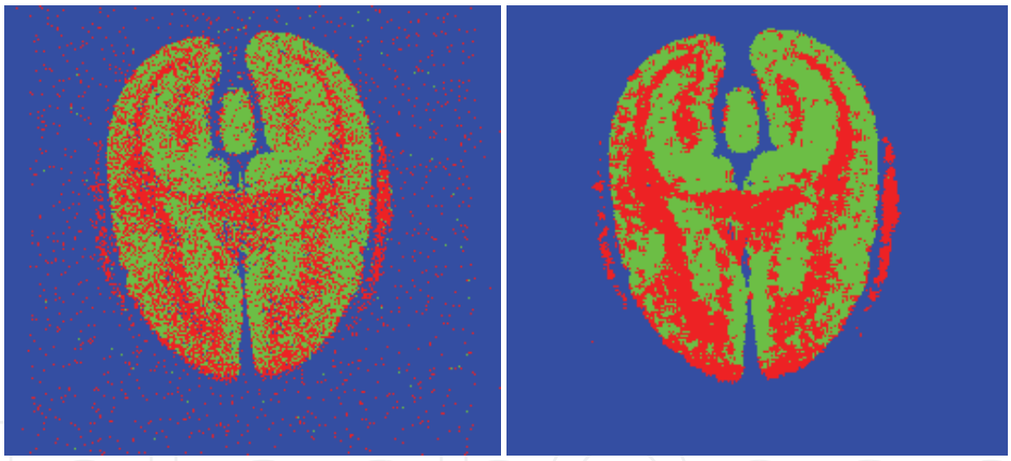


Fig. 10. Comparison between classification maps for the *marmocet* brain multispectral *NMR* image obtained using *LOGLC* and *LOGLC+ICM*

The hypothesis B was tested by simply constructing 95% confidence intervals (*CI*) for the respective *Kappa* coefficients. To illustrate the scenario, we compared the results of *KNNC+ICM* and *TREEC+ICM* classification form Table 9. The confidence intervals show that the results are statistically different, since for the *KNNC+ICM* we have [0.9387, 0.9713] and the *TREEC+ICM* provides a *Kappa* coefficient equal to one and with zero variance. Figure 11 compares the visual results. Actually, these results were expected since both combinatorial optimization algorithms *ICM* and *MPM* are sub-optimal, that is, they converge to different local maxima depending on the initialization.

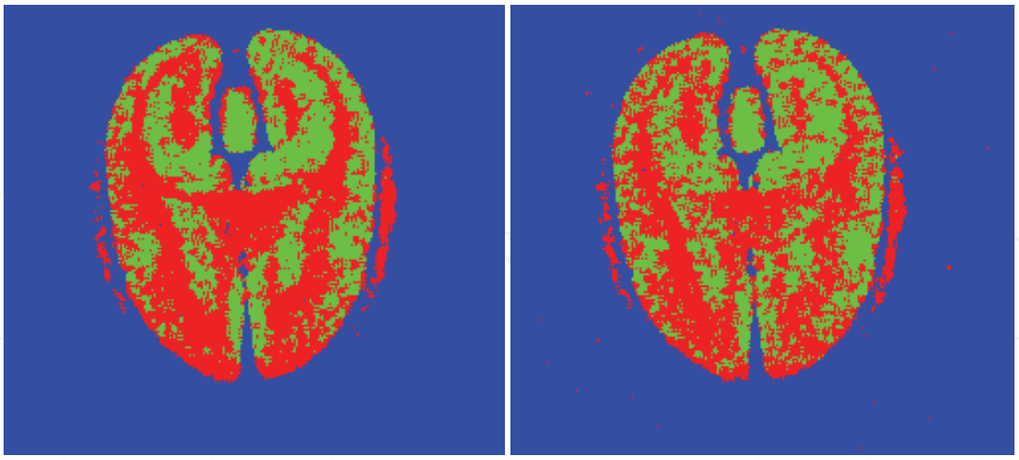


Fig. 11. Comparison between classification maps for the *marmocet* brain multispectral NMR image obtained using *KNNC+ICM* and *TREEC+ICM*.

To test the third hypothesis (C), a T test was performed in data from Tables 9 and 10 to compare the mean performances. The results show that the mean performances are significantly different, since $\bar{k} = -0.0288$, $T = -5.8115$, leading to a *p-value* smaller than 0.005 and once again, strong evidences against H_0. Figures 12 and 13 shows a comparison between *LOGLC+ICM* and *NMC+ICM* on second and third order systems.

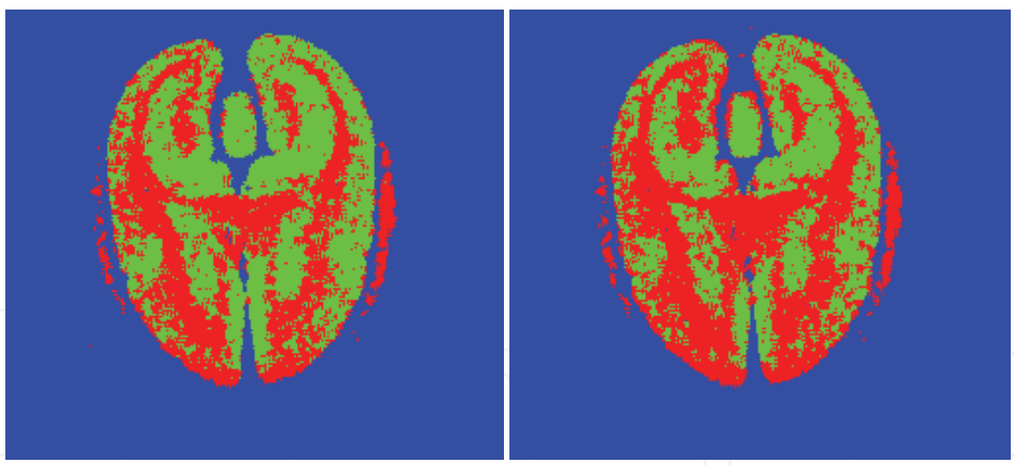


Fig. 12. Classification maps for *LOGLC+ICM* and *NMC+ICM* on second order systems.

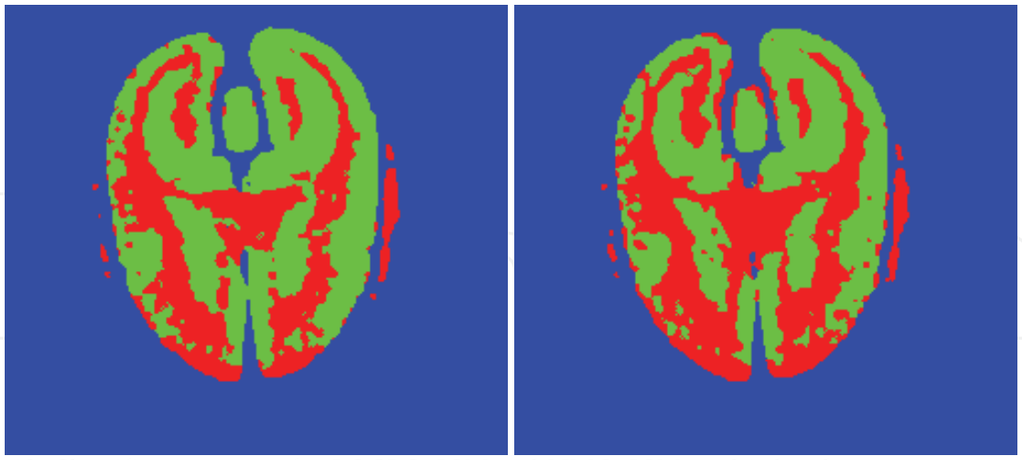


Fig. 13. Classification maps for *LOGLC+ICM* and *NMC+ICM* on third order systems.

Finally, to test the last hypothesis, 95% confidence intervals were constructed using the *Kappa* coefficient values together with the asymptotic variances. From Tables 7 and 9 it is possible to observe that *NMC+MPM* produces the interval [0.9920, 1], while *NMC+ICM* gives [0.9423, 0.9743], assessing that the performances are significantly different. This results and the result obtained on testing the hypothesis B suggest that the use of classifier combination rules can be explored in contextual classification problems, since there is significant differences in the results, providing complementary information that can be used to improve the performance even more. Figure 14 shows a comparison between the results of *MPM* and *ICM* for the same initialization (*NMC*).

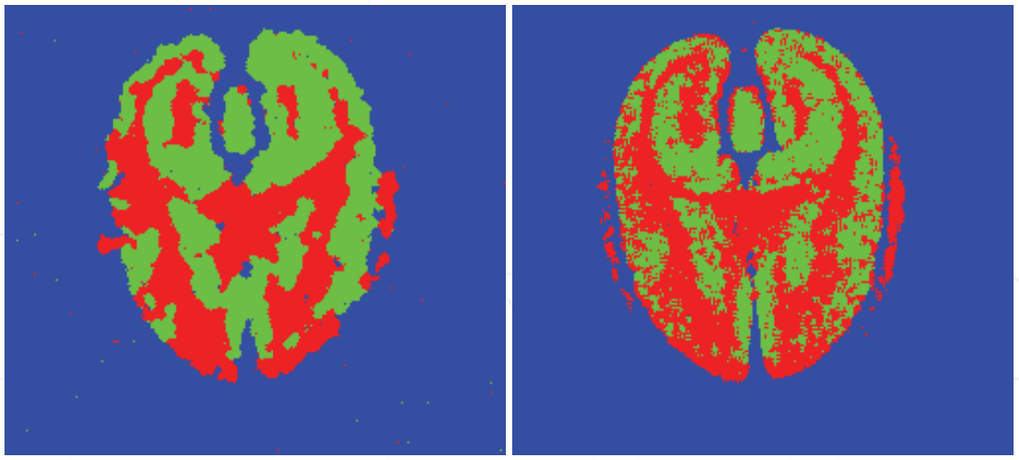


Fig. 14. Classification maps for *NMC+MPM* and *NMC+ICM* on second order systems.

7. Conclusions

In this chapter we discussed three important problems in pattern recognition. First, the derivation of novel pseudo-likelihood equations for Potts MRF model parameter estimation on higher-order neighbourhood systems. Then, the accuracy of MPL estimation was

assessed through approximations for the asymptotic variance of these estimators. Finally, multispectral image contextual classification was stated as a Bayesian inference problem. The obtained results show that the approach discussed here is valid, and more, is capable of significantly improving classification performance. Future works in this research area include the study about the efficiency of the MPL estimation through the analysis of necessary/sufficient conditions of information equality in MRF models, as well as the combination of contextual classifiers aiming for a further improvement in classification performance.

8. Acknowledgments

We would like to thank FAPESP for the financial support through Alexandre L. M. Levada student scholarship (grant n. 06/01711-4). We also would like to thank Hilde Buzzá for the NMR images acquisition and for all the support and assistance throughout this process.

9. References

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society – Series B,* vol.36, pp. 192-236.

Besag, J. (1986). On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society – Series B,* vol.48, n. 3, pp. 259-302.

Bickel, P. J. (1991). *Mathematical Statistics,* Holden Day, New York.

Brent, R. P. (1973). *Algorithms for minimization without derivatives,* Prentice Hall, New York.

Casella, G. & Berger, R. L. (2002). *Statistical Inference,* 2nd Edition, Duxbury, New York.

Cohen, J. A. (1960). A coefficient of agreement for nominal scales, *Educational and Psycological Measurement*, v. 20, n. 1, pp. 37-46.

Congalton, R. G. (1991). A review of assessing the accuracy of classification of remotely sensed data, *Remote Sensing of Enviroment*, v. 37, pp. 35-46.

Dubes, R. & Jain, A. (1989). Random field models in image analysis, *Journal of Applied Statistics*, v. 16, n. 2, pp. 131-164.

Efron, B. F. & Hinkley, D. V. (1978). Assessing the accuracy of the ML estimator : observed versus expected Fisher information, *Biometrika*, vol. 65, pp. 457-487.

Geman, S. & Geman, D. (1984). Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Trans. On Pattern Analysis and Machine Intelligence,* vol.6, n. 6, pp. 721-742.

Hammersley, J. M. & Clifford, P. (1971). Markov field on finite graphs and lattices. *Unpublished*.

Jensen, J. L. & Künsh, H. R. (1994). On asymptotic normality of pseudo likelihood estimates for pairwise interaction processes. *Annals of the Institute of Statistical Mathematics,* vol.46, n. 3, pp. 475-486.

Lehmann, E. L. (1983). *Theory of Point Estimation,* Wiley, New York.

Levada, A. L. M. ; Mascarenhas, N. D. A. & Tannús, A. (2008a). A novel pseudo-likelihood equation for Potts MRF model parameter estimation in image analysis, *Proceedings of the 15th International Symposium on Image Processing (ICIP)*, pp. 1828-1831, October 2008, IEEE, San Diego.

Levada, A. L. M. ; Mascarenhas, N. D. A. & Tannús, A. (2008b). Pseudo-likelihood equations for Potts MRF model parameter estimation on higher-order neighbourhood systems, *IEEE Trans. On Geoscience and Remote Sensing Letters*, vol. 5, pp. 522-526.

Levada, A. L. M. ; Mascarenhas, N. D. A. & Tannús, A. (2008c). On the asymptotic variances of Gaussian Markov Random Field model hyperparameters in stochastic image modeling, *Proceedings of the 19th International Symposium on Pattern Recognition (ICPR)*, pp. 1-4, December 2008, IEEE, Tampa.

Levada, A. L. M. ; Mascarenhas, N. D. A. ; Tannús, A. & Salvadeo, D. H. P. (2008d). Spatially non-homogeneous Potts model parameter estimation on higher-order neighbourhood systems by maximum pseudo-likelihood, *Proceedings of the 23rd Annual ACM Symposium on Applied Computing (ACM SAC)*, pp. 1733-1737, March 2008, ACM, Fortaleza, Brazil.

Liang, G. & Yu, B. (2003). Maximum pseudo likelihood estimation in network tomography, *IEEE Trans. On Signal Processing*, vol. 51, n. 8, pp. 2043-2053.

Marroquin, J. L.; Mitter, S. K. & Poggio, T. A. (1987). A probabilistic solution of ill-posed problems in computational vision. *Journal of the American Statistical Society,* vol.82, n. 397, pp. 76-89.

Metropolis, N.; Rosenbluth, A.; Rosenbluth, M.; Teller, A. & Teller, E. (1953). Equation of state calculations by fast computer machines. *Journal of Physical Chemistry,* vol. 21, n. 6, pp. 1087-1092.

Swendsen, R. & Wang, J. (1987). Nonuniversal critical dynamics in Monte Carlo simulations. *Physical Review Letters,* vol.58, pp. 86-88.

Winkler, G. (2006). *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods: A Mathematical Introduction,* 2nd Edition, *Springer-Verlag*, New York.

Wolff, U. (1989). Collective Monte Carlo updating for spin systems. *Physical Review Letters,* vol.62, pp. 361-364.

Won, C. S. & Gray, R. M. (2004). *Stochastic Image Processing,* Kluwer Academic/Plenum Publishers, New York.

Yamazaki, T. & Gingras, D. (1996). A contextual classification system for remote sensing using a multivariate Gaussian MRF model, *Proceedings of the 9th International Symposium on Circuits and Systems (ISCAS)*, pp. 648-651, May 1996, IEEE, Atlanta.

Yu, S. & Berthod, M. (1995). A game strategy approach for image labeling. *Computer Vision and Image Understanding,* vol.61, n. 1, pp. 32-37.

Pattern Recognition
Edited by Peng-Yeng Yin

ISBN 978-953-307-014-8
Hard cover, 568 pages
Publisher InTech
Published online 01, October, 2009
Published in print edition October, 2009

For more than 40 years, pattern recognition approaches are continuingly improving and have been used in an increasing number of areas with great success. This book discloses recent advances and new ideas in approaches and applications for pattern recognition. The 30 chapters selected in this book cover the major topics in pattern recognition. These chapters propose state-of-the-art approaches and cutting-edge research results. I could not thank enough to the contributions of the authors. This book would not have been possible without their support.

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

IntechOpen