



Inference Tools for Markov Random Fields on Lattices: The R Package `mrf2d`

Victor Freguglia 
University of Campinas

Nancy Lopes Garcia 
University of Campinas

Abstract

Markov random fields on two-dimensional lattices are behind many image analysis methodologies. `mrf2d` provides tools for statistical inference on a class of discrete stationary Markov random field models with pairwise interaction, which includes many of the popular models such as the Potts model and texture image models. The package introduces representations of dependence structures and parameters, visualization functions and efficient (C++-based) implementations of sampling algorithms, common estimation methods and other key features of the model, providing a useful framework to implement algorithms and working with the model in general. This paper presents a description and details of the package, as well as some reproducible examples of usage.

Keywords: Markov random fields, image analysis, R, Gibbs random fields, Potts model, texture.

1. Introduction

A Markov random field (MRF) is a generalization of the well-known concept of a Markov chain where variables are indexed by vertices of a graph instead of a sequence and the notion of memory is substituted by the neighborhood (edges) of that graph. Markov random fields on lattices, or more generally, Gibbs distributions, have been studied in statistical mechanics as models for interacting particle systems. They range from the basic Ising model (or its generalization Potts model) with pairwise nearest-neighbor interaction to models with more complex interaction types, presenting long-range and/or higher-order interaction. For an introduction to the subject we refer to [Liggett \(2012\)](#) and references therein.

A finite 2-dimensional lattice is a direct representation of pixel positions on a digital image. [Geman and Geman \(1984\)](#) make an analogy between image models and statistical mechanics systems, introducing probability-based computational methods for image restoration under a specific type of noise. Higher-order dependence structures are also described, for example,

interactions with pixels more distant than nearest-neighbors. Cross and Jain (1983) use MRFs with special interaction structures to model texture images.

Many modern image analysis methodologies in statistics and machine learning are grounded on Markov random field theory and the local dependence characteristic of image data. Common tasks in image analysis involve image segmentation (Zhang, Brady, and Smith 2001; Kato and Pong 2006; Roche, Ribes, Bach-Cuadra, and Krüger 2011; Cao, Zhou, Xu, Meng, Xu, and Paisley 2018; Ghamisi *et al.* 2018), texture synthesis (Gimel'farb 1996; Freeman and Liu 2011; Versteegen, Gimel'farb, and Riddle 2016) and statistical modeling (Derin and Elliott 1987; Guillot, Rajaratnam, and Emile-Geay 2015; Freguglia, Garcia, and Bicas 2020) all of which can be achieved with the use of MRFs. Some basic references are Blake, Kohli, and Rother (2011) and Kato and Zerubia (2012).

In this paper, when we refer to a MRF, we consider the particular case where variables are indexed by points of a 2-dimensional lattice, not a general graph structure. The regular grid naturally creates a spatial structure and notions of distance and direction for the variables, allowing models to be specified based on this spatial structure (see Besag 1974, for examples).

Parametric inference based on maximum likelihood for such models is difficult, even for the simple models, because of the intractable constant that appears in the likelihood. Inference for the simplest non-trivial case of the Ising model was first studied by Pickard (1987) and continues to present challenges, see for example Bhattacharya and Mukherjee (2018). On the other hand, while there is a continuous development of methodologies used in MRFs in the theoretical field, implementing new algorithms is a challenge in practice, mostly due to the high-dimensionality of the problem and the complexity of the data structures required to represent the data in this type of problem. An overview of this topic, mainly from the Bayesian perspective, can be found in Winkler (2012).

Most methodologies developed are based on Monte Carlo Markov chain methods, thus simple tasks like evaluating pairs of pixels or sampling individual pixels need to be repeated millions or billions of times in iterative methods, depending on the image size, making an efficient implementation of such methods one of the main demands for researchers of the topic.

R (R Core Team 2021) is one of the most used programming languages among statistics researchers, what makes the existence of good packages important for any field of statistics. In a general context (considering the definition of MRFs with general undirected graphs), packages like **graph** (Gentleman, Whalen, Huber, and Falcon 2021) and **network** (Butts 2008) provides tools for representation and manipulation of graph structures which can be used for constructing and visualizing graph-based models. Different versions of graph-based MRFs appear in many packages. For example, the **CRF** package (Wu 2019) has inferential tools Markov random fields with pairwise and unary interactions and their hidden MRF version, **MRFcov** (Clark, Wells, and Lindberg 2018) allows inference for the interaction parameter of between nodes of a graph considering covariates and **gamlss.spatial** (De Bastiani, Rigby, Stasinopoulous, Cysneiros, and Uribe-Opazo 2018) allows fitting Gaussian Markov random fields in a spatial context, similar to **INLA** (Rue, Martino, and Chopin 2009) and **mgcv** (Wood 2017).

Outside of the R ecosystem, there are powerful software in C++ used for image analysis related to Markov random fields, such as the **DGM** library (Kosov 2013) and **densecrf** (Krähenbühl and Koltun 2011), which also has a Python wrapper (Beyer 2015), and can be used for a variety of tasks and use extremely efficient computational methods.

For discrete MRFs on lattice data, closer to what is proposed in **mrf2d**, there are some R packages available. The **potts** package (Geyer and Johnson 2020), implements simulation algorithms and parameter estimation via composite-likelihood for a Potts model with nearest-neighbor interactions only. **PottsUtils** (Feng 2018) also implements simulation and tools for computing normalization constants in one, two and three-dimensional Potts model. The package **bayesImageS** (Moore, Nicholls, Pettitt, and Mengersen 2020) provides Bayesian image segmentation algorithms considering Gaussian mixtures driven by hidden Potts models with slightly more complex interaction neighborhood. **GiRaF** (Stoehr, Pudlo, and Friel 2020) allows calculation on, and sampling from general homogeneous Potts model. The **Pottslab** (Storath and Weinmann 2014) package for MATLAB also provides image segmentation algorithms using the Potts model, including for multivariate-valued data.

Although the available packages for discrete-valued MRFs offer efficient implementations of their methods, they do not provide an interface that allows simple extensions to different cases, for example, different interaction types for different positions and sparse long-range interaction neighborhoods. Some of the algorithms used also rely on specific characteristics of the specific setups they consider and cannot be applied more generally.

The **mrf2d** package (Freguglia 2022) provides a complete framework for statistical inference on discrete-valued MRF models on 2-dimensional lattice data, where all the elements used by algorithms (such as conditional probabilities, pseudo-likelihood function, simulation, sufficient statistics and more) are available for the user, as well as many built-in model fitting functions.

The package uses the model described in Freguglia *et al.* (2020) as a reference. Many other models, such as the Potts model and auto-models, are particular cases of our model obtained by including restrictions to the parameters or using specific interacting neighborhoods. These neighborhoods can be freely specified within the package and 5 families of parameter restrictions are available to cover the particular cases.

mrf2d (Freguglia 2022) is available from the Comprehensive R Archive Network (CRAN) at <https://CRAN.R-project.org/package=mrf2d> and can be installed and loaded with

```
R> install.packages("mrf2d")
R> library("mrf2d")
```

The development version is in its GitHub repository and can be installed with

```
R> devtools::install_github("Freguglia/mrf2d")
```

This paper is organized as follows. Section 2 describes the model considered in **mrf2d**, Section 3 presents the main functionalities of the package and details of the implementation, which are illustrated by examples in Section 4. We finish with a discussion in Section 5. All the results of example code in this article were obtained using R version 4.0.2 and **mrf2d** version 1.0.

2. Model description

Let $\mathcal{L} \subset \{\mathbf{i} = (i_1, i_2) \in \mathbb{N}^2\}$ be a finite set of locations in a two-dimensional lattice region and $\mathbf{Z} = \{Z_{\mathbf{i}}\}_{\mathbf{i} \in \mathcal{L}}$ a field of random variables indexed by those locations.

The main purpose of **mrf2d** is to provide a general framework for Markov random field models which satisfy the following assumptions:

- (a) **Finite support** Each $Z_{\mathbf{i}}$ can take values in $\mathcal{Z} = \{0, \dots, C\}$ for some finite $C > 0$.
- (b) **Pairwise interactions** The probability of a complete configuration $\mathbb{P}(\mathbf{Z} = \mathbf{z})$ can be decomposed into a product of functions of the pairs $(z_{\mathbf{i}}, z_{\mathbf{j}}), \mathbf{i} \neq \mathbf{j} \in \mathcal{L}$.
- (c) **Homogeneous interactions** For any two pixels \mathbf{i} and \mathbf{j} and any relative position \mathbf{r} , the interaction between pixels \mathbf{i} and $\mathbf{i} + \mathbf{r}$ is the same as for pixels \mathbf{j} and $\mathbf{j} + \mathbf{r}$, i.e., the interaction depends on the relative position of the pair of pixels, not on their position in the lattice.

These assumptions are satisfied by most commonly used models in image processing.

We use the representation in [Freguglia et al. \(2020\)](#) which expresses the probability distribution of the random field in the form of the exponential family and introduce additional constraints to parameter space and/or different dependence structures to include particular features of the model under study.

2.1. Homogeneous Markov random field with pairwise interactions

MRF models are characterized by their conditional independence property. Let \mathcal{N} be a neighborhood system on \mathcal{L} , then \mathbf{Z} is a *Markov random field* with respect to \mathcal{N} if $Z_{\mathbf{i}}$ given its neighbors $\mathbf{Z}_{\mathcal{N}_{\mathbf{i}}}$ is conditionally independent from all other variables

$$\mathbb{P}(Z_{\mathbf{i}} = z_{\mathbf{i}} \mid \mathbf{Z}_{-\mathbf{i}}) = \mathbb{P}(Z_{\mathbf{i}} = z_{\mathbf{i}} \mid \mathbf{Z}_{\mathcal{N}_{\mathbf{i}}}), \quad \mathbf{i} \in \mathcal{L}, \quad (1)$$

where $\mathbf{Z}_{-\mathbf{i}}$ denotes the set of variables $\{Z_{\mathbf{j}}, \mathbf{j} \neq \mathbf{i}\}$.

To start defining MRFs in an image processing context, a location of the lattice $\mathbf{i} \in \mathcal{L}$ will be referred as a *pixel* \mathbf{i} and an observed value of the variable $z_{\mathbf{i}} \in \mathcal{Z}$ as *pixel value* or *color*.

We denote by $\mathcal{R} \subset \mathbb{Z}^2$ a set of interacting relative positions such that, for no pair of elements $\mathbf{r}, \mathbf{r}' \in \mathcal{R}$ we have $\mathbf{r}' = -\mathbf{r}$ (no position in \mathcal{R} is a reflection of another). Based on \mathcal{R} , we can construct a neighborhood system (interaction structure) \mathcal{N} in such way that the set of neighbors of site \mathbf{i} , $\mathcal{N}_{\mathbf{i}}$ can be represented by a graph with vertices \mathcal{L} where there is an edge connecting \mathbf{i} and \mathbf{j} if, and only if, $\mathbf{j} = \mathbf{i} \pm \mathbf{r}$. For example, a nearest-neighbor structure corresponds to $\mathcal{R} = \{(1, 0), (0, 1)\}$.

Given an interaction structure \mathcal{R} , for any relative position $\mathbf{r} \in \mathcal{R}$ the interactions associated to that relative position are characterized by a map $\theta_{\mathbf{r}}(\cdot, \cdot)$, $\theta_{\mathbf{r}} : \mathcal{Z}^2 \rightarrow \mathbb{R}$. For $a, b \in \mathcal{Z}$, the value $\theta_{\mathbf{r}}(a, b)$ is called a potential.

The model in **mrf2d** considers a neighborhood system \mathcal{N} that connects pairs of pixel positions \mathbf{i}, \mathbf{j} such that $\mathbf{i} - \mathbf{j} \in \mathcal{R}$. Under assumptions (a), (b) and (c), the Hammersley-Clifford theorem ([Hammersley and Clifford 1971](#)) implies that the probability function for \mathbf{Z} belongs to the exponential family and can be described by a set of natural parameters $\boldsymbol{\theta} = \{\theta_{\mathbf{r}}(a, b), \mathbf{r} \in \mathcal{R}, a, b, \in \mathcal{Z}\}$,

$$\mathbb{P}(\mathbf{Z} = \mathbf{z}) = \frac{1}{\zeta_{\boldsymbol{\theta}}} e^{H(\mathbf{z}, \boldsymbol{\theta})}, \quad (2)$$

where

$$H(\mathbf{z}, \boldsymbol{\theta}) = \sum_{\mathbf{r} \in \mathcal{R}} \sum_{\mathbf{i}, \mathbf{j} \in \mathcal{L}} \theta_{\mathbf{r}}(z_{\mathbf{i}}, z_{\mathbf{j}}) \mathbb{1}_{(\mathbf{j}=\mathbf{i}+\mathbf{r})} \text{ and } \zeta_{\boldsymbol{\theta}} = \sum_{\mathbf{z}'} e^{H(\mathbf{z}', \boldsymbol{\theta})}. \quad (3)$$

Figure 1 illustrates how the function $H(\mathbf{z}, \boldsymbol{\theta})$ is computed for an example interaction structure \mathcal{R} and a field \mathbf{z} .

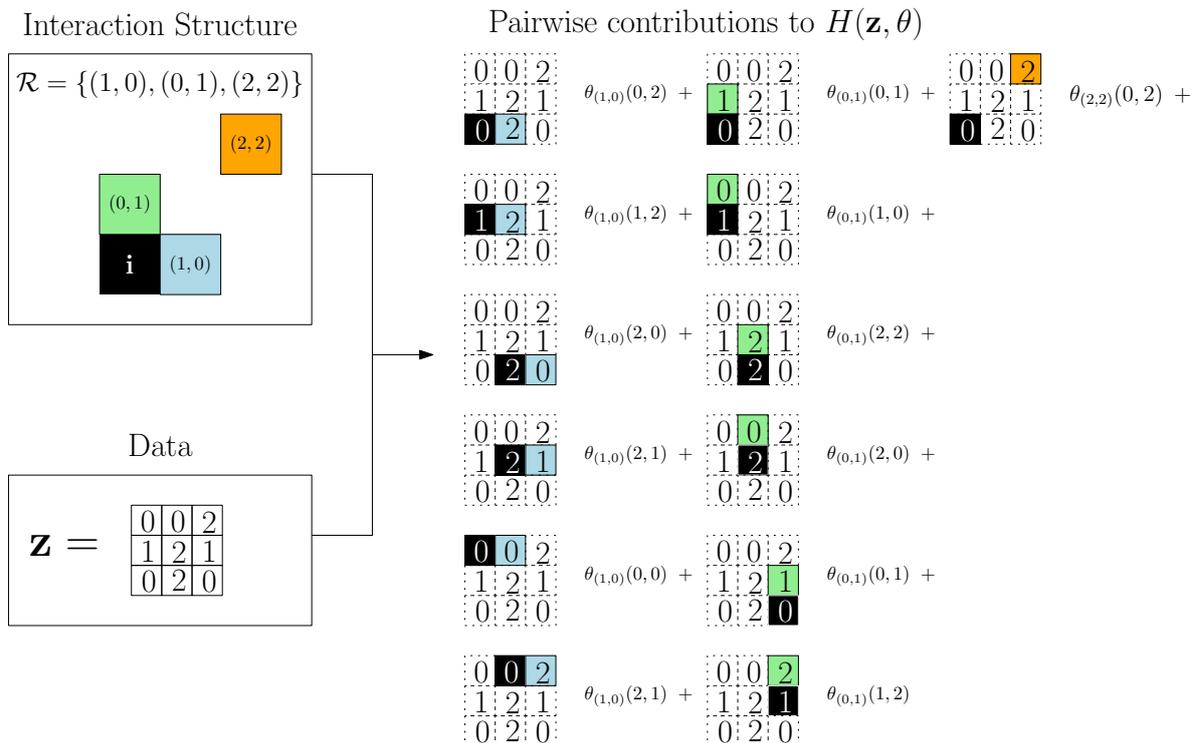


Figure 1: Example of interaction structure with three relative positions and example field on a 3 by 3 lattice (left) and contributions of each interacting pair to $H(\mathbf{z}, \theta)$ (right).

Note that adding a constant $c_{\mathbf{r}}$ to the potentials associated with a relative position $\mathbf{r} \in \mathcal{R}$ results in the same probability because the constant cancels when dividing by ζ_{θ} . Thus, constraints for the potentials $\theta_{\mathbf{r}}(a, b)$ are necessary to obtain identifiability in the model. We consider $\theta_{\mathbf{r}}(0, 0) = 0$ for all relative positions \mathbf{r} , which ensures identifiability and also gives an interpretation for interactions in terms of the pair $(0, 0)$: $\theta_{\mathbf{r}}(a, b) < 0$ (resp. > 0) means that the pair (a, b) is *less* (resp. *more*) likely to appear in a pair with relative position \mathbf{r} than $(0, 0)$.

Potts model as a particular case

The Potts model (Potts 1952) is one of the most important MRF model used in image segmentation because it can assign higher probability for equal-valued pairs of nearest-neighbors, creating large regions of pixels with the same values. The model has a single parameter ϕ that is interpreted as the inverse temperature in a mechanical statistics context.

A standard Potts model can be expressed as Equation 2 with the function $H(\mathbf{z}, \theta)$ taking the form

$$\sum_{(\mathbf{i}, \mathbf{j}) : \|\mathbf{i} - \mathbf{j}\| = 1} \phi \mathbb{1}_{(z_{\mathbf{i}} = z_{\mathbf{j}})}. \quad (4)$$

Assumptions (a), (b) and (c) are satisfied, thus, we can rewrite Equation 4 in terms of an interaction structure \mathcal{R} and potentials θ by noticing

- The set $\mathbf{i}, \mathbf{j} : \|\mathbf{i} - \mathbf{j}\| = 1$ are vertical and horizontal pairs of neighbors, therefore, the interaction structure \mathcal{R} is the set $\{(1, 0), (0, 1)\}$.

- The potential $\theta_{\mathbf{r}}(a, b)$ is equal to ϕ if $a \neq b$ and 0 otherwise, regardless of \mathbf{r} . The constraint $\theta_{\mathbf{r}}(0, 0) = 0$ is satisfied in this definition. Therefore, we have the parameter restriction

$$\theta_{\mathbf{r}}(a, b) = \phi \mathbb{1}_{(a \neq b)}$$

for all $\mathbf{r} \in \mathcal{R}$.

This parameter restriction corresponds to the "onepar" family described in Section 3.2.

2.2. Important elements of the model

The main inference challenge for MRFs lies in the normalizing constant ζ_{θ} appearing in Equation 2. It cannot be evaluated in practice as it requires summing over $\mathcal{Z}^{|\mathcal{L}|}$ possible field configurations and there is no analytical expression for it, except for trivial cases, leading to an intractable likelihood.

Being unable to evaluate the likelihood function hinders the use of most statistical methods. Inference under intractable likelihoods have been developed over the years. The main studies involve using conditional probability-based functions, like pseudo-likelihood (Jensen and Künsch 1994, for example) and Monte Carlo methods (Geyer and Thompson 1992; Møller, Pettitt, Reeves, and Berthelsen 2006, for example).

Although there is a wide variety of inferential methods available, most of them are built using the same pieces of the model. Thus, having access to each of these pieces is necessary to implement algorithms. We highlight important characteristics of the model available in **mrf2d** that are used by inference methods.

Conditional probabilities

A consequence of the Markov property (conditional independence) is a simple expression for conditional probabilities. $H(\mathbf{z}, \theta)$ is a sum of terms that only depends on pairs of pixel values, which implies that all terms not involving position \mathbf{i} cancel out when evaluating $\mathbb{P}(Z_{\mathbf{i}} | \mathbf{Z}_{-\mathbf{i}})$. Define the part of the sum that involves the pixel in position \mathbf{i} as

$$h_{\mathbf{i}}(k | \mathbf{z}) = \sum_{\mathbf{r} \in \mathcal{R}}^{\mathbf{(i+r)} \in \mathcal{L}} \theta_{\mathbf{r}}(k, z_{\mathbf{(i+r)}}) + \sum_{\mathbf{r} \in \mathcal{R}}^{\mathbf{(i-r)} \in \mathcal{L}} \theta_{\mathbf{r}}(z_{\mathbf{(i-r)}}, k). \quad (5)$$

The conditional probability of $Z_{\mathbf{i}} = k$ given all other locations, $\mathbb{P}(Z_{\mathbf{i}} = k | \mathbf{Z}_{\mathcal{N}_{\mathbf{i}}})$, is then given by the standard softmax of $h_{\mathbf{i}}(k | \mathbf{z})$,

$$\mathbb{P}(Z_{\mathbf{i}} = k | \mathbf{Z}_{-\mathbf{i}} = \mathbf{z}_{\mathcal{N}_{\mathbf{i}}}) = \frac{e^{h_{\mathbf{i}}(k|\mathbf{z})}}{\sum_{k'} e^{h_{\mathbf{i}}(k'|\mathbf{z})}}. \quad (6)$$

Pseudo-likelihood function

The pseudo-likelihood function (Besag 1974, 1975) is defined as the product of conditional probabilities of each variable given all other variables of a random field,

$$PL(\theta; \mathbf{z}) = \prod_{\mathbf{i} \in \mathcal{L}} \mathbb{P}(Z_{\mathbf{i}} = z_{\mathbf{i}} | \mathbf{Z}_{-\mathbf{i}} = \mathbf{z}_{-\mathbf{i}}) = \prod_{\mathbf{i} \in \mathcal{L}} \frac{e^{h_{\mathbf{i}}(z_{\mathbf{i}}|\mathbf{z})}}{\sum_{k'} e^{h_{\mathbf{i}}(k'|\mathbf{z})}}. \quad (7)$$

Algorithm 1: Approximate sampling algorithm for MRFs using T steps of Gibbs sampler.

Initialize \mathbf{z} with a starting configuration $\mathbf{z} = \mathbf{z}^{(0)}$;
Initialize the iteration counter $t = 0$; **while** $t \leq T$ **do**
 Sample $\{\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \dots, \mathbf{i}^{(|\mathcal{L}|)}\}$ a random permutation of the pixel positions \mathcal{L} ;
 for ℓ in $1, \dots, |\mathcal{L}|$ **do**
 Update $z_{\mathbf{i}^{(\ell)}}$ conditional to the rest of the field $\mathbf{z}_{-\mathbf{i}^{(\ell)}}$ with probabilities from
 Equation 6;
 end
 $t = t + 1$;
 Result: output the final configuration \mathbf{z} .
end

In the special case of an independent field, it is equivalent to the likelihood function. Notice that the pseudo-likelihood function does not depend on the intractable normalizing constant and Equation 7 is numerically equivalent to a logistic regression problem where each pixel values corresponds to independent observations and the interacting pixel values are covariates with coefficients corresponding to the associated potentials.

Generating MRFs via Gibbs sampler

While exact sampling from dependent and high-dimensional processes is a challenging task overall, the conditional independence of MRFs simplifies the implementation of the Gibbs sampler algorithm (Geman and Geman 1984). In the Gibbs sampler algorithm, each pixel value is updated conditionally to the current state of its neighbors and a Gibbs sampler *cycle* consists of updating each pixel exactly one time.

To avoid introducing any kind of bias due to updates order, a random permutation of \mathcal{L} is drawn to define the order in which pixels are updated at each cycle. After running a suitable number of cycles in Algorithm 1, the distribution of the resulting field sampled in the process is approximately the joint distribution of the MRF.

Sampling a field conditional to a subset of pixel values can be achieved with the same algorithm by skipping the updates for those pixels which are being conditioned on.

There exist faster mixing algorithms for particular cases such as Swendsen-Wang algorithm (Wang and Swendsen 1990), but they require specific conditions from the model and/or particular implementations to be efficient. Therefore, despite its slower mixing times in some scenarios, we keep the Gibbs sampler as the method of choice in this work due to its generalization ability as it only requires computing conditional distributions.

Sufficient statistics

An important computational consequence of the model assumptions is the fact that, in order to evaluate the probability (or likelihood) function for a particular observed field \mathbf{z} , it is not necessary to determine the values of each pixel individually, but only the co-occurrence counts for each relative position $\mathbf{r} \in \mathcal{R}$.

The function $H(\mathbf{z}, \theta)$ can be rewritten as

$$H(\mathbf{z}, \theta) = \sum_{\mathbf{r} \in \mathcal{R}} \sum_{a=0}^C \sum_{b=0}^C \theta_{\mathbf{r}}(a, b) n_{a,b,\mathbf{r}}(\mathbf{z}), \quad (8)$$

where $n_{a,b,\mathbf{r}}(\mathbf{z}) = \sum_{\mathbf{i} \in \mathcal{L}} \mathbb{1}_{(z_{\mathbf{i}}=a, z_{\mathbf{i}+\mathbf{r}}=b)}$ is the count of occurrences of the pair $(a, b) \in \mathcal{Z}^2$ in pairs of pixels with relative position \mathbf{r} . Therefore,

$$S_{\mathcal{R}}(\mathbf{z}) = \{n_{a,b,\mathbf{r}}(\mathbf{z}), a, b \in \mathcal{Z}, \mathbf{r} \in \mathcal{R}\}$$

is a vector of sufficient statistics, where each component $n_{a,b,\mathbf{r}}(\mathbf{z})$ is associated with a corresponding potential $\theta_{\mathbf{r}}(a, b)$. [Gimel'farb \(1996\)](#) calls this sufficient statistic the *co-occurrence histogram*.

Parameter constraints reduce the dimension of the sufficient statistic. Our identifiability constraint $\theta_{\mathbf{r}}(0, 0) = 0$ implies that all $n_{0,0,\mathbf{r}}(\mathbf{z})$ are excluded from $S_{\mathcal{R}}(\mathbf{z})$ and equality constraints require aggregating (sum) co-occurrence counts to match the parameter dimension. We shall keep the same notation for the constrained version of the sufficient statistics $S_{\mathcal{N}}(\mathbf{z})$ and potentials θ .

The main advantages of the representation with sufficient statistics are the reduced memory usage in Monte Carlo methods and a convenient representation of $H(\mathbf{z}, \theta)$ as an inner product that simplifies dealing with likelihood ratios as it is done in [Geyer and Thompson \(1992\)](#),

$$H(\mathbf{z}, \theta) = \langle S_{\mathcal{N}}(\mathbf{z}), \theta \rangle. \quad (9)$$

2.3. Gaussian mixtures driven by hidden MRFs

Another class of models present in the image processing field are hidden Markov random field models (HMRFs). The hidden version considers a latent (unobserved) process, denoted \mathbf{Z} and an observed field, denoted \mathbf{Y} , where \mathbf{Z} is distributed as a MRF and the distribution of $\mathbf{Y} \mid \mathbf{Z}$ is reasonably simple.

In this type of modeling, \mathbf{Z} is often considered the “true” image and \mathbf{Y} is a noisy image. Usually the goal of the analysis in this context is to recover the underlying field. Note that for the models considered in this work, where \mathbf{Z} has finite support, the hidden field defines a segmentation of the image, making it a suitable approach for image segmentation.

In **mrf2d**, we provide built-in tools for the case where $\mathbf{Y} \mid \mathbf{Z}$ is a finite Gaussian mixture with mixture components driven by the hidden field. Additional covariates can also be included as fixed effects for the mean,

$$Y_{\mathbf{i}} \mid Z_{\mathbf{i}} = a \sim N(\mu_a + \mathbf{x}_{\mathbf{i}} \top \boldsymbol{\beta}, \sigma_a^2), \quad a = 0, 1, \dots, C. \quad (10)$$

Given the latent field, observed values \mathbf{Y} are assumed to be independent leading to the conditional density

$$f(\mathbf{y} \mid \mathbf{Z} = \mathbf{z}) = \prod_{\mathbf{i} \in \mathcal{L}} \frac{1}{\sqrt{2\pi\sigma_{z_{\mathbf{i}}}^2}} \exp\left(-\frac{(y_{\mathbf{i}} - \mu_{z_{\mathbf{i}}} - \mathbf{x}_{\mathbf{i}} \top \boldsymbol{\beta})^2}{2\sigma_{z_{\mathbf{i}}}^2}\right) \quad (11)$$

and the complete likelihood function

$$L_{\theta}(\beta, \{(\mu_a, \sigma_a), a = 0, \dots, C\}; \mathbf{y}, \mathbf{z}) = \frac{1}{\zeta_{\theta}} e^{H(\mathbf{z}, \theta)} \prod_{\mathbf{i} \in \mathcal{L}} \frac{1}{\sqrt{2\pi\sigma_{z_i}^2}} \exp\left(\frac{(y_{\mathbf{i}} - \mu_{z_i} - \mathbf{x}_{\mathbf{i}}^{\top} \beta)}{2\sigma_{z_i}^2}\right) \quad (12)$$

Inference for this models involves estimating the parameters $(\mu_k, \sigma_k)_{k=0,1,\dots,C}$ and β associated with the Gaussian Mixture and predicting the labels of the latent field \mathbf{z} simultaneously. Bayesian methods and the EM algorithm are the most common approaches. The parameters of the latent field distribution θ are fixed a priori and considered tuning hyper-parameters of the algorithm.

3. Using the package

3.1. Model representation

The model described in Section 2 can be completely characterized by three components: the random field \mathbf{z} , the interaction structure \mathcal{R} and the potentials θ . Additionally, \mathbf{y} and the mixture parameters $(\mu_k, \sigma_k^2)_{k=0,\dots,C}$ are included in hidden MRFs.

A consistent representation of each component is provided in the package so that inputs and outputs of built-in functions, as well as methods the user may implement, are compatible and usable in the analysis pipeline. Representations are described in Table 1.

Random fields \mathbf{z} and \mathbf{y} .

Realizations of a random fields \mathbf{z} and \mathbf{y} are represented by simple `matrix` objects with dimension $N \times M$, where $N \geq \max_{i_1}(i_1, i_2) \in \mathcal{L}$ and $M \geq \max_{i_2}(i_1, i_2) \in \mathcal{L}$, i.e., the maximal

Model component	Function argument	Representation in <code>mrf2d</code>
<code>z</code> : Discrete-valued field	<code>Z</code>	A <code>matrix</code> object with values in $\{0, \dots, C\}$, where <code>Z[w,q]</code> represents the pixel value in position (w, q) of the lattice. <code>NA</code> values are used for positions that do not belong to \mathcal{L} when it is not a rectangular region.
<code>y</code> : Continuous-valued field	<code>Y</code>	A <code>matrix</code> object with real values, where <code>Y[u,v]</code> represents the pixel value in position (u, v) of the lattice. <code>NA</code> values are used for positions that do not belong to \mathcal{L} when it is not a rectangular region.
<code>R</code> : Interaction structure	<code>mrfi</code>	An object of the S4 class ‘ <code>mrfi</code> ’. It can be created with the <code>mrfi()</code> and <code>rpositions()</code> functions.
$\theta_r(a, b)$: Array of potentials	<code>theta</code>	A three-dimensional <code>array</code> object with dimensions $(C + 1) \times (C + 1) \times \mathcal{R} $. For a pair of values (a, b) and the s -th interacting relative position \mathbf{r}_s of \mathcal{R} , the corresponding potential is mapped at <code>theta[a+1, b+1, s]</code> .

Table 1: Model representation summary.

coordinates. This matrix represents a rectangular set of pixels that contains \mathcal{L} . The value in row i_1 and column i_2 represents the observed value of the random field in position (i_1, i_2) : an integer in $\{0, 1, \dots, C\}$ for \mathbf{z} or a real number for \mathbf{y} .

We do not require \mathcal{L} to be a complete rectangular region. Pixels which position does not belong to \mathcal{L} are assigned the NA value.

Two functions are available for visualizing random fields: `dplot()` and `cplot()`. `dplot()` should be used for discrete-valued fields \mathbf{z} while `cplot()` is used for continuous-valued matrices \mathbf{y} . These functions provide an alternative to base R `image()` function, producing elegant images in the form of `ggplot` objects. The main advantage is that they allow the use of the `ggplot2` package (Wickham 2016) to customize the image using the grammar of graphics. Details and examples of customization of the images produced using `ggplot2` can be found in Appendix A.

Interaction structures \mathcal{R}

Interaction structures are represented by objects of the S4 class ‘`mrfi`’ implemented in `mrfd`. These objects can be created with the `mrfi()` or `rpositions()` functions, which have arguments `max_norm`, `norm_type` and `positions`. In `mrfi()`, the interaction structure created will include all relative positions which satisfy $\|(i_1, i_2)\| \leq \text{max_norm}$ for the specified norm type. `positions` can be passed as a list containing length 2 integer vectors with relative positions to include when using `rpositions()`. The function automatically checks for repeated and opposite relative positions to ensure the structure is valid.

`norm_type` options are the same as R built-in `norm()` function, mainly, "1", "2" and "m" are used for ℓ_1 , ℓ_2 and the maximum norm, respectively. The default is ℓ_1 norm.

An algebra of `mrfi` objects is implemented for manipulating these objects. `+` is used to perform union of two `mrfi` objects or a `mrfi` object and a `numeric` vector with 2 integers can be used to add a single interacting position to an existing `mrfi` object. Similarly, the `-` operator can be used to perform set difference between two `mrfi` objects or to remove a single position if a vector with 2 integers is used in the right-hand-side.

Some examples for creating different \mathcal{R} are detailed below.

- `mrfi(max_norm = 1)` creates an interaction structure with all positions with $\|(i_1, i_2)\|_1 \leq 1$, which corresponds to a nearest-neighbor structure $\mathcal{R} = \{(1, 0), (0, 1)\}$.
- `rpositions(positions = list(c(1,0), c(0,1)))` is an alternative way of specifying the same structure of the previous example.
- `mrfi(max_norm = 1) + rpositions(positions = list(c(2,0)))` results in the interaction structure $\mathcal{R} = \{(1, 0), (0, 1), (2, 0)\}$.
- `mrfi(max_norm = 1) + rpositions(positions = list(c(-1,0)))` results in $\mathcal{R} = \{(0, 1), (-1, 0)\}$. The norm-based and position-based positions had an intersection (reflected position at $(-1, 0)$), so the redundant position $(1, 0)$ was removed. In case of opposite directions being added together, right-hand size argument is prioritized.

Additionally, conversion of `mrfi` objects to `list` is implemented in the `as.list()` method. Subsetting methods are also available with the `"["` and `"[["` operators. These methods

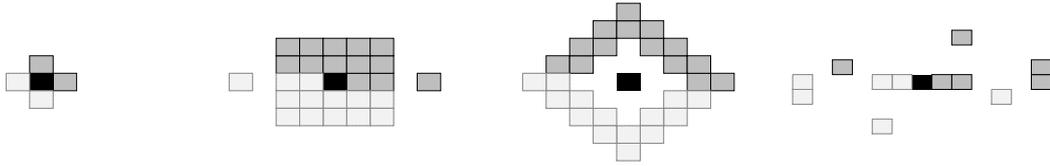


Figure 2: Examples of interaction structures \mathcal{R} created and their visualization.

are particularly important for model selection algorithms, as many distinct sparse interaction structures can be obtained by using different subsets of a large reference base structure.

A `plot` method is available for `mrfi` objects. The code chunk below exemplifies the usage of plotting functions and manipulation of `mrfi` objects. The resulting plots are presented in Figure 2. The black square represents the origin position $(0, 0)$, positions included in the interaction structure \mathcal{R} are represented by the dark-gray squares with black borders, while their opposite directions are the light-gray squares.

```
R> plot(mrfi(max_norm = 1))
R> plot(mrfi(max_norm = 2, norm_type = "m") + c(4, 0))
R> plot(mrfi(4) - mrfi(2))
R> plot(mrfi(6, norm_type = "m")[c(1, 2, 6, 9, 19, 41)])
```

Potentials array θ

The collection of potentials, $\theta_{\mathbf{r}}(a, b)$, is represented by an `array` object with dimensions $(C + 1) \times (C + 1) \times |\mathcal{R}|$. Rows and columns are used to map a and b , respectively, while slices are used to map relative positions \mathbf{r} . A set of potentials $\{\theta_{\mathbf{r}}(a, b), a, b \in \mathcal{Z}, \mathbf{r} \in \mathcal{R}\}$ is always related to an interaction structure $\mathcal{R} = \{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_{|\mathcal{R}|}\}$. Since the i -th slice maps the i -th relative position of \mathcal{R} , \mathbf{r}_i , this is the minimal representation required to store all parameters required by the model.

An important detail is that array indices in R start at 1, while we consider our set of possible values $\mathcal{Z} = \{0, 1, \dots, C\}$, therefore we need to shift a and b by one position when accessing their value in the R array. Figure 3 illustrates how potentials can be represented as an array in R in the $C = 2$ case. Two elements are highlighted and the associated indices used to access them are shown as examples.

3.2. Parameter restriction families

Parameter restrictions play an important role in the inference process of our Markov random field models. `mrfd` functions support 5 families of parameter restrictions for the array of potentials to be considered in inference algorithms. They are specified by the `family` argument of functions to ensure the resulting output array (`theta`) respects those constraints. A brief description of each interaction structure is given next. Table 2 presents the mathematical definitions, number of free parameters and an example of a slice of the array of potentials for the case with $C = 2$ in each family.

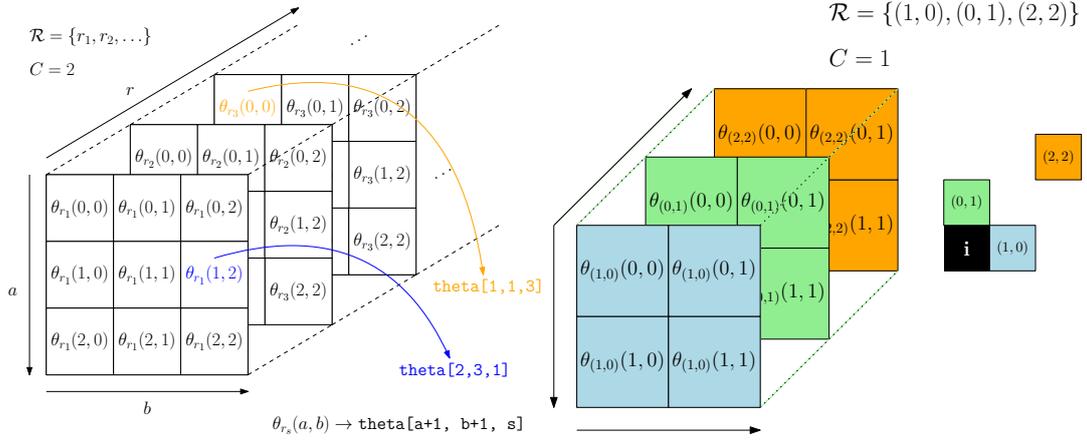


Figure 3: Left: Example of array representation of potentials with $C = 2$ for a generic interaction structure \mathcal{R} . Right: Array representation of potentials for the example from Figure 1.

Family	Restriction	Free parameters	Example slice
"onepar"	$\theta_{\mathbf{r}}(a, b) = \phi \mathbb{1}_{(a \neq b)}$	1	$\begin{bmatrix} 0 & \phi & \phi \\ \phi & 0 & \phi \\ \phi & \phi & 0 \end{bmatrix}$
"oneeach"	$\theta_{\mathbf{r}}(a, b) = \phi_{\mathbf{r}} \mathbb{1}_{(a \neq b)}$	$ \mathcal{R} $	$\begin{bmatrix} 0 & \phi_{\mathbf{r}} & \phi_{\mathbf{r}} \\ \phi_{\mathbf{r}} & 0 & \phi_{\mathbf{r}} \\ \phi_{\mathbf{r}} & \phi_{\mathbf{r}} & 0 \end{bmatrix}$
"absdif"	$\theta_{\mathbf{r}}(a, b) = \sum_{d=1}^C \phi_{\mathbf{r},d} \mathbb{1}_{(b-a =d)}$	$ \mathcal{R} C$	$\begin{bmatrix} 0 & \phi_{\mathbf{r},1} & \phi_{\mathbf{r},2} \\ \phi_{\mathbf{r},1} & 0 & \phi_{\mathbf{r},1} \\ \phi_{\mathbf{r},2} & \phi_{\mathbf{r},1} & 0 \end{bmatrix}$
"dif"	$\theta_{\mathbf{r}}(a, b) = \sum_{d=-C, d \neq 0}^C \phi_{\mathbf{r},d} \mathbb{1}_{(b-a=d)}$	$ \mathcal{R} 2C$	$\begin{bmatrix} 0 & \phi_{\mathbf{r},1} & \phi_{\mathbf{r},2} \\ \phi_{\mathbf{r},-1} & 0 & \phi_{\mathbf{r},1} \\ \phi_{\mathbf{r},-2} & \phi_{\mathbf{r},-1} & 0 \end{bmatrix}$
"free"	$\theta_{\mathbf{r}}(0, 0) = 0$	$ \mathcal{R} (C^2 - 1)$	$\begin{bmatrix} 0 & \phi_{\mathbf{r},0,1} & \phi_{\mathbf{r},0,2} \\ \phi_{\mathbf{r},1,0} & \phi_{\mathbf{r},1,1} & \phi_{\mathbf{r},1,2} \\ \phi_{\mathbf{r},2,0} & \phi_{\mathbf{r},2,1} & \phi_{\mathbf{r},2,2} \end{bmatrix}$

Table 2: Description of parameter restriction families.

"onepar" A single-parameter (ϕ) model, where interactions depend only on the fact that values are equal or different, regardless of their relative position. This restriction corresponds to the classical Ising and Potts model.

"oneeach" The same interaction type as "onepar", but allowing different values $\phi_{\mathbf{r}}$ for different interacting positions $\mathbf{r} \in \mathcal{R}$.

"absdif" For each $\mathbf{r} \in \mathcal{R}$, the potentials $\theta_{\mathbf{r}}(a, b)$ depend only on the absolute differences of their pixel values $d = |b - a|$. Note that "absdif" is equivalent to "oneeach" when $C = 1$.

"dif" Generalizes the "absdif" family allowing opposite signal differences to have different interactions.

"free" No restrictions, except for the identifiability constraint $\theta_{\mathbf{r}}(0, 0) = 0$.

Families "dif" and "absdif" should only be used when pixel values represent quantities and their differences are well-defined, for instance, in grayscale images with few levels, as in the example analysed in Gimel'farb (1996). If relabeling the values does not change the interpretation of the problem, then these restrictions are probably not suitable.

The function `smr_array(theta, family)` can be used to transform a parameter array into a vector of appropriate length containing only the free parameters corresponding to the provided array (`theta`) and the restriction `family`. The opposite operation is also available as the `expand_array(theta_vec, family, mrfi, C)` function. These transformations use a simpler and less memory consuming structure so they are particularly useful for optimization problems as most functions, for example R built-in `optim` function, require a vector of parameters. Also, they are convenient for storing multiple vectors, for example in Monte Carlo methods.

3.3. Random field sampler

Being able to sample observations of Markov random fields is a key component of many inference methods that aim to avoid the intractable normalizing constant. In **mrfd**, a complete and efficient routine to sample fields using the Gibbs sampler algorithm described in Algorithm 1 is provided by the `rmrf2d()` function. Its arguments are:

- `init_Z`: The initial field configuration, or a length-2 vector with the dimensions of the field to be sampled. If the dimensions are provided, the initial configuration is randomly sampled from independent discrete uniform distributions.
- `mrfi`: A `mrfi` object representing the interaction structure \mathcal{R} .
- `theta`: An array of potentials.
- `cycles`: The number of Gibbs sampler cycles.
- `sub_region`: Optional argument used for non-rectangular images when `init_Z` is a vector holding the dimensions. A logical matrix with the same dimensions as specified in `init_Z`. Pixels with `FALSE` value are not included in the image.
- `fixed_region`: Optional. A matrix with logical values. Pixel positions with `TRUE` value are conditioned on their initial configuration (`init_Z`) value and are not updated.

We illustrate below the use of the sampling function on two fields: a 200×200 field (`z_sample`) without conditioning on any pixel (nothing specified in `fixed_region`) and a 100×100 field (`z_border`) conditioning on the boundary values being fixed as 0, sampled from a random initial configuration. The resulting images are presented in Figure 4.

```
R> th <- expand_array(-1, family = "onepar", mrfi(1), C = 1)
R> z_sample <- rmrf2d(init_Z = c(200,200), mrfi = mrfi(1), theta = th)
```

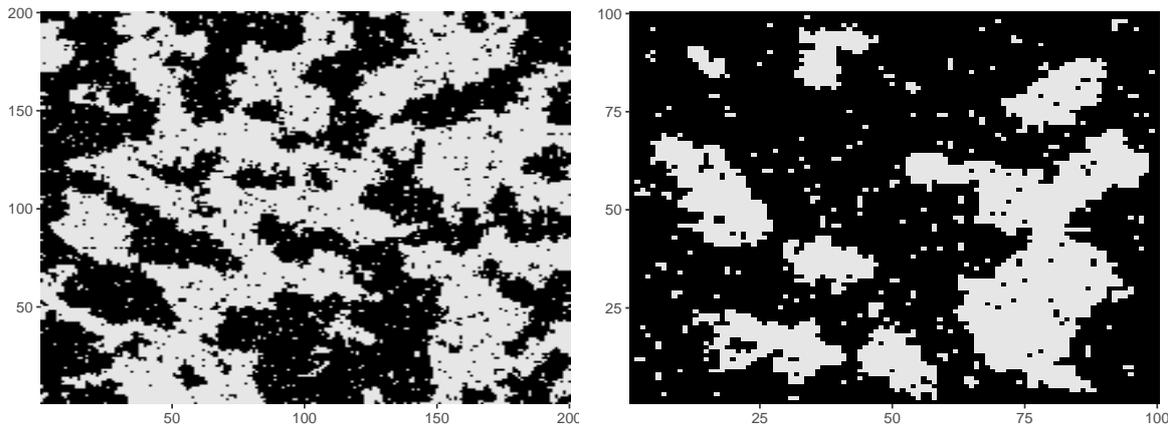


Figure 4: Simulated random fields from a nearest-neighbor structure. Left: no boundary conditions. Right: conditional to all border values being 0 (black).

```
R> border <- matrix(FALSE, nrow = 100, ncol = 100)
R> border[1, ] <- border[100, ] <- border[, 1] <- border[, 100] <- TRUE
R> initial <- matrix(sample(0:1, 100*100, replace = TRUE),
+   nrow = 100, ncol = 100)
R> initial[border] <- 0
R> z_border <- rmrf2d(initial, mrfi = mrfi(1), theta = th,
+   fixed_region = border)
```

Non-rectangular fields can be sampled either by passing a non-rectangular field as the `init_Z` argument or by using the `sub_region` argument and specifying the dimensions of the sampled field.

Another important feature is conditioning on a subset of pixel values. There are many situations where keeping a subset of pixels fixed during the sampling process can be useful, for example, filling a region of missing pixel values via simulation, defining boundary conditions (our model corresponds to a free boundary condition, but other types such as fixed or periodic boundary can be sampled with proper manipulation of the initial configuration and conditioning region) or performing block-wise updates of the data using conditionally independent blocks (for parallelization of algorithms).

Since the Gibbs sampler algorithm updates each pixel value multiple times, performance is one of our main implementation concerns. To improve the performance and speed up computations considerably, the internals of the sampling function, as well as most other computationally intensive functions are written in C++ with the use of **Rcpp** (Eddelbuettel and François 2011) and **RcppArmadillo** (Eddelbuettel and Sanderson 2014) packages.

3.4. Statistical inference in `mrf2d`

Inference methods for MRF models are diverse and their suitability highly depend on the type of data being analyzed. The framework provided by `mrf2d` can be used to implement all sorts of algorithms that are built from a common stack of components: simulation, conditional probabilities and sufficient statistics. It also provides complete built-in routines for some estimation algorithms.

Function	Use
<i>Miscellaneous</i>	
<code>rmrf2d</code>	Generates samples of a MRF via Gibbs sampler. Used for Monte Carlo based methods.
<code>cp_mrf2d</code>	Computes the conditional probabilities for a pixel position given its neighbors.
<code>pl_mrf2d</code>	Computes pseudo-likelihood value for an observed field considering interaction structure <code>mrfi</code> and array of potentials <code>theta</code> .
<code>cohist</code>	Creates the co-occurrence histogram of an observed field given an interaction structure. Can be converted to a vector of sufficient statistics given a restriction family with the <code>smr_stat</code> function.
<code>smr_array</code> and <code>expand_array</code>	Conversions between array and vector representation of potentials given a parameter restriction family.
<i>Built-in inference algorithms</i>	
<code>fit_pl</code>	Estimates the parameter array given an observed field via pseudo-likelihood optimization. Returns a <code>mrfout</code> object.
<code>fit_sa</code>	Estimates the parameter array given an observed field via stochastic approximation algorithm. Returns a <code>mrfout</code> object.
<code>fit_ghm</code>	Fits a Gaussian mixture driven by a given hidden MRF model using the EM algorithm from Zhang <i>et al.</i> (2001) . <code>polynomial_2d</code> and <code>fourier_2d</code> can be used to create polynomial and 2-dimensional Fourier basis functions, respectively, to be used as a fixed effect. Returns a <code>hmrfout</code> object.

Table 3: List of available functions used for inference in `mrfd` with a brief description of each one.

Table 3.4 presents a list of functions available in the package that can be used to construct inference algorithms, as well as built-in functions for parameter estimation for MRF and for the hidden MRF models defined in Section 2.3 that we describe next. The built-in inference functions return objects of class ‘`mrfout`’ (MRF data) or ‘`hmrfout`’ (hidden MRF models), which contains the information about the fitted model, as well as `summary` and `plot` methods associated for interpretation of the results.

Maximum pseudo-likelihood estimation

The pseudo-likelihood function in Equation 7 can be evaluated efficiently because it does not depend on the intractable normalizing constant. A common estimation procedure for intractable likelihood problems is optimizing the pseudo-likelihood with respect to the parameters. The pseudo-likelihood estimator is given by

$$\hat{\theta}_{PL} = \arg \max_{\theta} PL(\theta; \mathbf{z}). \quad (13)$$

The function `fit_pl()` from `mrfd` implements an optimization of the pseudo-likelihood function using R built-in `optim()` function. It handles the conversions between array and vector representation of potentials automatically, respecting the restriction family selected

and returns the estimated array of potentials and maximum value of the pseudo-likelihood in logarithmic scale. The arguments of `fit_pl` are:

- **Z**: The observed random field \mathbf{z} .
- **mrfi**: A `mrfi` representing an interaction structure \mathcal{R} .
- **family**: A parameter restriction family.
- **init**: An array with the initial configuration used in the optimization. 0 can be used to start from the independent model.
- **optim_args**: A named list with additional arguments passed to the `optim()` function call.

Stochastic approximation algorithm

Given an observed field $\mathbf{z}^{(0)}$, the stochastic approximation algorithm (Robbins and Monro 1951) seeks to create a Markov chain of parameter vectors $\{\boldsymbol{\theta}^{(t)}\}_{t \geq 1}$ that converges to the maximum likelihood estimate of $\boldsymbol{\theta}$, which is the solution of the zero gradient condition $\mathbb{E}_{\boldsymbol{\theta}}(S_{\mathcal{R}}(\mathbf{Z})) = S_{\mathcal{R}}(\mathbf{z}^{(0)})$, derived from Equation 9.

The algorithm is defined by the recurrence

$$\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} + \gamma^{(t)}(S_{\mathcal{R}}(\mathbf{z}^{(0)}) - S_{\mathcal{R}}(\mathbf{z}^{(t)})), \quad (14)$$

where $\mathbf{z}^{(t)}$ is a field sampled using $\boldsymbol{\theta}^{(t)}$ and $\gamma^{(t)}$ is a sequence of positive constants that satisfies $\sum_{t=1}^{\infty} \gamma^{(t)} = \infty$ and $\sum_{t=1}^{\infty} (\gamma^{(t)})^2 < \infty$.

Stochastic approximation is implemented in **mrfd** as the `fit_sa()` function. It samples $\mathbf{z}^{(t)}$ via Gibbs sampler considering the previous field $\mathbf{z}^{(t-1)}$ as the initial configuration. Periodically, the field samples are refreshed, starting from an independent discrete uniform distribution and running a greater number of Gibbs sampler cycles, this procedure prevents the algorithms from getting stuck in problematic field samples. Its arguments are:

- **Z**: The observed field $\mathbf{z}^{(0)}$.
- **mrfi**: The interaction structure \mathcal{R} .
- **family**: The family of parameter restrictions considered when converting the potentials array to a vector.
- **gamma_seq**: A sequence of step size values to be used as $\gamma^{(t)}$. These values are divided by the number of pixels $|\mathcal{L}|$ internally to be invariant with respect to the image size. The typical sequence recommended is `seq(from = M, to = 0, length.out = B)`, with M ranging from 0.5 to 2 and large number of iterations (B).
- **init**: The initial array of parameters or the value 0 to start from the independent model.
- **cycles**: Number of Gibbs sampler ran between iterations.

- **refresh_each**: Restarts the sample $\mathbf{z}^{(t)}$ from a random configuration each **refresh_each** iterations.
- **refresh_cycles**: When a refresh happens, how many Gibbs sampler cycles are ran in the current parameter configuration.

Among the elements of the returned **mrfout** object, **theta** contains the estimated potential array and **metrics** a data frame with the Euclidean distances between $S_{\mathcal{R}}(\mathbf{z}^{(0)})$ and $S_{\mathcal{R}}(\mathbf{z}^{(t)})$ for each iteration. This sequence of distances is used to monitor the convergence of the algorithm as a form of diagnostics analysis.

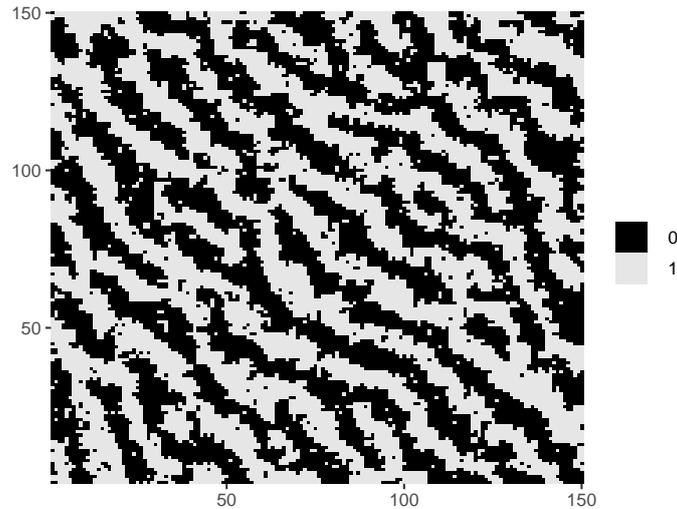
EM algorithm for HMRF models

Gaussian Mixtures driven by hidden MRFs can be fitted in **mrf2d** with an extension of the EM algorithm from Zhang *et al.* (2001) to include a fixed effect (see Freguglia *et al.* 2020, for details). The probabilities computed for the latent label of each pixel in the E-step are conditioned on the global maximum probability configuration of its neighbors, obtained via iterated conditional modes (ICM) algorithm at each iteration (Besag 1986).

The complete algorithm is available in the **fit_ghm** function. Its main arguments are

- **Y**: The observed continuous-valued field **y**.
- **mrfi**: Interaction structure of the latent field \mathcal{R} .
- **theta**: The array of potentials that defines the latent field distribution.
- **fixed_fn**: A list of functions of pixel positions $f(i_1, i_2)$ to be used as fixed effect. Constructors for 2-dimensional polynomials and Fourier basis are available in the functions **polynomial_2d** and **fourier_2d**, respectively.
- **equal_vars**: A logical value indicating if mixture components are forced to have equal variances.
- **init_mus** and **init_sigmas**: Optional initial values of $(\mu_a, \sigma_a)_{a=0, \dots, C}$. If none is passed, an independent Gaussian mixture is fitted with initial values based on quantiles and the estimates of this fitted model are used as (often good) starting values in the main procedure.
- **maxiter**: Maximum number of iterations before stopping.
- **max_dist**: Defines a stopping condition for the EM algorithm. For consecutive iterations t and $t+1$, the absolute difference in each parameter, $|\mu_k^{(t)} - \mu_k^{(t+1)}|$ and $|\sigma_k^{(t)} - \sigma_k^{(t+1)}|$ are computed for $k = 0, 1, \dots, C$. The algorithm stops if all differences are less than **max_dist**.
- **icm_cycles**: Number of cycles of Iterated Conditional Modes algorithm executed in each iteration.

fit_ghm() returns a **hmrfout** object represented by a **data.frame** containing **par** – the estimates of the mixture parameters $\{(\hat{\mu}_a, \hat{\sigma}_a), a = 0, 1, \dots, C\}$, **Z_pred** – the highest probability

Figure 5: Visualization of `field1`.

configuration of the latent field computed via ICM algorithm $\hat{\mathbf{z}}$, `fixed` – a matrix with the estimated fixed effects $(\mathbf{x}_i^\top \hat{\beta})$ for each pixel, and `predicted` – a matrix with the predicted mean for each pixel $(\mathbf{x}_i^\top \hat{\beta} + \hat{\mu}_{\hat{z}_i})$.

4. Data analysis using `mrf2d`

4.1. Example 1: A binary image with texture-like pattern

Description

To illustrate the usage of `mrf2d` for finite-valued images, we use the object `field1` available in the package, which contains a binary field with anisotropic pattern as seen in Figure 5. It is a synthetic texture image of the same type as the binary texture data presented in [Cross and Jain \(1983\)](#). The data can be loaded and viewed using the code chunk below.

```
R> data("field1", package = "mrf2d")
R> dplot(field1, legend = TRUE)
```

Our goal is to fit a MRF model to this data and sample images from the fitted model to evaluate if the patterns achieved in the generated data are similar to the original data. This is the typical setup of a texture synthesis problem with finite-valued images.

This analysis involves three main stages: Specifying the model (interaction structure and parameter restrictions), estimating the parameters and evaluating the fitted model. All of the run times described in this section were obtained using an Intel Core i7-7500U 2.70GHz CPU processor.

Specifying \mathcal{R} and parameter family

Model selection under intractability is a challenging problem because most algorithms require comparing (maximum) likelihood functions for different models, what cannot be done exactly and/or have a high computational cost for MRF models.

The main routes in the model specification stage are: using prior information of the data or problem to select what type of restrictions and interaction structure are best suited, using the most general model (e.g., no restrictions and a complex interaction structure) as in [Freguglia et al. \(2020\)](#) or using some estimation technique.

This image presents a diagonal pattern what indicates a nearest-neighbor interaction structure may not be appropriate to capture all the dependence present in the field. We first choose what kind of parameter restriction family will be considered by checking that relabeling the values \mathcal{Z} does not change the patterns in the image indicating symmetric potentials should be suited for this image, and there is a clear difference in the interactions when considering pixels in different directions, what indicates we need different types of interaction for different relative positions. These characteristics match the "oneeach" family that will be used in this example.

In order to estimate the set of interacting positions \mathcal{R} , we use a naive algorithm which consists of performing 300 steps of stochastic approximation considering a large set of candidate interacting positions (all positions with maximum norm less or equal 6, 84 positions total) and then select the positions with absolute value of the associated potential higher than a threshold value. This is a strategy similar to the heuristic search algorithm from [Gimel'farb \(1996\)](#). Stochastic approximation was preferred over maximum pseudo-likelihood, for example, because it is computationally more suited for high-dimensional situations and we are not requiring a very accurate estimation at this point, so we can use a reasonably low number of iterations.

The code below implements this naive interaction selection algorithm in a few lines using the tools available in **mrf2d** considering a threshold value of 0.10. A large set of interaction position (`candidates`) is defined and the stochastic approximation algorithm is executed based on this complete interaction structure, obtaining `complete_sa`, then a selection based on thresholding absolute values of interactions is performed (`selected`).

```
R> candidates <- mrfi(6, norm_type = "m")
R> set.seed(1)
R> complete_sa <- fit_sa(field1, candidates, family = "oneeach",
+   gamma_seq = seq(from = 1, to = 0, length.out = 300),
+   cycles = 2, refresh_each = 301)
```

The complete stochastic approximation procedure in the `fit_sa()` call took 168 seconds to complete in total.

```
R> plot(complete_sa)
R> thr_value <- 0.1
R> theta_vec <- smr_array(complete_sa$theta, "oneeach")
R> selected <- which(abs(theta_vec) > thr_value)
R> R1 <- candidates[selected]
R> R1
```

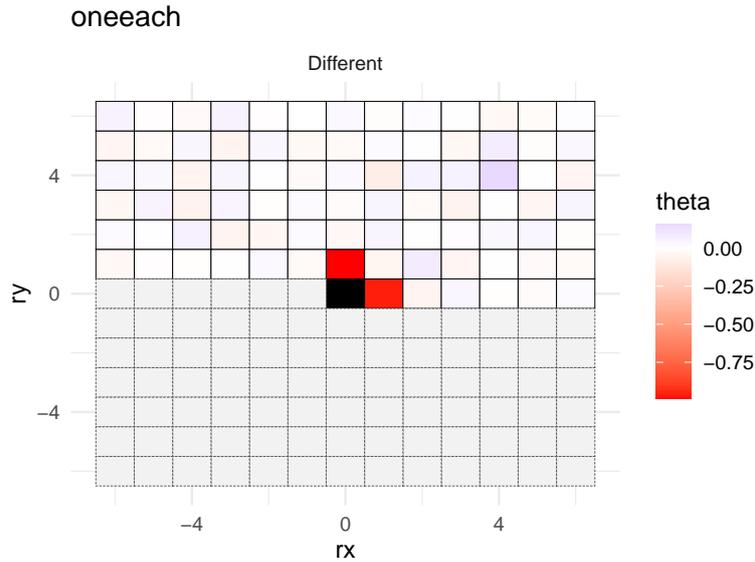


Figure 6: Plot of the `complete_sa` object. Nearest-neighbor positions have strong interactions (with negative potential) for different-valued pairs, while position (4, 4) has a weaker positive potential. This can be interpreted as nearest-neighbors having more weight when they are equal, while pixel with relative position (4, 4) have more weight when they are different.

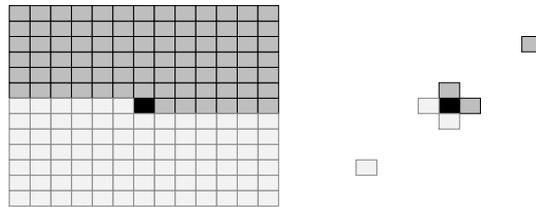


Figure 7: Candidate positions for the interaction structure (left) and selected positions (right).

3 interacting positions.

rx	ry
1	0
0	1
4	4

Estimating θ

Considering the parameter restriction family "`onepar`" and the selected interaction structure $\mathcal{R} = \{(1, 0), (0, 1), (4, 4)\}$, we have a model with 3 free parameters. A 3-dimensional optimization problem is simple enough to be solved using the built-in pseudo-likelihood optimization function. We also fit the model via stochastic approximation, now only considering the selected interaction structure, for comparison. The results are compared with the `summary()` method for the `'mrfout'` class and presented below.

```
R> pl <- fit_pl(field1, R1, family = "oneeach")
```

```
R> summary(pl)
```

```
Model adjusted via Pseudolikelihood
```

```
Image dimension: 150 150
```

```
2 colors, distributed as:
```

```
      0      1
11083 11417
```

```
Interactions for different-valued pairs:
```

```
Position| Value  Rel. Contribution
(1,0)| -0.993  1.000 ***
(0,1)| -1.021  0.995 ***
(4,4)|  0.183  0.735 **
```

```
R> sa <- fit_sa(field1, R1, family = "oneeach",
+   gamma_seq = seq(from = 1, to = 0, length.out = 300))
R> summary(sa)
```

```
Model adjusted via Stochastic Approximation
```

```
Image dimension: 150 150
```

```
2 colors, distributed as:
```

```
      0      1
11083 11417
```

```
Interactions for different-valued pairs:
```

```
Position| Value  Rel. Contribution
(1,0)| -0.964  1.000 ***
(0,1)| -0.983  0.987 ***
(4,4)|  0.183  0.756 ***
```

The resulting estimates were roughly the same using the two functions. The `fit_pl` call ran in 2.371 seconds while the `fit_sa` call took 56.627 seconds to complete. The optimization process from maximum pseudo-likelihood estimation was substantially faster, mainly due to the low-dimensionality of the problem, than running a satisfactory number of stochastic approximation steps to achieve reasonable precision.

Evaluating the fitted model

To evaluate how well the estimated parameters fit the data, we generate a new sample from the fitted model. Figure 8 shows the original image and the image simulated from the fitted model for comparison. The patterns created are visually very similar. Therefore the fitted MRF model successfully describes the characteristics of the data and is capable of synthesizing new images with the same texture pattern.

```
R> z_sim <- rmrf2d(dim(field1), R1, pl$theta)
```

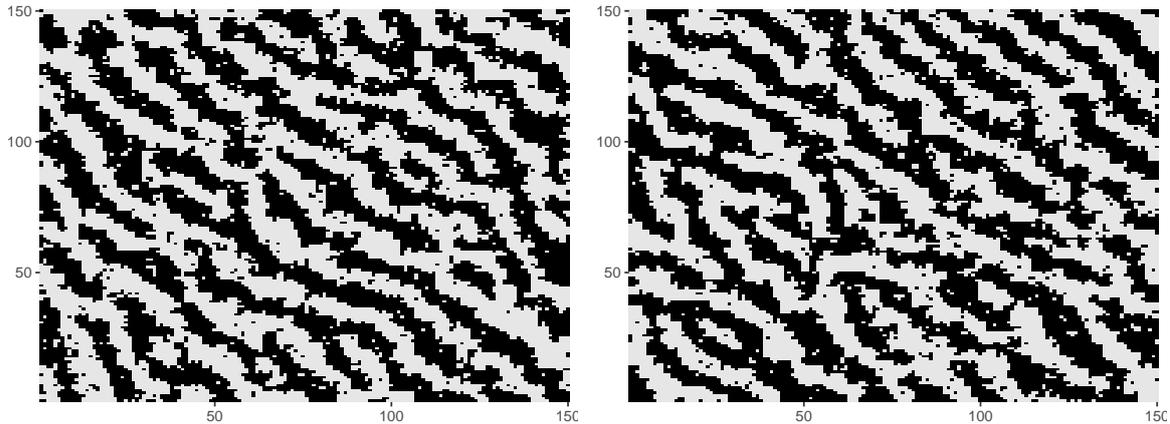


Figure 8: Original data (left) and random field simulated from the fitted model (right).

4.2. Example 2: Image segmentation of a hidden MRF

Description

The data available in the object `hfield1` in the package will be used to illustrate the use of `mrf2d` for Gaussian mixtures driven by hidden MRFs. It consists of an image with continuous-valued pixels ranging from 0.3 to 15.2. A pattern similar to the previous example can be observed with the addition of a continuous noise.

```
R> data("hfield1", package = "mrf2d")
R> cplot(hfield1)
```

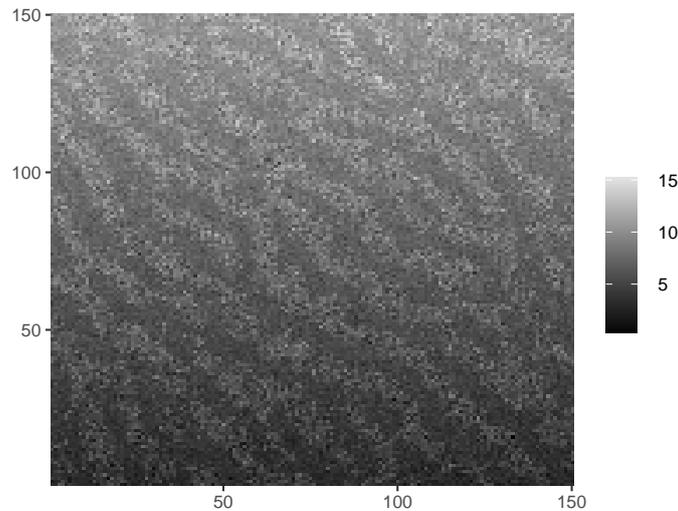
We consider an image composed by a latent (hidden) MRF plus a random noise whose distribution for each pixel may depend on the pixel value in the latent field. The main goal for this type of data is to recover the segmentation of the underlying pixel labels, an image segmentation problem (Li, Wu, and Zhang 2009; Shah and Chauhan 2015, for example). This is a typical problem where Gaussian mixtures driven by hidden Markov random fields are well suited.

Fitting a hidden MRF with no fixed effect

The built-in function for fitting hidden MRFs (`fit_ghm()`), like most algorithms used for Gaussian mixtures driven by HMRFs, considers the distribution of the underlying field as a hyper-parameter specified a priori. In this example, since we observe an underlying pattern similar to one in Example 1, we will reuse the model estimated by penalized likelihood as the MRF distribution.

We fit a HMRF model to the data using the `fit_ghm` function. The mixture parameters estimates are shown below and the resulting segmentation is presented in Figure 10(b).

```
R> hmrf_nofixed <- fit_ghm(hfield1, mrfi = R1, theta = pl$theta)
R> summary(hmrf_nofixed)
```

Figure 9: Image data for `hfield1`.

Gaussian mixture model driven by Hidden MRF fitted by EM-algorithm.

Image dimensions: 150 150

Predicted mixture component table:

0	1
10988	11512

Number of covariates (or basis functions): 0

Interaction structure considered: (1,0) (0,1) (4,4)

Mixture parameters:

Component	mu	sigma
0	5.29	1.39
1	9.19	1.46

Model fitted in 4 iterations.

The labels in the segmentation follow the expected pattern only in the middle part of the image. Two large clusters without the pattern appear at the upper and lower parts of the image, what indicates there might be some missing spatial information not included in the model.

Adding a polynomial trend as fixed effect

A HMRF model without covariates has an intrinsic assumption that the mean values of pixel intensities, given their labels, are homogeneous along the image region. This is not the case for the data in Figure 9, as a vertical gradient effect can be observed.

In order to incorporate this spatial effect not captured in the model, we include spatial covariates, in the form of polynomial functions of pixel positions (i_1, i_2) as a fixed effect. These covariates can be specified in `fixed_fn` argument of the function and a (centered) polynomial can be created with the `polynomial_2d()` function from **mrf2d**.

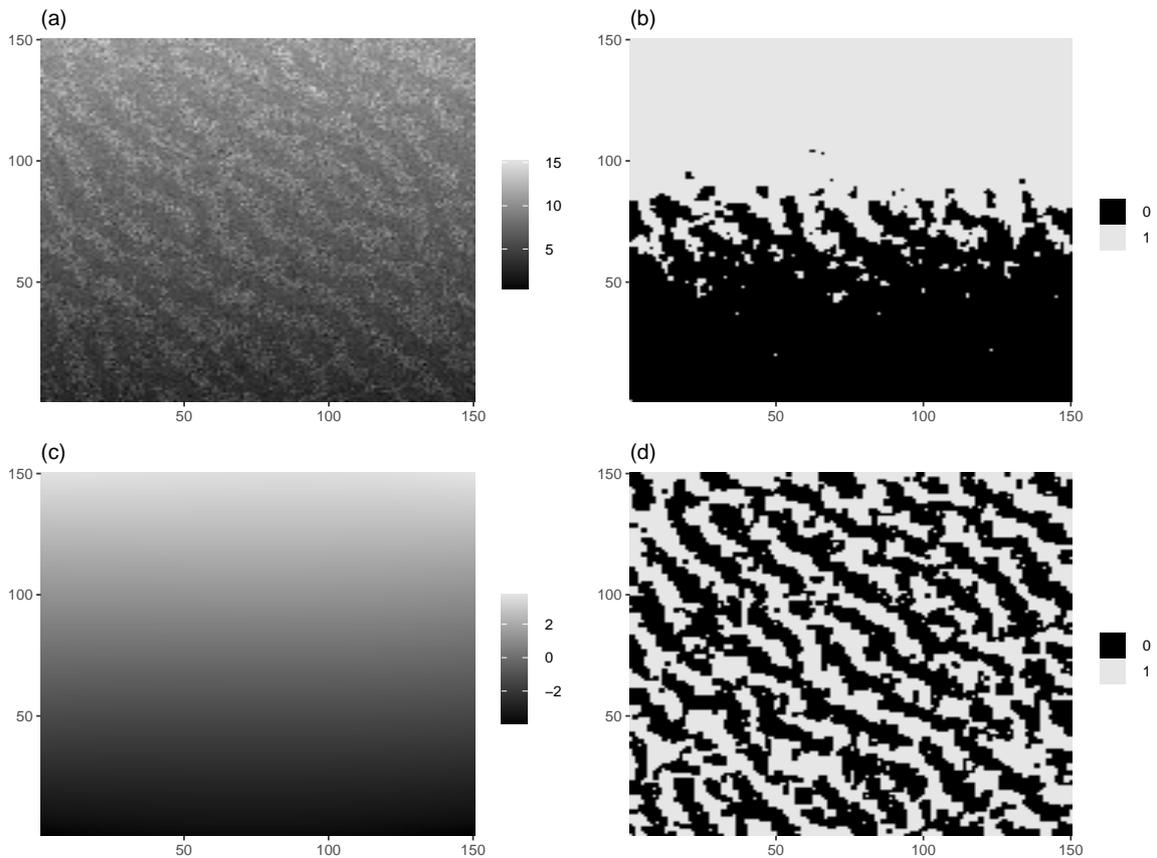


Figure 10: Results from the hidden MRF fits: (a) the original image data, (b) segmentation obtained without adding a polynomial effect, (c) polynomial fitted as a fixed effect, (d) image segmentation when the polynomial effect is included.

In this example, we include all terms of a two-dimensional centered cubic polynomial, that is,

$$p(i_1, i_2) = \sum_{d_1=0}^3 \sum_{d_2=0}^3 \beta_{i_1, i_2} (i_1 - c_1)^{d_1} (i_2 - c_2)^{d_2}, \quad (15)$$

where the centering position (c_1, c_2) is the middle pixel position of the image.

```
R> hmrf_poly <- fit_ghm(hfield1, mrfi = R1, theta = pl$theta,
+   fixed_fn = polynomial_2d(c(3, 3), dim(hfield1)))
R> summary(hmrf_poly)
```

Gaussian mixture model driven by Hidden MRF fitted by EM-algorithm.

Image dimensions: 150 150

Predicted mixture component table:

0	1
11718	10782

Number of covariates (or basis functions): 15

Interaction structure considered: (1,0) (0,1) (4,4)

Mixture parameters:

Component	mu	sigma
0	6.79	0.62
1	7.82	1.20

Model fitted in 4 iterations.

The code chunk below illustrates how the resulting fields available in the function output can be visualized. The results are presented in Figure 10.

```
R> cplot(hfield1)
R> dplot(hmrf_nofixed$Z_pred, legend = TRUE)
R> cplot(hmrf_poly$fixed)
R> dplot(hmrf_poly$Z_pred, legend = TRUE)
```

This example highlights two features of **mrf2d** that are not available in other packages: The possibility to specify a distribution for the underlying field that is more flexible than a simple Potts model and the option to include covariates (in this example, the polynomial trend) that are estimated simultaneously to the mixture parameters, preventing undesired effects in the segmentation results.

4.3. Example 3: Neuroimaging segmentation with BOLD5000 data

Neuroimaging is one of the most frequent applications of HMRF models (Zhang *et al.* 2001; Shah and Chauhan 2015). We illustrate a brain magnetic resonance image segmentation using a sample of the BOLD5000 dataset (Chang, Pyles, Marcus, Gupta, Tarr, and Aminoff 2019) available in the `bold5000` object in the package.

```
R> data("bold5000", package = "mrf2d")
R> cplot(bold5000)
```

Our main goal in this problem is to segment the brain image into large regions corresponding to different elements, like a background, bones, fat, grey matter, white matter, etc.

The most common approach for the segmentation using HMRFs is to consider a simple Potts model (nearest-neighbor interaction structure $\mathcal{R} = \{(1, 0), (0, 1)\}$) and the "onepar" parameter restriction family. The potential associated with different-valued pairs controls, as well as the number of components are considered fixed a priori and will not be discussed in this paper. For the purpose of illustration, we use 4 components ($C = 3$) and the value -1 for the potentials of different-valued pairs.

```
R> Rnn <- mrfi(1)
R> theta_nn <- expand_array(-1, family = "onepar", C = 3, mrfi = Rnn)
```

We add a constraint that all variance parameters of the mixture components must be equal by setting the `equal_vars` parameter to `TRUE`. This improves the results in this problem by preventing some of the mixture components to be estimated with too high variance, what may causes pixels with large and small values to be predicted in the same class with high probability.

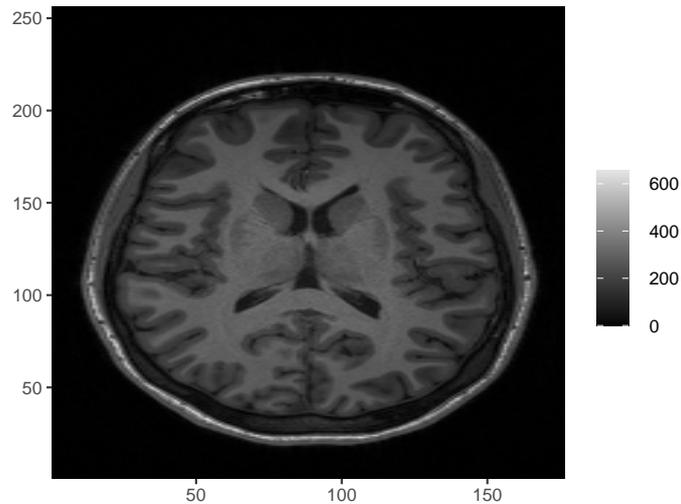


Figure 11: Brain magnetic resonance image in `bold5000` data.

We also fit an independent Gaussian mixture (by multiplying all potentials by zero) for a comparison with the HMRF model. Segmentation results are presented in Figure 12 and the parameter estimates are shown below.

```
R> set.seed(1)
R> fit_brain <- fit_ghm(bold5000, Rnn, theta_nn, equal_vars = TRUE)
R> fit_brain_ind <- fit_ghm(bold5000, Rnn, theta_nn*0, equal_vars = TRUE)
```

```
R> summary(fit_brain)
```

Gaussian mixture model driven by Hidden MRF fitted by EM-algorithm.

Image dimensions: 176 256

Predicted mixture component table:

	0	1	2	3
	22921	6829	7305	8001

Number of covariates (or basis functions): 0

Interaction structure considered: (1,0) (0,1)

Mixture parameters:

Component	mu	sigma
0	7.23	28.78
1	128.90	28.78
2	207.53	28.78
3	294.71	28.78

Model fitted in 11 iterations.

```
R> summary(fit_brain_ind)
```

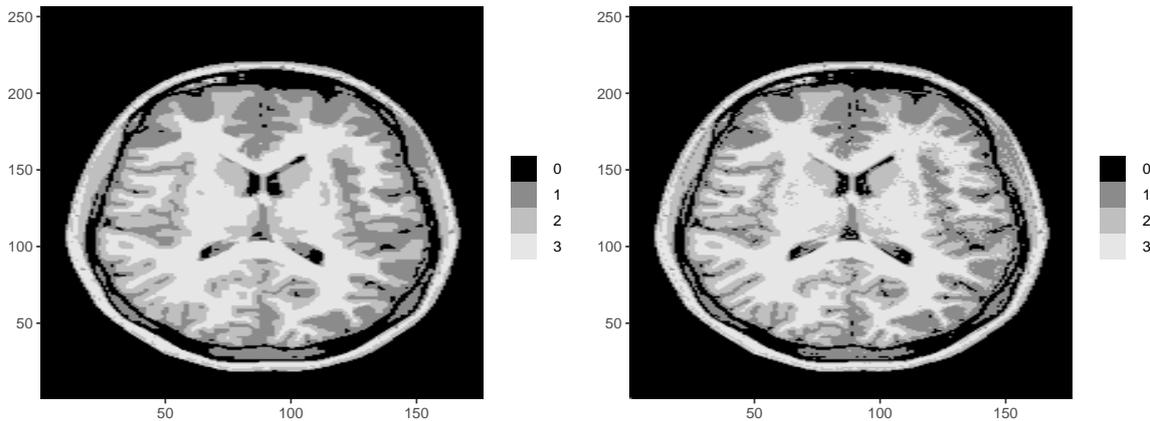


Figure 12: Image segmentation predicted by the hidden MRF fitted (left) and the independent mixture model (right).

Gaussian mixture model driven by Hidden MRF fitted by EM-algorithm.

Image dimensions: 176 256

Predicted mixture component table:

0	1	2	3
23013	7021	7065	7957

Number of covariates (or basis functions): 0

Interaction structure considered: (1,0) (0,1)

Mixture parameters:

Component	mu	sigma
0	7.51	30.23
1	133.10	30.23
2	211.72	30.23
3	295.33	30.23

Model fitted in 4 iterations.

The resulting parameter estimates are not much different when comparing the independent mixture model and the HMRF and both ran in approximately 85 seconds, but the segmentation is cleaner when using the HMRF model, without sparse different-labeled pixels inside regions.

```
R> dplot(fit_brain$Z_pred, legend = TRUE)
```

```
R> dplot(fit_brain_ind$Z_pred, legend = TRUE)
```

5. Discussion

mr2d provides a consistent programming interface for statistical inference in a large class of discrete Markov random field models defined on 2-dimensional lattices. It has an efficient and

simple to use implementation of the main stack of computations used by most of inference algorithms, as well as complete routines for some commonly used and more complex estimation methods. The objects used for representing each model component have been carefully designed and tuned over several iterations to achieve a balance between performance and usability in the stable version.

The model featured in the package generalizes Potts model from other available packages in different ways, such as allowing a flexible definition of interacting pixel positions and interaction types, with the drawback that it cannot take advantage of algorithms that require the setup of a Potts model to improve their efficiency.

The versatility from the non-parametric model behind **mrf2d** and the flexible representation proposed in the package allows us to create special 2-dimensional structures that are equivalent to 3-dimensional representations of data. This can make **mrf2d** also an interesting option for applications with 3-dimensional data where the assumptions of the model also hold. A vignette is available with a deeper explanation on how to reshape 3-dimensional problems for the **mrf2d** framework and it can be viewed by using the `vignette()` function.

```
R> vignette("three-dimensions-on-mrf2d", package = "mrf2d")
```

We currently have over 160 unit tests supported by the **testthat** package (Wickham 2011) and more than 90% of the code covered in the tests. These tests were designed to verify mathematical correctness of functions, the behavior of functions with unexpected input and the consistency of error messages. The package is in constant development and new tests are added whenever new functionalities are implemented to ensure its reliability over time.

For these reasons, **mrf2d** is an important tool for making statistical inference in images using Markov random fields more accessible, allowing researchers to perform data analysis and implement new algorithms in R with a simple and consistent framework.

Acknowledgments

This work was funded by Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP grant 2017/25469-2 and by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

We are thankful to the editors and reviewers for their valuable comments that were essential to improve the manuscript, especially the anonymous referee who commented on 3-dimensional applications that motivated us to add the new vignette.

References

- Besag J (1974). “Spatial Interaction and the Statistical Analysis of Lattice Systems.” *Journal of the Royal Statistical Society B*, **36**(2), 192–225. doi:10.1111/j.2517-6161.1974.tb00999.x.
- Besag J (1975). “Statistical Analysis of Non-Lattice Data.” *Journal of the Royal Statistical Society D*, **24**(3), 179–195. doi:10.2307/2987782.

- Besag J (1986). “On the Statistical Analysis of Dirty Pictures.” *Journal of the Royal Statistical Society B*, **48**(3), 259–279. doi:10.1111/j.2517-6161.1986.tb01412.x.
- Beyer L (2015). “**PyDenseCRF**.” URL <https://github.com/uzeful/pydensecrf>.
- Bhattacharya BB, Mukherjee S (2018). “Inference in Ising Models.” *Bernoulli*, **24**(1), 493–525. doi:10.3150/16-bej886.
- Blake A, Kohli P, Rother C (2011). *Markov Random Fields for Vision and Image Processing*. MIT Press. doi:10.7551/mitpress/8579.001.0001.
- Butts CT (2008). “**network**: A Package for Managing Relational Data in R.” *Journal of Statistical Software*, **24**(2). doi:10.18637/jss.v024.i02.
- Cao X, Zhou F, Xu L, Meng D, Xu Z, Paisley J (2018). “Hyperspectral Image Classification with Markov Random Fields and a Convolutional Neural Network.” *IEEE Transactions on Image Processing*, **27**(5), 2354–2367. doi:10.1109/tip.2018.2799324.
- Chang N, Pyles JA, Marcus A, Gupta A, Tarr MJ, Aminoff EM (2019). “BOLD5000, a Public fMRI Dataset While Viewing 5000 Visual Images.” *Scientific Data*, **6**(1), 1–18. doi:10.1038/s41597-019-0052-3.
- Clark NJ, Wells K, Lindberg O (2018). “Unravelling Changing Interspecific Interactions Across Environmental Gradients Using Markov Random Fields.” *Ecology*, **99**(6), 1277–1283. doi:10.1002/ecy.2221.
- Cross GR, Jain AK (1983). “Markov Random Field Texture Models.” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **5**(1), 25–39. doi:10.1109/tpami.1983.4767341.
- De Bastiani F, Rigby RA, Stasinopoulous DM, Cysneiros AH, Uribe-Opazo MA (2018). “Gaussian Markov Random Field Spatial Models in GAMLSS.” *Journal of Applied Statistics*, **45**(1), 168–186. doi:10.1080/02664763.2016.1269728.
- Derin H, Elliott H (1987). “Modeling and Segmentation of Noisy and Textured Images Using Gibbs Random Fields.” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **9**(1), 39–55. doi:10.1109/tpami.1987.4767871.
- Eddelbuettel D, François R (2011). “**Rcpp**: Seamless R and C++ Integration.” *Journal of Statistical Software*, **40**(8), 1–18. doi:10.18637/jss.v040.i08.
- Eddelbuettel D, Sanderson C (2014). “**RcppArmadillo**: Accelerating R with High-Performance C++ Linear Algebra.” *Computational Statistics & Data Analysis*, **71**, 1054–1063. doi:10.1016/j.csda.2013.02.005.
- Feng D (2018). *PottsUtils: Utility Functions of the Potts Models*. R package version 0.3-3, URL <https://CRAN.R-project.org/package=PottsUtils>.
- Freeman WT, Liu C (2011). “Markov Random Fields for Super-Resolution and Texture Synthesis.” In *Advances in Markov Random Fields for Vision and Image Processing*. MIT Press.

- Freguglia V (2022). **mrf2d**: *Markov Random Field Models for Image Analysis*. R package version 1.0, URL <https://CRAN.R-project.org/package=mrf2d>.
- Freguglia V, Garcia NL, Bicas JL (2020). “Hidden Markov Random Field Models Applied to Color Homogeneity Evaluation in Dyed Textile Images.” *Environmetrics*, **31**(4), e2613. doi:10.1002/env.2613.
- Geman S, Geman D (1984). “Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images.” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**(6), 721–741. doi:10.1109/tpami.1984.4767596.
- Gentleman R, Whalen E, Huber W, Falcon S (2021). **graph**: *A Package to Handle Graph Data Structures*. R package version 1.72.0, URL <https://bioconductor.org/packages/graph/>.
- Geyer CJ, Johnson L (2020). **potts**: *Markov Chain Monte Carlo for Potts Models*. R package version 0.5-9, URL <https://CRAN.R-project.org/package=potts>.
- Geyer CJ, Thompson EA (1992). “Constrained Monte Carlo Maximum Likelihood for Dependent Data.” *Journal of the Royal Statistical Society B*, **54**(3), 657–683. doi:10.1111/j.2517-6161.1992.tb01443.x.
- Ghamisi P, Maggiori E, Li S, Souza R, Tarablaka Y, Moser G, De Giorgi A, Fang L, Chen Y, Chi M, Serpico SB, Benediktsson JA (2018). “New Frontiers in Spectral-Spatial Hyperspectral Image Classification: The Latest Advances Based on Mathematical Morphology, Markov Random Fields, Segmentation, Sparse Representation, and Deep Learning.” *IEEE Geoscience and Remote Sensing Magazine*, **6**(3), 10–43. doi:10.1109/mgrs.2018.2854840.
- Gimel’farb GL (1996). “Texture Modeling by Multiple Pairwise Pixel Interactions.” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(11), 1110–1114. doi:10.1109/34.544081.
- Guillot D, Rajaratnam B, Emile-Geay J (2015). “Statistical Paleoclimate Reconstructions via Markov Random Fields.” *The Annals of Applied Statistics*, **9**(1), 324–352. doi:10.1214/14-aos794.
- Hammersley JM, Clifford P (1971). “Markov Fields on Finite Graphs and Lattices.” Unpublished Manuscript.
- Jensen JL, Künsch HR (1994). “On Asymptotic Normality of Pseudo Likelihood Estimates for Pairwise Interaction Processes.” *Annals of the Institute of Statistical Mathematics*, **46**(3), 475–486. URL <https://link.springer.com/article/10.1007/BF00773511>.
- Kato Z, Pong TC (2006). “A Markov Random Field Image Segmentation Model for Color Textured Images.” *Image and Vision Computing*, **24**(10), 1103–1114. doi:10.1016/j.imavis.2006.03.005.
- Kato Z, Zerubia J (2012). “Markov Random Fields in Image Segmentation.” *Foundations and Trends in Signal Processing*, **5**(1–2), 1–155. doi:10.1561/20000000035.

- Kosov S (2013). “Direct Graphical Models C++ Library.” URL <http://research.project-10.de/dgm/>.
- Krähenbühl P, Koltun V (2011). “Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials.” In *Advances in Neural Information Processing Systems*, pp. 109–117.
- Li M, Wu Y, Zhang Q (2009). “SAR Image Segmentation Based on Mixture Context and Wavelet Hidden-Class-Label Markov Random Field.” *Computers & Mathematics with Applications*, **57**(6), 961–969. doi:10.1016/j.camwa.2008.10.042.
- Liggett TM (2012). *Interacting Particle Systems*, volume 276. Springer-Verlag.
- Møller J, Pettitt AN, Reeves R, Berthelsen KK (2006). “An Efficient Markov Chain Monte Carlo Method for Distributions with Intractable Normalising Constants.” *Biometrika*, **93**(2), 451–458. doi:10.1093/biomet/93.2.451.
- Moore M, Nicholls G, Pettitt A, Mengersen K (2020). “Scalable Bayesian Inference for the Inverse Temperature of a Hidden Potts Model.” *Bayesian Analysis*, **15**(1), 1–27. doi:10.1214/18-ba1130.
- Pickard DK (1987). “Inference for Discrete Markov Fields: The Simplest Nontrivial Case.” *Journal of the American Statistical Association*, **82**(397), 90–96. doi:10.1080/01621459.1987.10478394.
- Potts RB (1952). “Some Generalized Order-Disorder Transformations.” In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 48, pp. 106–109. Cambridge University Press.
- R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Robbins H, Monro S (1951). “A Stochastic Approximation Method.” *The Annals of Mathematical Statistics*, pp. 400–407. doi:10.1214/aoms/1177729586.
- Roche A, Ribes D, Bach-Cuadra M, Krüger G (2011). “On the Convergence of EM-Like Algorithms for Image Segmentation Using Markov Random Fields.” *Medical Image Analysis*, **15**(6), 830–839. doi:10.1016/j.media.2011.05.002.
- Rue H, Martino S, Chopin N (2009). “Approximate Bayesian Inference for Latent Gaussian Models by Using Integrated Nested Laplace Approximations.” *Journal of the Royal Statistical Society B*, **71**(2), 319–392. doi:10.1111/j.1467-9868.2008.00700.x.
- Shah SA, Chauhan NC (2015). “An Automated Approach for Segmentation of Brain MR Images Using Gaussian Mixture Model Based Hidden Markov Random Field with Expectation Maximization.” *Journal of Biomedical Engineering and Medical Imaging*, **2**(4), 57–57. doi:10.14738/jbemi.24.1411.
- Stoehr J, Pudlo P, Friel N (2020). **GiRaF**: A Toolbox for Gibbs Random Fields Analysis. R package version 1.0.1, URL <https://CRAN.R-project.org/package=GiRaF>.
- Storath M, Weinmann A (2014). “Fast Partitioning of Vector-Valued Images.” *SIAM Journal on Imaging Sciences*, **7**(3), 1826–1852. doi:10.1137/130950367.

- Versteegen R, Gimel'farb G, Riddle P (2016). "Texture Modelling with Nested High-Order Markov-Gibbs Random Fields." *Computer Vision and Image Understanding*, **143**, 120–134. doi:[10.1016/j.cviu.2015.11.003](https://doi.org/10.1016/j.cviu.2015.11.003).
- Wang JS, Swendsen RH (1990). "Cluster Monte Carlo Algorithms." *Physica A: Statistical Mechanics and Its Applications*, **167**(3), 565–579. doi:[10.1016/0378-4371\(90\)90275-w](https://doi.org/10.1016/0378-4371(90)90275-w).
- Wickham H (2011). "**testthat**: Get Started with Testing." *The R Journal*, **3**, 5–10. doi:[10.32614/rj-2011-002](https://doi.org/10.32614/rj-2011-002).
- Wickham H (2016). **ggplot2: Elegant Graphics for Data Analysis**. Springer-Verlag, New York.
- Winkler G (2012). *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods: A Mathematical Introduction*, volume 27. Springer-Verlag.
- Wood SN (2017). *Generalized Additive Models: An Introduction with R*. 2nd edition. Chapman & Hall/CRC.
- Wu LY (2019). **CRF: Conditional Random Fields**. R package version 0.4-3, URL <https://CRAN.R-project.org/package=CRF>.
- Zhang Y, Brady M, Smith S (2001). "Segmentation of Brain MR Images through a Hidden Markov Random Field Model and the Expectation-Maximization Algorithm." *IEEE Transactions on Medical Imaging*, **20**(1), 45–57. doi:[10.1109/42.906424](https://doi.org/10.1109/42.906424).

A. Customizing visualizations with ggplot2

A.1. Random field visualization

The two plotting functions `dplot()` and `cplot()` return an object of the ‘`ggplot`’ class, what allows users to produce customized visualizations by changing scales, legend characteristics, titles, themes and much more by using the grammar of graphics from the `ggplot2` package (Wickham 2016).

Random fields (represented by `matrix` objects) are transformed into a `data.frame` structure with columns `x`, `y` and `value`. `x` and `y` are the indices of the matrix object while `value` maps the pixel-value in that position ($Z[x,y]$).

The plots are constructed using a tile plane with rectangles (`geom_tile()` from `ggplot2`) with the `value` column map to the `fill` aesthetics. In `dplot()`, which is used for finite-valued fields, `value` is treated as a factor, while in `cplot()` it is a continuous numeric. This is the only difference between the functions and it should be kept in mind when defining custom color scales.

Figure 13 created with the code chunk below shows examples of customized versions of a random field visualization built from the same base `dplot()` result. Modifications include adding a title, removing all scale-related information (keeping only the actual image), using a custom color-scale and changing the legend position, respectively.

```
R> library("ggplot2")
R> base_plot <- dplot(field1, legend = TRUE)
R> base_plot + ggtitle("This is a custom title")
R> base_plot + theme_void() + theme(legend.position = "none")
R> base_plot + scale_fill_manual(values = c("red", "blue"))
R> base_plot + theme(legend.position = "bottom")
```

A.2. Interaction structure visualization

Similarly to the functions used to visualize random fields, a `plot()` method is available for `mrfi` objects. A `data.frame` with columns named `rx` and `ry` is created internally with the coordinates of each interacting relative position. These columns are mapped to the `x` and `y` axis, respectively and a `geom_tile` is used to produce the plot. The reverse positions are included automatically with a light-gray color, but this can be prevented by setting `include_opposite = FALSE` in the plot call. It also returns a `ggplot` object that can be customized. The code chunk next presents examples of how plots can be customized by adding custom colors, text labels to the interacting positions and a custom title.

```
R> mrfi_plot <- plot(mrfi(3) + c(5,1))

R> mrfi_plot + geom_tile(fill = "orange", color = "blue")
R> mrfi_plot + geom_text(aes(label = paste0("(", rx, ", ", ry, ")")))
R> mrfi_plot + ggtitle("Add a custom title") +
+   theme(plot.title = element_text(hjust = 0.5, size = 24))
```

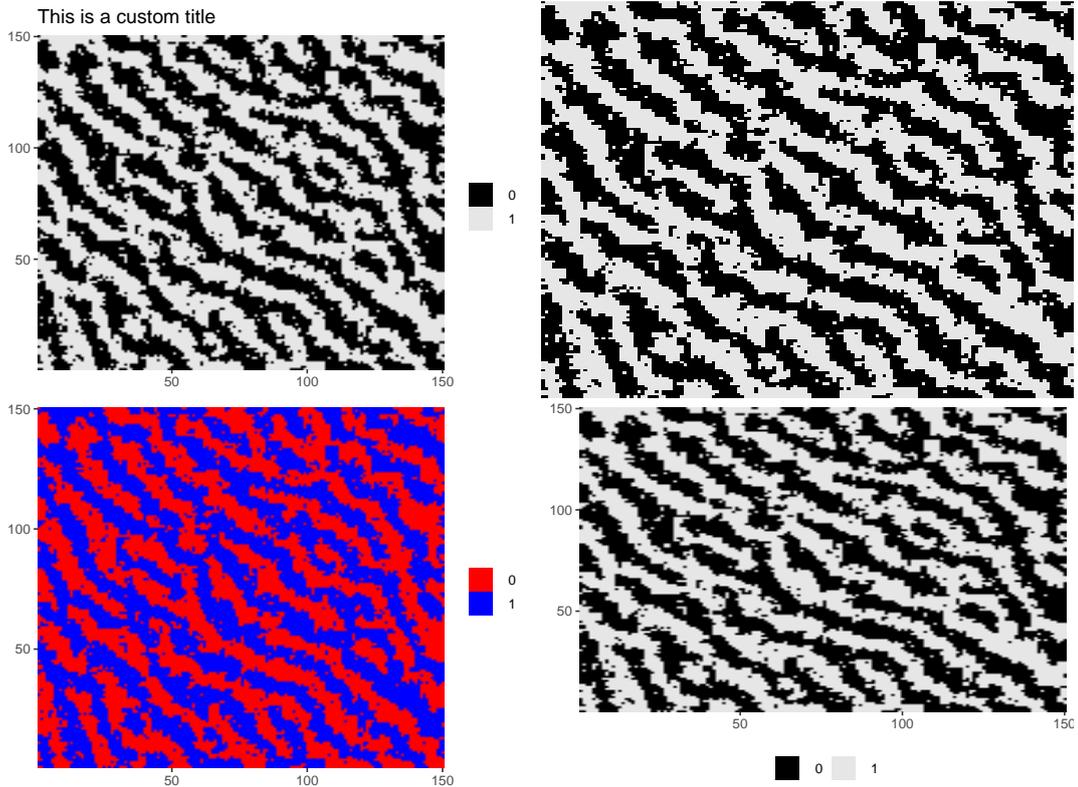


Figure 13: Four examples of field visualizations achieved by adding `ggplot2` layers to a base plot produced in `mrf2d`.

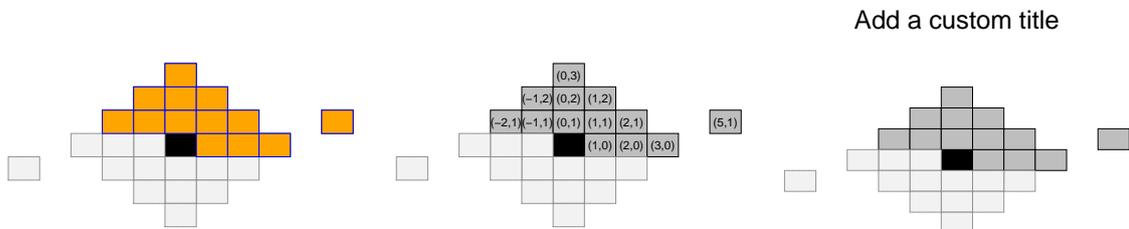


Figure 14: Examples of customized visualizations of `mrfi` objects.

B. Comparing methods between packages for a Potts model

One of the most important particular cases for the model presented in Section 2.1.1 is the Potts model, which is used in the theory supporting many available packages. In particular, the `potts` package (Geyer and Johnson 2020) provides functionalities similar to some of the ones implemented in `mrf2d`, mainly random sampling and computing the composite likelihood (comparable to the pseudo-likelihood function). The main difference between the packages is that `mrf2d` uses a Gibbs sampler scheme, with sequential updates of individual pixel while `potts` uses the Swendsen-Wang algorithm (Wang and Swendsen 1990), that updates blocks of pixels at each iteration.

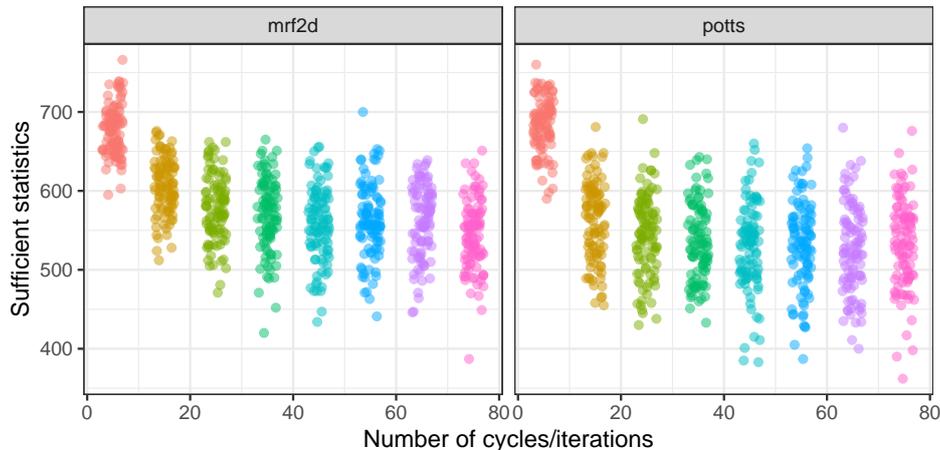


Figure 15: Sufficient statistics for 100 simulated realizations of the Potts Model with varying number of iterations using both **mrf2d** and **potts** packages.

We conducted a simulation study in order to compare **mrf2d** and the **potts** package version 0.5-9 with respect to the Potts model simulation algorithms. Our goal was not to compare implementations as the algorithms are different, but to check whether the simulated fields are comparable between packages as a form of validation and understanding the behavior of our simulation method with respect to the number of Gibbs sampler cycles.

Considering the parametrization of the Potts model in Equation 4, 100 realizations of a three-color ($C = 2$) Potts model with parameter $\phi = -1$ (corresponds +1 in the parametrization used in the **potts** package) were simulated considering number of iterations equal to 5, 15, 25, 35, 45, 55, 65 and 75 in each package. We use the term iteration referring to the `cycles` parameter in **mrf2d** and `nspace` in **potts**. For each simulated field, we computed the sufficient statistics $T(\mathbf{z}) = \sum_{\|i-j\|=1} \mathbb{1}_{(z_i \neq z_j)}$. The results are presented in Figure 15 and the code for the simulations is available in the replication script for this article.

We can consider the algorithms achieved equilibrium when executing more iterations do not change the distribution of the samples, which we summarize by the distribution of their sufficient statistics. In this case, the Gibbs sampler from **mrf2d** seems to achieve equilibrium within 35 to 45 cycles while roughly a value of 25 for the `nspace` parameter of **potts** seems to be enough to achieve equilibrium in the Swendsen-Wang algorithm. Finally, we verify that the distribution of the sufficient statistics after reaching an equilibrium state is approximately the same in both packages and we conclude that even though **potts** uses the more specialized and efficient Swendsen-Wang algorithm, **mrf2d** with a Gibbs sampler provides a valid alternative simulation tool for the model that can also be extended, even within the scope of the Potts model, to cases such as positive ϕ , not covered by the Swendsen-Wang algorithm.

We also compare estimation algorithms in both packages for the maximum pseudo-likelihood method, which is a particular case of composite likelihood, thus available in **potts**. We simulated 100 realizations of a 3 color Potts model in a 64×64 lattice again with the parameter $\phi = -1$. For each realization, we computed maximum pseudo-likelihood estimates using **mrf2d** and **potts** to compare the results. Results are presented in Figure 16. In our comparison, estimates obtained with **potts** were consistently greater (smaller absolute value) than

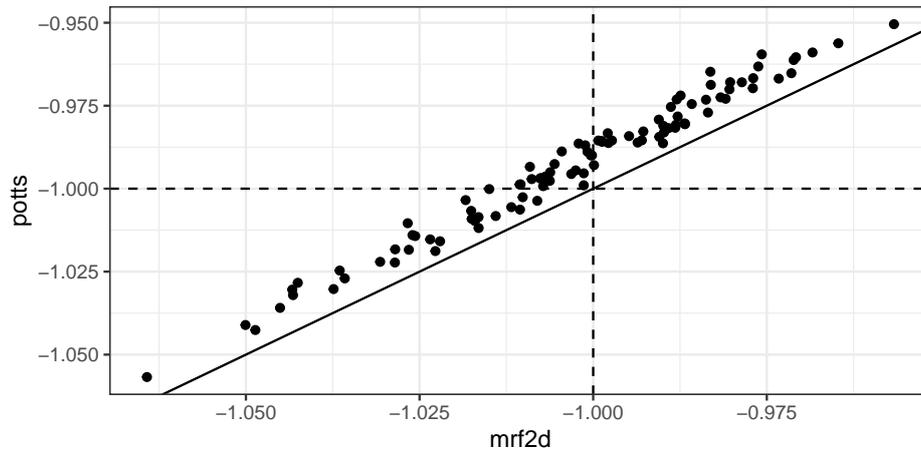


Figure 16: Estimated parameter via pseudo-likelihood optimization for 100 simulated realizations of the Potts model using the **mrf2d** and **potts** packages. The solid line represents the $y = x$ equation and dashed lines mark the parameter value used in the simulations.

the ones obtained via **mrf2d**. The average estimate for **mrf2d** was -1.0027996 while for **potts** the average was -0.9930211 , with sample deviations equal to 0.0213258 and 0.0216528 , respectively. We conclude that maximum pseudo-likelihood estimation methods are roughly equivalent in both packages.

Affiliation:

Victor Freguglia, Nancy Lopes Garcia
 University of Campinas
 Institute of Mathematics, Statistics and Scientific Computation
 Campinas, Brazil
 E-mail: victorfreguglia@gmail.com, nancyg@unicamp.br
 URL: <https://freguglia.github.io>