Unsupervised image segmentation with Gaussian Pairwise Markov Fields

Hugo Gangloff, Jean-Baptiste Courbot, Emmanuel Monfrini, Christophe Collet

Unsupervised Image Segmentation with Gaussian Pairwise Markov Fields

Hugo Gangloff[a,b,*], Jean-Baptiste Courbot[c], Emmanuel Monfrini[d], Christophe Collet[a]

[a] *ICube - CNRS UMR 7357, Université de Strasbourg - CNRS, Illkirch, France*
[b] *GEPROVAS, Strasbourg, France*
[c] *IRIMAS UR 7499, Université de Haute-Alsace, Mulhouse, France*
[d] *SAMOVAR - CNRS UMR 5157, Télécom SudParis, Institut Polytechnique de Paris, Évry, France*

Abstract

Modeling strongly correlated random variables is a critical task in the context of latent variable models. A new probabilistic model, called Gaussian Pairwise Markov Field, is presented to generalize existing Markov Fields latent variables models, and to introduce more correlations between variables. This is done by considering the correlations within Gaussian Markov Random Fields models which are much richer than in the classical Markov Field models. The assets of the Gaussian Pairwise Markov Field model are explained. In particular, it offers a generalization of the classical Markov Field modelization that is highlighted. The new model is also considered in the practical case of unsupervised segmentation of images corrupted by long-range spatially-correlated noise, producing interesting new results.

Keywords: unsupervised image segmentation, Pairwise Markov Fields, Gaussian Markov Fields, parameter estimation

1. Introduction

Unsupervised image segmentation is a vast field which deals with the task of labelling each pixel appropriately, in homogeneous and meaningful regions, without any ground truth available [41]. Popular approaches to solve this problem include clustering-based methods [1], graph-based methods [6] or Bayesian methods [4]. There have been many developments of this topic over the years but very few seem to deal with the processing of images corrupted with long-range, spatially-correlated noise. Such problem might be approached with an explicit and precise modelization of the noise through the development of a new Bayesian model, as we propose in this paper.

Bayesian image segmentation is often based on the classical Hidden Markov Field (HMF) model. In the latter, the image is treated as an observed random variable, while the desired segmentation map is hidden. When strong spatially-correlated noise corrupts the image, classical approaches reach their limit and new dedicated models need to be considered to improve the accuracy of the segmentation.

On the one hand, more recent probabilistic models have been developed in order to handle rich correlations between random variables. Within latent variables models, Pairwise Markov Fields (PMF) [14] [25] [34] are a model family generalizing the HMF model, thus enriching the law of the sample. In this case, the latent and observed random variables jointly form the Markov process. An appealing property of PMF is that they relax the Markovian assumption on the hidden process, which enables the modeling of more complex correlations, while keeping Bayesian inference easily available. Figure 1 depicts the graphical models of the classical HMF, of PMF and of the numerous direct correlations that can be introduced between variables with the latter model. PMF have in turn been extended to Triplet Markov Fields [2][13] in which a third auxiliary process is added to improve the modelization possibilities, such as non-stationarities in the model parameters.

On the other hand, Gaussian Markov Random Fields (GMRF) [37] are a powerful probabilistic model to deal with a large variety of correlated random variables, and especially with long-range spatial correlation between pixels in images. Studies have been done for regression and classification using *conditional* GMRF, *i.e.* these studies consider a discriminative probabilistic model [33][35][40]. However, to the best of our knowledge, there exists no literature on generative probabilistic models using GMRF for image segmentation. Generative models might be advantageous when no dataset is available, for example, in the case of unsupervised image segmentation. This article aims at proposing such a new probabilistic modelization.

Therefore, we introduce a new model, called Gaussian Pairwise Markov Fields (GPMF), belonging to the PMF family, and defined such that the GPMF conditional likelihood is a GMRF. The new model then combines a generalization of the classical Gaussian HMF model and the ability to model strongly correlated variables as a GMRF, while preserving tractability. Indeed, we will see that the PMF hypothesis is a natural way to answer the problem of modeling correlated noise and introducing long correlations by using the GMRF model as conditional likelihood.

*Corresponding author. Hugo Gangloff (hugogangloff@unistra.fr) is with ICube UMR 7357, Université de Strasbourg - CNRS, 300 Bd Sébastien Brant, 67400 Illkirch-Graffenstaden, France.

We start by deriving the general probability distribution that defines a GPMF. We then propose four intances of the GPMF model whose expressiveness varies, in terms of direct dependencies that the model can handle. Those models are studied both theoretically and through experiments. Besides, we handle the unsupervised image segmentation problem which requires a crucial step of parameter estimation. To this end, we propose a stochastic parameter estimation algorithm for the PMF models. In the case of unsupervised image segmentation, the GPMF model performs better than other classical unsupervised approaches.

The outline of the paper is the following. We first describe the new model, its core equations and main properties (Section 2). Then we illustrate the successive generalizations made with GPMF by studying models which are special cases of the GPMF (Section 3). We develop a parameter estimation procedure to solve the problem of unsupervised segmentation (Section 4). Finally, the models are successfully evaluated on synthetic and on real world images linked with a medical application (Section 5).

Remark: They are other works in the literature called Pairwise Markov Fields, for example, [18] and [32]. However, those works mostly focus on structure learning with only visible variables. Thus, they should not be confused with the latent variables models presented in this article.

2. Gaussian Pairwise Markov Fields

2.1. Pairwise Markov Fields

This first section introduces main definitions, Hidden Markov Fields (HMF) and Pairwise Markov Fields (PMF).

$\boldsymbol{X} = (X_1, \ldots, X_N)$ is a discrete-valued random vector with values in Ω^N, with $\Omega = (\omega_1, \ldots, \omega_K)$. $\boldsymbol{Y} = (Y_1, \ldots, Y_N)$ is a real-valued random vector with values in $\mathbb{R}^N$. In this article, we deal with stationary Markov processes defined on a graph, whose vertices are indexed by $\mathcal{S}$ such that $|\mathcal{S}| = N$, with respect to a neighborhood $\mathcal{N}$. A neighborhood $\mathcal{N}_s$ is defined as the set of vertices in relation with a given vertice s ($s \notin \mathcal{N}_s$). We also denote by $\mathcal{C}_n$ the set of cliques of size n [28].

We use the notation shortcut $p(\boldsymbol{X} = \boldsymbol{x}) = p(\boldsymbol{x})$ to refer to the probability of the realizations of random vectors and variables and $\tilde{p}$ to refer to an unnormalized probability distribution. The classical Hidden Markov Field (HMF) model with independent noise is defined by the joint distribution [26]:

$$p(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{Z} \prod_{s \in \mathcal{S}} \tilde{p}(x_s | \boldsymbol{x}_{\mathcal{N}_s}) \tilde{p}(y_s | x_s), \tag{1}$$

where Z is the normalization constant. In this model, $\boldsymbol{X}$ is a Markovian process with respect to the neighborhood $\mathcal{N}$.

The PMF family of models generalizes the HMF model [14] [34]. In the PMF family, the assumption of $(\boldsymbol{X}, \boldsymbol{Y})$ being a Markov process is made. Note that this implies that $\boldsymbol{X}$ given some realizations $\boldsymbol{y}$ of $\boldsymbol{Y}$ is a Markov process. A PMF is defined by the distribution:

$$p(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{Z} \prod_{s \in \mathcal{S}} \tilde{p}(x_s, y_s | \boldsymbol{x}_{\mathcal{N}_s}, \boldsymbol{y}_{\mathcal{N}_s}), \tag{2}$$

where Z is the normalization constant. The greater generality of the pairwise fields hypothesis (that can be seen in Equation 2) enables the modeling of more complex correlations between variables. Indeed, neither $\boldsymbol{X}$ nor $\boldsymbol{Y}$ is necessarly a Markovian process.

In the context of unsupervised image segmentation, the $\{X_n\}_{n=1}^N$ are the *hidden* variables, and the $\{Y_n\}_{n=1}^N$ are the *visible* or *observed* variables.

In the context of Bayesian image segmentation, one can use the Maximum A Posteriori (MAP) estimator [28]:

$$\hat{\boldsymbol{x}}^{MAP} = \text{argmax}_{\boldsymbol{x} \in \Omega^N} p(\boldsymbol{x} | \boldsymbol{y}), \tag{3}$$

as well as the Maximum Posterior Mode (MPM) estimator [29]:

$$\hat{x}_s^{MPM} = \text{argmax}_{x_s \in \Omega} p(x_s | \boldsymbol{y}), \forall s \in \mathcal{S}. \tag{4}$$

In the context of random fields, the MAP estimator is classicaly approximated with an algorithm using simulated annealing approaches [24]. The MPM can be approximated by sampling techniques based on the Gibbs sampler [29] and the following local expression of the posterior Markov field:

$$p(x_s | \boldsymbol{x}_{\mathcal{N}_s^X}, \boldsymbol{y}) = \frac{\tilde{p}(x_s, y_s | \boldsymbol{x}_{\mathcal{N}_s}, \boldsymbol{y}_{\mathcal{N}_s})}{\sum_{x'_s} \tilde{p}(x'_s, y_s | \boldsymbol{x}_{\mathcal{N}_s}, \boldsymbol{y}_{\mathcal{N}_s})}. \tag{5}$$

2.2. Graphical representations

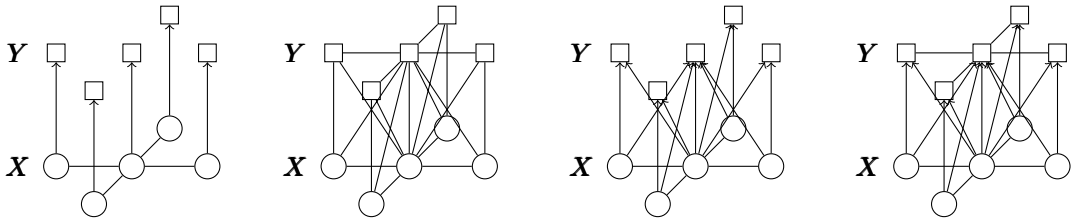
Figure 1 illustrates the graphical representations, for a four-neighbor system, of the HMF and PMF models defined so far as well as intermediate models in terms of dependencies that can be modeled. Note that an edge represents a dependency that can possibly be directly modeled, while in the absence of edge, the direct dependency cannot be modeled. It appears clearly that the PMF model is the most general.

In the case when $\boldsymbol{Y}$ is defined as a continuous Gaussian process, probabilistic models from each of the four types shown in Figure 1 will be defined and used in experiments later in this article.

2.3. Gaussian Pairwise Markov Fields

In this section we define the new Gaussian Pairwise Markov Field (GPMF) model. While it is a subclass of the PMF model, it is still a very rich family which can rigorously model complex situations.

We state that a GPMF is a PMF where the distribution of $\boldsymbol{Y}$ given a realization $\boldsymbol{X} = \boldsymbol{x}$ is a GMRF. Then we have the following property: $(\boldsymbol{X}, \boldsymbol{Y})$ is a GPMF with respect to



(a) HMF (b) PMF (c) PMF-UN (d) HMF-CN

Figure 1: Graphical models of (a) the classical Hidden Markov Field (HMF) with independent noise, (b) the Pairwise Markov Field (PMF) with full dependencies, (c) the Pairwise Markov Field-Uncorrelated Noise (PMF-UN), (d) and the Hidden Markov Field-Correlated Noise (HMF-CN). A graph edge indicates a variable upon which the local conditional distribution of the Markov process is directly conditioned. An arrow indicates a one-way dependency while an edge indicates a two-way dependency.

a neighborhood $\mathcal{N}$ if, and only if, $p(\boldsymbol{x},\boldsymbol{y})$ has the following joint Gibbs distribution with respect to $\mathcal{N}$:

$$p(\boldsymbol{x},\boldsymbol{y}) = \frac{1}{Z}\exp\big(-E(\boldsymbol{x},\boldsymbol{y})\big), \tag{6}$$

where Z is the normalization constant and $E(\boldsymbol{x},\boldsymbol{y})$ is the joint energy with:

$$E(\boldsymbol{x},\boldsymbol{y}) = \sum_{n=1}^{|\mathcal{N}|}\Big(\sum_{\boldsymbol{c}\in\mathcal{C}_n} V_n(\boldsymbol{x_c},\boldsymbol{y_c})\Big). \tag{7}$$

In Equation 7, the V_n are potentials functions subject to constraints:

- $\forall n \in \{1,2\}, \sum_{\boldsymbol{c}\in\mathcal{C}_n} V_n(\boldsymbol{x_c},\boldsymbol{y_c}) = \sum_{\boldsymbol{c}\in\mathcal{C}_n} \bar{V}_n(\boldsymbol{y_c},\boldsymbol{x_c}) + \sum_{\boldsymbol{c}\in\mathcal{C}_n} \tilde{V}_n(\boldsymbol{x_c})$, where $\sum_{\boldsymbol{c}\in\mathcal{C}_n} \bar{V}_n(\boldsymbol{y_c},\boldsymbol{x_c})$ is a positive semidefinite quadratic form in the $\boldsymbol{y}$ variables and $\tilde{V}_n$ are potential functions where the $\boldsymbol{y}$ variables have no play.
- $\forall n \in \{3,\dots,|\mathcal{N}|\}, V_n$ are potential functions where the $\boldsymbol{y_c}$ variables have no play.

Equivalently, the GPMF can be defined by its local conditional factorization (Equation 2). It can be shown, starting from Equation 7, that:

$$\tilde{p}(x_s,y_s|\boldsymbol{x}_{\mathcal{N}_s},\boldsymbol{y}_{\mathcal{N}_s}) = \exp\Bigg(-\Big(\sum_{n=1}^{|\mathcal{N}|}\big(n\sum_{\substack{\boldsymbol{c}\in\mathcal{C}_n\\ \text{s.t. } s\in\boldsymbol{c}}} V_n(\boldsymbol{x_c},\boldsymbol{y_c})\big)\Big)\Bigg), \tag{8}$$

This proposition is proven in Appendix A.

Remark: One can draw a parallel with a specific type of PMF developed in [34], which is also called "Gaussian". Indeed, the decomposition of the model given in [34] makes their model seem to appear as a special case of our new GPMF model. Note that, unlike the article we mention, our definition ensures that $p(\boldsymbol{y}|\boldsymbol{x})$ is defined and is a GMRF.

3. Models from the GPMF family

After this general introduction of GPMF, we now define models of varying generalities, which belong to the GPMF family. Those models will subsequently be used in image segmentation experiments.

3.1. A general instance of the GPMF model

This is the most general instance of GPMF that we will define. Since there will be no ambuigity, we will denote this model as GPMF in the rest of this article. Its graphical representation is given in Figure 1b: we will take advantage of all the dependencies that can be modeled.

We define the GPMF model energy by:

$$E(\boldsymbol{x},\boldsymbol{y}) = \sum_{s\in\mathcal{S}}\sum_{s'\in\mathcal{N}_s} V(x_s,y_s,x_{s'},y_{s'}) + \sum_{s\in\mathcal{S}}\sum_{\substack{s'\in\\ \mathcal{N}_s\cup\{s\}}}\Big[\mathbb{1}_{[s'\in\mathcal{N}_s^2]}\frac{1}{2}Q_{s,s'}\bar{y}_s\bar{y}_{s'}\Big], \tag{9}$$

and the potential function V is given by:

$$V(x_s,y_s,x_{s'},y_{s'}) = -\mathbb{1}_{[s'\in\mathcal{N}_s^1]}\delta_{x_s}^{x_{s'}}\beta\left(1-\frac{1}{2}(\bar{y}_s-\bar{y}_{s'})^2\right), \tag{10}$$

where we denote $\bar{y}_s = y_s - \mu_{x_s}$ and δ is the Kronecker delta function.

Equivalently, the corresponding unnormalized local conditional probabilities are:

$$\tilde{p}(x_s,y_s|\boldsymbol{x}_{\mathcal{N}_s},\boldsymbol{y}_{\mathcal{N}_s}) = \exp\Bigg(-\Big(2\sum_{s'\in\mathcal{N}_s} V(x_s,y_s,x_{s'},y_{s'}) + \frac{1}{2}Q_{s,s}\bar{y}_s^2 + \sum_{s'\in\mathcal{N}_s}\mathbb{1}_{[s'\in\mathcal{N}_s^2]}Q_{s,s'}\bar{y}_s\bar{y}_{s'}\Big)\Bigg), \tag{11}$$

In Equations 9, 10 and 11, we encounter several parameters, a granularity parameter $\beta\in\mathbb{R}_+^*$, a spatially varying vector $\boldsymbol{\mu_x}$ (which depend on the realization $X=x$) and a symmetric positive semidefinite (SPD) matrix $\boldsymbol{Q}$ such that:

$$\forall s\in\mathcal{S},\ \text{if } s'\notin\mathcal{N}_s^2, Q_{s,s'}=0. \tag{12}$$

Moreover $\mathcal{N}^1$ and $\mathcal{N}^2$ are neighborhood systems, which we call *subneighborhoods*, which enable to modulate the summation ranges.

This GPMF model then offers new modelizations properties, notably, it is able to handle Gaussian spatially correlated noise. Moreover, in addition to the correlated

noise, $\boldsymbol{X}$ is clearly not Markovian due to mixed products between terms depending on $\boldsymbol{x}$ and $\boldsymbol{y}$. In Appendix B, we show that the proposed model complies with the definition of the GPMF model given in Section 2.3.

Remark: The potential function of Equation 10 can be seen as an enhanced Potts potential [28] where the granularity coefficient β is adjusted to the observations available in the neighborhood of each site.

3.2. Details on the GPMF model parameters

We now detail the model parameters that define the most general instance of GPMF model proposed in the previous section. Similar parameters will be found in the other models we will define.

We first need to explain an approximation that is made for the model to be tractable. This approximation also explains the origin of the parametrization seen in Section 3.1: it justifies the link between the model parameters and a GMRF conditional likelihood. Indeed, $\boldsymbol{Q}$ is a sparse SPD matrix which is approximated to be the precision matrix of the GMRF defined by the conditional likelihood of the GPMF model. By definition of a GMRF we have:

$$p(\boldsymbol{y}|\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^N \det(\boldsymbol{R}^{-1})}} \exp\left(\frac{1}{2}\bar{\boldsymbol{y}}^T \boldsymbol{R}\bar{\boldsymbol{y}}\right). \quad (13)$$

In the latter equation, $\boldsymbol{R}$ is the original precision matrix of the GMRF conditional likelihood and we have $\boldsymbol{R} = \boldsymbol{P} + \boldsymbol{Q}$. $\boldsymbol{P}$ (which will be defined in Equation B.4) approximates a diagonally banded matrix whose non-zero entries are much smaller than $\boldsymbol{Q}$ entries (practical β values are much lower than 1), hence $\boldsymbol{Q} \approx \boldsymbol{R}$. So by making the assumption $\boldsymbol{Q} = \boldsymbol{R}$, *i.e.*, $\boldsymbol{Q}$ is the precision matrix of a GMRF field, we can derive a computationally efficient parameter estimation procedure that is described in Section 4. Therefore, from now on, we will carry the description of the parameters in terms of GMRF vocabulary.

Remark: In the following, the variance of the GMRF is stationary, as opposed to the mean of the GMRF. The simulation of the GMRF must be carried through its conditional equations [7][37] which are updated at each iterations since the neighbor values change. In this context it is known that introducing a non-stationary variance is very complex [23], and this is out of scope of the article.

• r and σ. Here we describe the precision matrix $\boldsymbol{Q}$ of the GMRF. $\boldsymbol{Q}$ is linked to the covariance matrix $\boldsymbol{\Sigma}$ such that $\boldsymbol{Q}^{-1} = \boldsymbol{\Sigma}$. The latter is assumed to be generated by an exponential correlation function with decay $r \in \mathbb{R}_+^*$. Let $\|.\|_t$ be the Euclidean distance on the torus (of dimensions $L_1 \times L_2$). For $\boldsymbol{a} = (a_1, a_2) \in \mathbb{R}^2, \boldsymbol{b} = (b_1, b_2) \in \mathbb{R}^2$, it is defined by:

$$\|\boldsymbol{a} - \boldsymbol{b}\|_t = \Big(\min(|a_1 - b_1|, L_1 - |a_1 - b_1|)^2 + \min(|a_2 - b_2|, L_2 - |a_2 - b_2|)^2 \Big)^{\frac{1}{2}}. \quad (14)$$

Then, $\forall a \in \mathbb{R}^2, \forall b \in \mathbb{R}^2$, the correlation function is:

$$\rho(a, b; r) = \exp\left(-\frac{\|a - b\|_t}{r}\right). \quad (15)$$

Finally the covariance between two observed variables is:

$$\forall (s, s') \in \mathcal{S}^2, Cov(Y_s, Y_{s'}) = \Sigma_{s,s'} = \sigma^2 \rho(s, s'; r), \quad (16)$$

where σ is another parameter of the model. It is a strictly positive scalar that we associate with the (constant) variance of the GMRF. It is important to note that r and σ totally define $\boldsymbol{\Sigma}$ and then $\boldsymbol{Q}$.

• $\mathcal{N}^2$. Following the common practice in GMRF studies, $\boldsymbol{\Sigma}$ is a dense matrix but its values are getting arbitrarly small according to the range r. The same is true for $\boldsymbol{Q} = \boldsymbol{\Sigma}^{-1}$. Then, the values of $\boldsymbol{Q}$ whose magnitude drops below a given theshold are forced to 0 to induce sparsity. The values that are set to 0 are those outside $\mathcal{N}^2$. The Markovianity of $\boldsymbol{Y}$ given a realization $\boldsymbol{x}$ of $\boldsymbol{X}$ is then reflected in $\boldsymbol{Q}$ because we assumed $\boldsymbol{Q} = \boldsymbol{R}$.

• $\boldsymbol{\mu}$ is a spatially-varying vector that we associate with the non-stationary mean of the GMRF. It depends on $\boldsymbol{x}$, the realizations of $\boldsymbol{X}$, such that $\forall s \in \mathcal{S}, \mu_s = \mu_{x_s} \in \mathbb{R}$.

• $\beta \in \mathbb{R}_+^*$ is a coefficient whose role is similar to the granularity parameter in a classic Potts model [28].

• $\mathcal{N}^1$ is a subneighborhood that can be freely chosen. In the following we choose the classic eight-nearest-neighbor system.

Remark: The periodic boundary assumption [37] (also called torus assumption) is made. In this context, the borders of the image are wrapped to induce a continuity along the rows and columns of the image. When the assumption is done, $\boldsymbol{Q}$ is a block-circulant matrix with circulant blocks. It follows that the operations involving matrices (inverses, determinants, ...) become computationally efficient since they rely on Fourier transforms. We also gain in efficiency for the storing of $\boldsymbol{Q}$: because of the special matrix structure, we only need to store the *base* matrix of $\boldsymbol{Q}$ which has the same dimension as the image. The periodic boundary assumption is then crucial to build a computationally tractable model and, in practice, it was not a limiting factor to the quality of the results.

3.3. Other GPMF submodels

In this section, we propose three other models that are particular, specific cases of GPMF model. We first present some new local distributions and then we use them to define the particular GPMF models.

Let us recall the Potts prior for Markov fields [26], which is based on the following local conditional probabilities:

$$\tilde{p}(x_s|\boldsymbol{x}_{\mathcal{N}_s}) = \exp\left(-2 \sum_{s' \in \mathcal{N}_s} V^P(x_s, x_{s'})\right), \quad (17)$$

where V^P is the Potts potential function making use of a granularity coefficient β:

$$V^P(x_s, x_{s'}) = -\mathbb{1}_{[s' \in \mathcal{N}_s^1]} \delta_{x_s}^{x_{s'}} \beta. \tag{18}$$

Let us also recall the independent Gaussian conditional likelihood [26] which leads to the local conditional probabilities:

$$\tilde{p}(y_s|x_s) = \exp\left(-\ln\left(\sqrt{2\pi}\sigma\right) - \frac{\bar{y}_s^2}{2\sigma^2} \right). \tag{19}$$

We also have the strict generalization of Equation 19 to the conditional GMRF likelihood in its local form [7] [37]:

$$\tilde{p}(y_s|x_s, \boldsymbol{x}_{\mathcal{N}_s}, \boldsymbol{y}_{\mathcal{N}_s}) = \exp\left(-\ln\left(\sqrt{2\pi v_s}\right) - \frac{(y_s - m_s)^2}{2v_s} \right), \tag{20}$$

with

$$m_s = \mu_{x_s} - v_s \sum_{s' \in \mathcal{N}_s} \mathbb{1}_{[s' \in \mathcal{N}_s^2]} Q_{s,s'} \bar{y}_{s'} \text{ and } v_s = Q_{s,s}^{-1}. \tag{21}$$

Let us also introduce a generalized version of the Potts potential taking into account the spatial context, using the function V defined in Equation 10:

$$\tilde{p}(x_s|y_s, \boldsymbol{x}_{\mathcal{N}_s}, \boldsymbol{y}_{\mathcal{N}_s}) = \exp\left(-\left(2 \sum_{s' \in \mathcal{N}_s} V(x_s, y_s, x_{s'}, y_{s'})\right)\right). \tag{22}$$

Given the local conditional probabilities stated so far, three particular GPMF models can be designed:

- *Potts-Independent Noise* (P-IN), with local distributions given by Equation 17 and Equation 19, which is a HMF model (Figure 1a). It corresponds to the most popular model of HMF with independent noise [26] [28].
- *Pairwise Markov Field with Uncorrelated Noise* (PMF-UN), with local distributions given by Equation 22 and Equation 19. It belongs to the models depicted by Figure 1c). A similar model has already been studied in [14].
- *Potts-Gaussian Markov Random Field* (P-GMRF), with local distributions given by Equation 17 and Equation 20, which belongs to the HMF-CN family (depicted in Figure 1d). We are not aware of similar models in the literature.

Table 1 summarizes the local distributions of these models. Note that the GPMF (Section 3.1), PMF-UN and P-GMRF models all introduce more dependencies than the P-IN model. PMF-UN enhances the local law for the hidden variables and P-GMRF for the observed variables. Besides, note that the PMF-UN, P-GMRF and P-IN models are submodels of the GPMF in the sense that they have conditional dependencies that are ignored with respect to the GPMF dependencies.

Model	$p(x_s, y_s \vert \boldsymbol{x}_{\mathcal{N}_s} \boldsymbol{y}_{\mathcal{N}_s})$ factorizes using
P-IN	Eq. 17 and Eq. 19
PMF-UN	Eq. 22 and Eq. 19
P-GMRF	Eq. 17 and Eq. 20
GPMF	Eq. 11

Table 1: Factorization of the GPMF submodels.

4. Parameter estimation in the GPMF model

In this section, we develop an algorithm for the unsupervised parameter estimation task in the GPMF model. Without loss of generality, in the following development we consider $K = 2$, i.e. $\Omega = \{\omega_0, \omega_1\}$. Therefore the model is described with 5 parameters. Let $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ be the vector of parameters, then:

$$\begin{aligned} &\boldsymbol{\theta} = \{\mu_{\omega_0}, \mu_{\omega_1}, \beta, \sigma, r\} \triangleq \{\mu_0, \mu_1, \beta, \sigma, r\}, \\ &\text{with } \boldsymbol{\Theta} = \mathbb{R}^2 \times (\mathbb{R}_+^*)^3. \end{aligned} \tag{23}$$

We develop a variation of the Stochastic Expectation Maximization algorithm [9], which we call Stochastic Parameter Estimation (SPE). We first give the statistical estimators of the parameters given the *complete* data $(\boldsymbol{x}, \boldsymbol{y})$.

A generalization of the approach of [17] is established to retrieve the parameter β in the PMF family, using the Least Squares (LS) estimator [22]. This estimation makes use of the probabilities $p(x_s, \boldsymbol{x}_{\mathcal{N}_s^X} | \boldsymbol{y})$, $\forall x_s \in \Omega$. The latter denote the probability of encountering a certain configuration of hidden variables at site s and its neighboring sites. These probabilities are independent of s and are estimated using the frequency estimator.

In the case of the GPMF model we show in Appendix C that:

$$\hat{\beta} = (\boldsymbol{A}^T \boldsymbol{A})^{-1} \boldsymbol{A}^T \boldsymbol{B}, \tag{24}$$

where $\beta \in \mathbb{R}_+^*$ and $\boldsymbol{A}, \boldsymbol{B}$ are real vectors with N elements with generic term, $\forall s \in \mathcal{S}$, $\forall (x_s, x_s') \in \Omega^2$:

$$\begin{aligned} A_s = \ln\left(\frac{p(x_s, \boldsymbol{x}_{\mathcal{N}_s}|\boldsymbol{y})}{p(x_s', \boldsymbol{x}_{\mathcal{N}_s}|\boldsymbol{y})}\right) + \frac{1}{2} Q_{s,s} \bar{y}_s^2 - \frac{1}{2} Q_{s,s} (\bar{y}_s')^2 + \\ \sum_{t \in \mathcal{N}_s} \mathbb{1}_{[t \in \mathcal{N}_s^2]} Q_{s,t} \bar{y}_s \bar{y}_t - \sum_{t \in \mathcal{N}_s} \mathbb{1}_{[t \in \mathcal{N}_s^2]} Q_{t,s} \bar{y}_s' \bar{y}_t, \end{aligned} \tag{25}$$

and

$$\begin{aligned} B_s = \sum_{t \in \mathcal{N}_s} \mathbb{1}_{[t \in \mathcal{N}_s^1]} \Big(2\delta_{x_s}^{x_t} - 2\delta_{x_s'}^{x_t} + (\mu_{x_s'} - \mu_{x_s}) \times \\ (-\bar{y}_s' - \bar{y}_s + 2\bar{y}_t) \Big). \end{aligned} \tag{26}$$

with $\bar{y}_s' = (y_s - \mu_{x_s'})$.

The Maximum Likelihood (ML) estimator is used to estimate μ_0, μ_1, σ. The expressions are, for $i \in \{0, 1\}$:

$$\hat{\mu}_i = \frac{1}{\sum_{s \in \mathcal{S}} \mathbb{1}_{\{x_s = i\}}} \sum_{s \in \mathcal{S}} y_s \mathbb{1}_{\{x_s = i\}}, \tag{27}$$

and

$$\hat{\sigma} = \left(\frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} (y_s - \hat{\mu}_{x_s})^2 \right)^{\frac{1}{2}}. \quad (28)$$

Let us now write an estimator of the range of the exponential correlation function, r, given the complete data $(\boldsymbol{x}, \boldsymbol{y})$. As given in [15][Section 2.4], the correlogram of the field can be estimated by:

$$\hat{C}(d) = \frac{1}{|\mathcal{D}(d)|} \sum_{(s,s') \in N(d)} \frac{1}{\hat{\sigma}^2} (y_s - \hat{\mu}_{x_s})(y_{s'} - \hat{\mu}_{x_{s'}}), \quad (29)$$

where $\mathcal{D}(d)$ is the set of pixel pairs whose Manhattan distance on torus is $d \in \mathbb{N}^+$, *i.e.*:

$$\mathcal{D}(d) = \{(s, s') \in \mathcal{S}^2 : \min(|s_1 - s'_1|, L_1 - |s_1 - s'_1|) + \min(|s_2 - s'_2|, L_2 - |s_2 - s'_2|) = d\}. \quad (30)$$

The exponential correlation function has the form $u = e^{-\frac{d}{r}}$, where u is the correlation value, d the distance (between 0 and N) and r the range to estimate. There is no constant in front of the exponential since the GRF has been standardized, so we will estimate $r \in \mathbb{R}^{+*}$ by fitting the exponential correlation function to the data points $\hat{C}(d)$ [3]. The LS estimator yields:

$$\hat{r} = ((\boldsymbol{E}^T \boldsymbol{E})^{-1} \boldsymbol{E}^T \boldsymbol{w})^{-1}, \quad (31)$$

with vectors of N elements such that: $\boldsymbol{E}^T = (0, \ldots, N)$ and $\boldsymbol{w}^T = (\log \hat{C}(0), \ldots, \log \hat{C}(N))$.

In the context of latent variables, we only possess the observations and not the complete data. Thus the SPE algorithm successively repeats two steps. The first consists in simulating an estimation of the hidden layer, $\hat{\boldsymbol{x}}$, which is then used to form the *completed* data $(\hat{\boldsymbol{x}}, \boldsymbol{y})$, to approximate the complete data. The second step uses the parameter estimators defined previously over the completed data. Algorithm 1 details the steps of the procedure.

SPE is an iterative procedure inspired by the Stochastic Expectation Maximization (SEM) algorithm [9]. SPE differs from SEM because β and r are not estimated using a maximum likelihood estimator.

In SPE, at each iteration t we sample from the posterior law to get a completed data, using Gibbs sampling, run through $K(t)$ iterations. We set $K(t) = t$ as proposed in [8].

Remark: The same estimation procedure can be used to estimate the parameters of the related models described in Section 3.3.

Remark: The computational time of the SPE algorithm is directly linked to the time needed for the posterior samplings since none of the estimator computation is computationally intensive. We have that at iteration t of the SPE algorithm, we have $\frac{t(t+1)}{2}$ Gibbs sampler iterations done. The algorithm is stopped when the estimated parameters reach stability, in practice, we found that ~ 30 SPE iterations are sufficient.

Algorithm 1: The Stochastic Parameter Estimation (SPE) procedure to train the GPMF model.

Data: $\boldsymbol{\theta}^0 = \{\beta^0, \mu_0^0, \mu_1^0, \sigma^0, r^0\}$, the initial set of parameters,
$\boldsymbol{y}$, the observations,
$\hat{\boldsymbol{x}}^0$, the initial configuration for Gibbs sampler.
Result: $\hat{\boldsymbol{\theta}} = \{\hat{\beta}, \hat{\mu}_0, \hat{\mu}_1, \hat{\sigma}, \hat{r}\}$, the estimated parameters.
$t \leftarrow 1$
while *convergence is not attained* **do**
  /* Posterior sampling with a Gibbs sampler initialized at $\hat{\boldsymbol{x}}^{t-1}$ and run during $K(t)$ steps: */
  $\hat{\boldsymbol{x}}^t \sim p(\boldsymbol{x}|\boldsymbol{y}; \boldsymbol{\theta}^{t-1})$
  /* Estimation: */
  • LS estimator for β^t (Eq.24).
  • ML estimator for μ_0^t and μ_1^t (Eq.27).
  • ML estimator for σ^t (Eq.28).
  • Estimation via correlogram for r^t (Eq:31).
  $\boldsymbol{\theta}^t \leftarrow \{\beta^t, \mu_0^t, \mu_1^t, \sigma^t, r^t\}$
  $t \leftarrow t + 1$
end

5. Experiments and Results

This section illustrates the models from the PMF family in the practical task of image segmentation in situations of various complexity. These models are also compared to other methods from the literature of unsupervised image segmentation.

In this section we consider real-world images but also *semi-real* images. The latter are real binary images artificially corrupted by correlated noise, in order to test the relevance and tractability of the new model and its counterparts. The images used come from the "1070-Binary Shape Database" [1]. We want to evaluate the capacity of our model to handle correlated noise, the main purpose of this work, but also to test the parameter estimation procedures with the proposed SPE algorithm (Algorithm 1).

Thus, from the binary image we construct an observation with an additive correlated Gaussian noise over the image. Such a noise is modeled by the conditional likelihood $p(\boldsymbol{y}|\boldsymbol{x})$ of a GMRF. This GMRF has a non-stationary mean $\boldsymbol{\mu_x}$ parametrized by the states of the natural binary image and a stationary covariance matrix $\boldsymbol{\Sigma}$ parametrized by a variance σ^2 and exponential correlation rates of range

[1] https://vision.lems.brown.edu/content/available-software-and-databases

r (as described in Section 3.2). Thus, the artificial images can be considered as sampled from :

$$\boldsymbol{y} \sim p(\boldsymbol{y}|\boldsymbol{x}) = \frac{\exp\left((\boldsymbol{y}-\boldsymbol{\mu_x})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu_x})\right)}{\sqrt{(2\pi)^N \det(\boldsymbol{\Sigma})}}. \tag{32}$$

In all the experiments, we call *error rate* the proportion of misclassified pixels: for an estimate $\hat{\boldsymbol{x}}$ and a ground truth $\boldsymbol{x}$, the error rate e is given by:

$$e = \frac{1}{|\mathcal{S}|}\sum_{s\in\mathcal{S}} \mathbb{1}_{[x_s \neq \hat{x}_s]}. \tag{33}$$

Beforehand, we discuss the efficiency of sampling in the probabilistic models.

5.1. Improved sampling with Tempered-Gibbs sampler

In this section we study more in depth the problematic of sampling in the new distributions. Indeed, it is known that the classic Gibbs sampler procedure (or other MCMC-based sampling approaches) suffers from a poor exploration of the probability distribution. The algorithm is moreover dependent on the initialization of the Gibbs sampler. Running several Gibbs sampler with different random initializations might not be satisfactory as well as being costly. These issues are discussed in [30] which also introduces *parallel tempering*. The idea of parallel tempering is to run in parallel several Gibbs sampler at different temperatures (in the same meaning as in simulated annealing [24]). Samples from Gibbs samplers at high temperature can swap and become the current state of a Gibbs sampler of low temperature. At high temperature the distribution is less severely peaked and the Gibbs sampler explores more easily this distribution. This approach is called the Tempered Gibbs sampler (T-Gibbs sampler).

In our case we use the parallel tempering approach to improve the sampling procedures from the P-GMRF and GPMF models. We develop a similar methodology as [11] where the temperature factor was used to rescale some precise terms of the energy function rather than the whole energy. We then build a tempered energy so that, at high temperature, the probability distribution of the complex model tends towards the distribution of the P-IN model since sampling is easier in the P-IN model. This uses the fact that the PMF models are more general than the P-IN model. To do so we modify the potentials and make them dependent on a temperature parameter that we explicitly add in the notation.

The tempered version of the potentials are given in Appendix D. The T-Gibbs sampler can replace the classic Gibbs sampler in the Marroquin algorithm for MPM computation [29]. In the T-Gibbs algorithm we then proceed to samplings using the tempered versions of the local unnormalized probabilities.

Remark: The tempered Gibbs sampler is clearly much more computationally intensive than the classical Gibbs sampler. This constitutes an important drawback despite the improved performance that will study in the next sections.

5.2. Supervised segmentation on semi-real images with the PMF models

In this section the task of image segmentation is performed with the models from the PMF family in the context of supervised segmentation. The pair of complete data $(\boldsymbol{x}, \boldsymbol{y})$ is then available to estimate $\hat{\boldsymbol{\theta}}$ from it (with the estimators described in Section 4) before estimating $\hat{\boldsymbol{x}}$ using $\boldsymbol{y}$ and $\hat{\boldsymbol{\theta}}$.

The PMF models are here tested in the case of supervised image segmentation according to the two segmentation criteria, MAP and MPM. The MPM will be computed either with Marroquin's algorithm [29] or the T-Gibbs algorithm (Section 5.1). We compute the average error rate for each PMF model in the segmentation of a series of images from the dataset with varying noise level. Recall that the noise is modeled by a GMRF for the semi-real image formation. In the first case, the noise parameter that is varying is $\Delta\mu = |\mu_1 - \mu_0|$, in the second case, the range r of the noise is the variable. The experiment is summed up in Figure 2.

Remark: r represents the scale of the "textures" in the image, while $\Delta\mu$ represents the scale of the pixel intensities. A smaller $\Delta\mu$ means that the image is harder to restore (bigger noise level) but this is not necesarly true for a larger r.

First of all, we note the overall reduced error rate thanks to the new pairwise models over the classical P-IN model. The GPMF model yields the best segmentations for all noise levels. This behaviour illustrates the capacity for the new model to take into account the correlated noise.

Note that in the case of varying range, for all three segmentation algorithms (especially for the MAP), the GPMF and P-GMRF models perform worse than the P-IN and PMF-UN models at the smallest correlation ranges. This leads us to believe that in such situations the models which can handle correlated noise are prone to overfitting. Indeed, at fixed parameters, the GPMF (resp. P-GMRF) models is equivalent to the PMF-UN (resp. P-IN) model expect for the term which models the correlated noise (last term of Equation 8). Thus we would except the error rates of these equivalent to finally become similar for the smallest ranges. It is, however not happening, and might be due to overfitting for the more complex models. This issue is, however, unlikely to appear in practical cases when the noise is known to be correlated.

5.3. Unsupervised segmentation on semi-real images

In this section, we address the problem of unsupervised image segmentation. We use the semi-real images

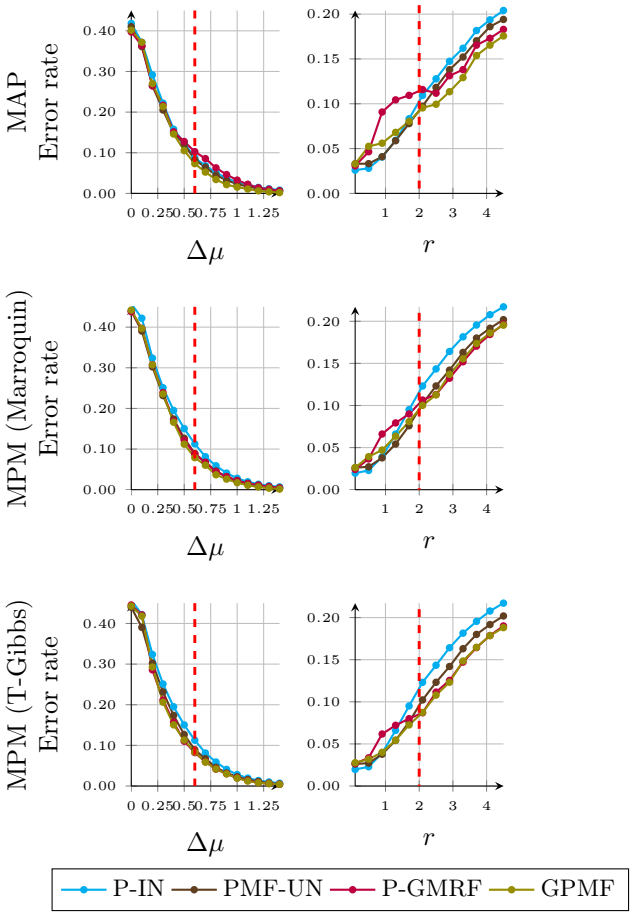


Figure 2: Error rate in function of varying model parameters and of several estimators in the supervised segmentation of the 'dude' images of the dataset. Each line, from top to bottom, respectively corresponds to the MAP, MPM (with Marroquin's algorithm) and MPM (with T-Gibbs algorithm) segmentation criteria. Each column, from left to right, respectively corresponds to $\Delta\mu$ variable and r variable. For the $\Delta\mu$ column, the other parameters of the GRF noise were fixed to $\sigma = 0.5$ and $r = 2$. For the r column, they were fixed to $\mu_0 = 0, \mu_1 = 0.6$ and $\sigma = 0.5$. The dashed red lines are common between the graphs on each line. In such case we notice that the models have identical performances relative to each other. Each point of the curves is the average error rate (Equation 33) for the segmentation of the 30 'dude' images of the dataset, repeated 10 times.

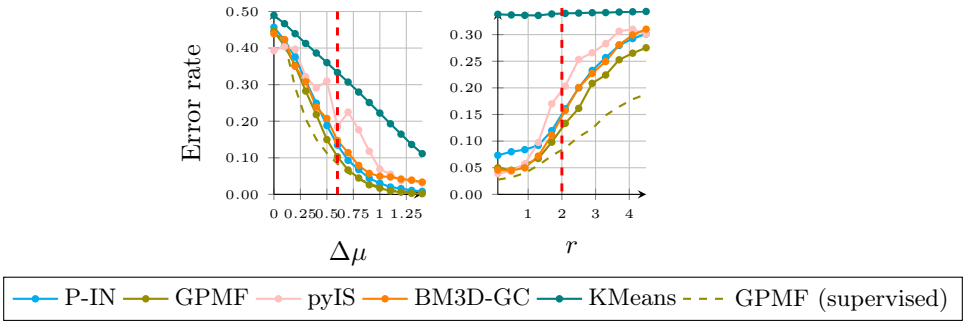


Figure 3: MPM (T-Gibbs) unsupervised segmentation over the 'dude' images. For the $\Delta\mu$ column, the other parameters of the GRF noise were fixed to $\sigma = 0.5$ and $r = 2$. For the r column, they were fixed to $\mu_0 = 0, \mu_1 = 0.6$ and $\sigma = 0.5$. The dashed vertical red lines highlight parameter the situation when the parameter configuration is identical. In such case we notice that the models have identical performances relative to each other. Tha dashed green lines reproduce the results of the GPMF model from the supervised case experiment (last row of Figure 2). Each point of the curves is the average error rate (Equation 33) for the segmentation of the 30 'dude' images ofthe dataset, repeated 10 times.

previously introduced but consider only the observations $\boldsymbol{Y}$. We here only consider the GPMF model and the MPM (T-Gibbs) segmentation criterion which were the best performing approaches of the previous section. To compare with the GPMF model, we used 3 approaches in unsupervised image segmentation:

- The classical Hidden Markov Field model with Potts prior and Independent Noise (P-IN) [28] [26], which is also a the simplest member of the PMF model family (see Section 3.3).
- The KMeans clustering[2] algorithm [1].
- The pyImSegm[3] (pyIS) segmentation algorithm proposed in [5]. The core of this technique also relies on a Markov Random Field energy minimization problem using super-pixel based and graph-cut based approaches.
- Based on the recent developments of [36] for correlated noise reduction (which were not available online), we propose to combine the Block-Matching 3D filter[4] from [16] and the graph-cut algorithm[5] from [6]. This approach is then the combination of a popular noise reduction technique and a popular graph-cut based segmentation. We call this approach BM3D+GC.

Let us now compare the unsupervised segmentations obtained with the GPMF model with the results given by other models. Figure 3 illustrates the segmentation performance of the models for a varying noise level. Moreover, since the settings of the supervised experiment (Section 5.2) and of this unsupervised experiment are comparable, we reproduced the result of the GPMF model in the supervised case. By comparing the GPMF model performances in both cases we can see the stable degradation of the performance in the unsupervised case, which illustrates the robustness of the SPE method.

Figure 4 depicts some segmentations of images from the database. We notice, in all cases, the superiority of the new probabilistic model GPMF in its capacity to handle the correlated noise, improving the segmentation error and giving stable results. The GPMF model always performs best or equally best at all noise level. In the best scenarii, the GPMF model increases the average error rate by around 4 points. The graph-cut approaches, BM3D+GC and pyIS, are more unstable and perform worse when dealing with an image corrupted with correlated noise, despite manually tuned hyperparameters. Finally, note that, as expected, the overall error rates are higher in this unsupervised segmentation experiment than in the supervised segmentation experiment of Section 5.2.

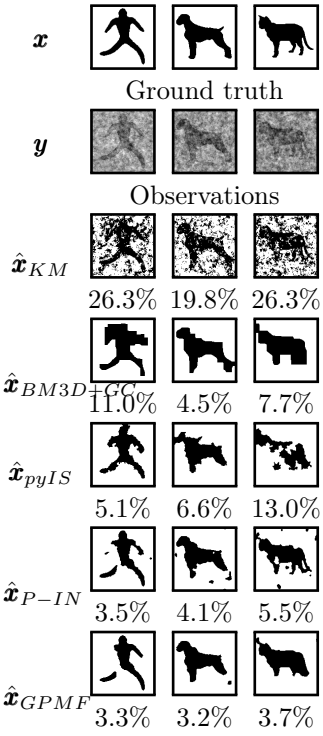


Figure 4: Unsupervised segmentation of images from the dataset with the new model and 4 other models. The error rates with respect to the ground truth appear below each sample. For the 3 images, the GMF parameters are $\mu_0 = 0$, $\mu_1 = 2$, $\sigma = 1$, $r = 3$. The segmentation is here done under the MPM criterion computed with Marroquin's algorithm [29]: it is the best compromise between segmentation performance and computation time as explained in our study.

[2]Implementation from `https://opencv.org`.
[3]Implementation from `https://github.com/Borda/pyImSegm`.
[4]Implementation from `http://www.cs.tut.fi/~foi/GCF-BM3D/`.
[5]Implementation from `https://github.com/shaibagon/GCMex/`.

5.4. On real world images

We now present the model in a real world application from the medical field, where unsupervised image segmentation needs to be done on strongly spatially corrupted data.

We consider micro-computed tomography X-rays scans [20] of human arteries containing a metallic biomaterial (called a *stent*). The scarcity of this data obliges the use of an unsupervised approach. These images are very noisy because of the beam hardening artifacts caused by the interactions between X-rays and the metallic stent. To improve the design of biomaterials and their implantation in the human body, it is crucial to analyze, *in situ*, the biomaterial when it fails and need to be explanted from the patient body. In such process of analysis, we need to precisely segment stent, organic material and background despite the strong noise in the image. In this image processing problem, the stent beam hardening artifacts can be modeled as a correlated noise. An automatization of the segmentation process could help processing more data and create enhanced inputs for biomechanical studies carried to increase the knowledge about the vascular diseases.

Some works have been proposed using different techniques for stent artifact reduction, see for example, [10] [19]. However these works do not propose a segmentation step and they often rely on physical additional data about the acquisition technique. Moreover, to the best of our knowledge, there is no counterpart of our model in the literature involving spatially-correlated noise. Thus, a comparison with the techniques developed for this precise application is out of the scope of this article.

The goal of the experiment is to segment precisely the organic material from the background of the images. The two classes we wish to distinguish are visible by the naked eye, but they are very corrupted by the artifacts and are challenging to segment using automatic unsupervised algorithms. We treat 512×512-pixel 2D images.

Figure 5 depicts the results obtained by three methods described in the previous section (BM3D+GC, P-IN and GPMF). While the overall error rate is in favor of the GPMF model, it does not truly reflects the capacity of the model to resolve correlated noise and offer a proper segmentation. Therefore we consider the organic material as the *true* class and we compute the False Negative (FN) and False Positive (FP) percentage of pixels in areas around the stent where the correlations are the strongest. We can see that the GPMF model best captures the correlated noise and resolve much of the stent hardening beam problem: it offers the best compromise between the FN and FP scores. On the other hand, the KMeans, P-IN and, to a smaller extent, the BM3D+GC results are particularly prone to misclassifications because of the artifacts. The results of the pyIS method were essentially similar to that of BM3D+GC and are omitted here for brevity, as well as the results from the KMeans algorithm which are omitted because of the poor performance of the algorithm studied in Section 5.3. In addition the results obtained with the GPMF model were also judged very satisfying by the pathologists.

6. Conclusion

In this article, we presented a new kind of probabilistic latent variable model called Gaussian Pairwise Markov Fields. We established its formal definition and considered fours models that are GPMF to address the problem of unsupervised image segmentation corrupted with correlated noise. We showed in theory and in practice that these models are generalizations of the classical hidden Markov field model. We exhibited important links between this extension of pairwise Markov fields and the Gaussian fields theory. We also developed an efficient parameter estimation procedures to answer the problem of unsupervised image segmentation. In this setting, the GPMF model which was able to model the more dependencies was the best performing method to segment images in presence of correlated noise on the considered dataset. The same conclusion was true considering real world data, with convincing results in a medical imaging application.

Future work will consider modeling the non-stationarities in the observed process, which seems to be the major limitation in the currrent model. To this end, works have been carried directly in the modelization of the GRF, through a complexification of the correlation function that we might investigate [27] [21] [31]. Another approach would be that of building a "superposition" of GRFs [12] [39]. Improvements of the model in these directions might lead to better performances, especially on real world images.

Acknowledgement

The authors want to thank Pr. Nabil Chakfé (head of Vascular Surgery and Kidney Transplantation Department and GEPROVAS, Strasbourg, France) for supporting this work and Salomé Kuntz, Matthew D. Kutyna and Dr. Renu Virmani from CVPath Institute Inc. (Gaithersburg, Maryland, USA) for their collaboration to this work and for the acquisition of the micro-computed tomography images.

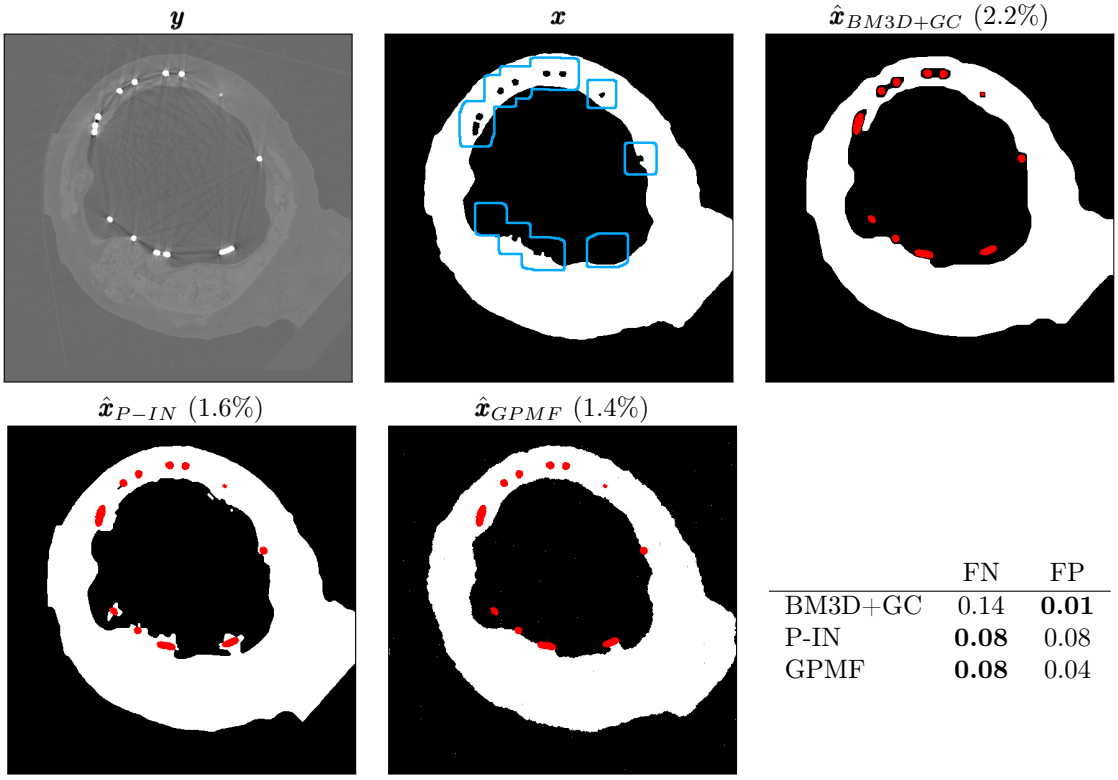
References

[1] David Arthur and Sergei Vassilvitskii. K-means++: The advantages of careful seeding. In *Proceedings of the 18th ACM-SIAM symposium on Discrete algorithms*, pages 1027–1035, 2007.

[2] Dalila Benboudjema and Wojciech Pieczynski. Unsupervised statistical segmentation of nonstationary images using triplet Markov fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(8):1367–1378, 2007.

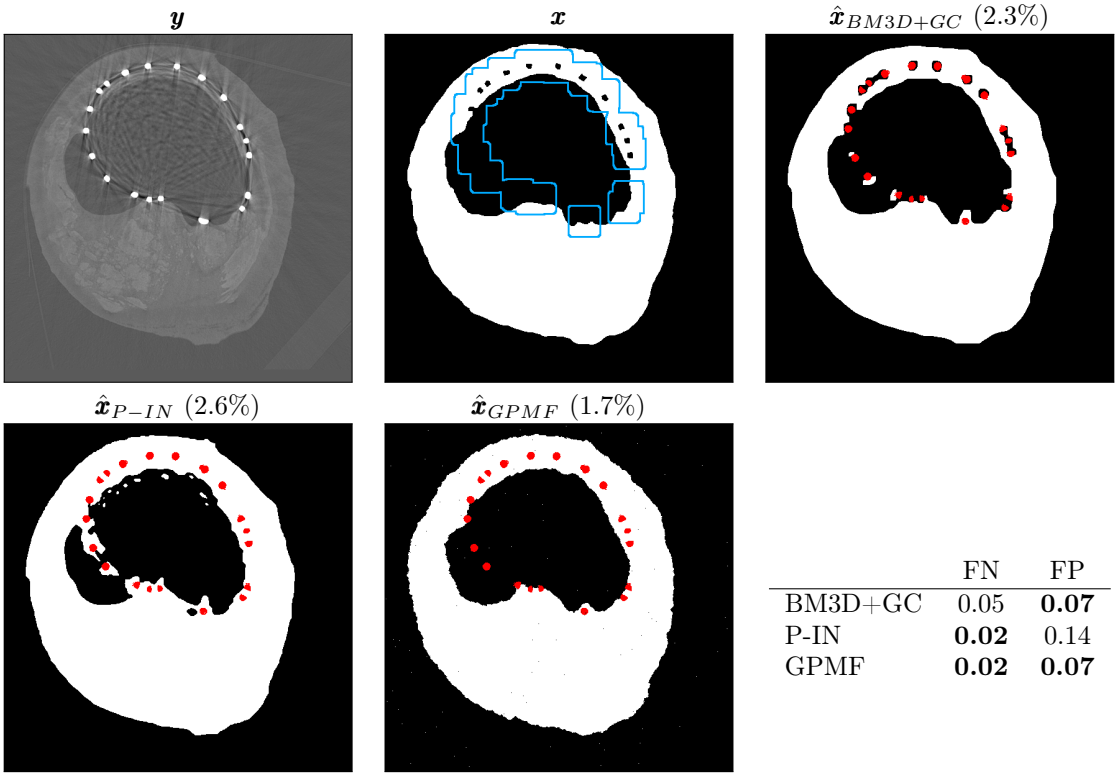
[3] Ottar N Bjørnstad and Wilhelm Falck. Nonparametric spatial covariance functions: estimation and testing. *Environmental and Ecological Statistics*, 8(1):53–70, 2001.

[4] Andrew Blake, Pushmeet Kohli, and Carsten Rother. *Markov Random Fields for Vision and Image Processing*. The MIT Press, 2011. ISBN 0262015773, 9780262015776.



	FN	FP
BM3D+GC	0.14	**0.01**
P-IN	**0.08**	0.08
GPMF	**0.08**	0.04

(a) Case 1



	FN	FP
BM3D+GC	0.05	**0.07**
P-IN	**0.02**	0.14
GPMF	**0.02**	**0.07**

(b) Case 2

Figure 5: Unsupervised segmentations of organic material in corrupted X-rays images. For these 2 cases, which illustrate different explanted biomaterials, we have: the real image $\boldsymbol{y}$, the ground truth image $\boldsymbol{x}$, the BM3D-GC segmentation, the P-IN segmentation and the GPMF segmentation. The segmentation criterion for the probabilistic models is the MPM (Marroquin). The stent parts (brightest pixels in $\boldsymbol{y}$) were segmented beforehand by a thresholding technique and then considered as image borders. They appear in red on the segmented images and did not take part in the segmentation. The ground truth could be provided by experts since an histological analysis is available. Error rates with respect to the ground truth on the whole image appear in parenthesis. FN and FP rates computed in the blue areas for each model appear in the table.

[5] Jiří Borovec, Jan Švihlík, Jan Kybic, and David Habart. Supervised and unsupervised segmentation using superpixels, model estimation, and graph cut. *Journal of Electronic Imaging*, 26 (6):061610, 2017.

[6] Yuri Boykov and Vladimir Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 26(9):1124–1137, September 2004.

[7] D. Andraw Brown, Christopher S. McMahan, and Stella Watson. Sampling strategies for fast updating of Gaussian Markov random fields. *The American Statistician*, (just-accepted):1–32, 2019.

[8] Miguel A Carreira-Perpinan and Geoffrey E Hinton. On contrastive divergence learning. In *AISTATS*, volume 10, pages 33–40. Citeseer, 2005.

[9] Gilles Celeux. The sem algorithm: a probabilistic teacher algorithm derived from the em algorithm for the mixture problem. *Computational statistics quarterly*, 2:73–82, 1985.

[10] Yang Chen, Yinsheng Li, Hong Guo, Yining Hu, Limin Luo, Xindao Yin, Jianping Gu, and Christine Toumoulin. CT metal artifact reduction method based on improved image segmentation and sinogram in-painting. *Mathematical Problems in Engineering*, 2012, 2012.

[11] KyungHyun Cho, Tapani Raiko, and Alexander Ilin. Parallel tempering is efficient for learning restricted boltzmann machines. In *The 2010 International Joint Conference on Neural Networks*, pages 1–8. IEEE, 2010.

[12] Jean-Baptiste Courbot, Emmanuel Monfrini, Vincent Mazet, and Christophe Collet. Oriented triplet Markov fields. *Pattern Recognition Letters*, 103:16–22, 2018.

[13] Jean-Baptiste Courbot, Emmanuel Monfrini, Vincent Mazet, and Christophe Collet. Triplet Markov trees for image segmentation. In *2018 IEEE Statistical Signal Processing Workshop (SSP)*, pages 233–237. IEEE, 2018.

[14] Jean-Baptiste Courbot, Vincent Mazet, Emmanuel Monfrini, and Christophe Collet. Pairwise Markov fields for segmentation in astronomical hyperspectral images. *Signal Processing*, 163: 41 – 48, 2019.

[15] Noel Cressie. *Statistics for spatial data*. Wiley Online Library, 1992.

[16] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. BM3D image denoising with shape-adaptive principal component analysis. 2009.

[17] Haluk Derin and Howard Elliott. Modeling and segmentation of noisy and textured images using Gibbs random fields. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (1): 39–55, 1987.

[18] Tamara Dimitrova and Ljupco Kocarev. Graphical models over heterogeneous domains and for multilevel networks. *IEEE Access*, 6:69682–69701, 2018.

[19] Elena Faggiano, Tommaso Lorenzi, and Alfio Quarteroni. Metal artefact reduction in computed tomography images by a fourth-order total variation flow. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 4(3-4): 202–213, 2016.

[20] Brian P Flannery, Harry W Deckman, Wayne G Roberge, and Kevin L D'Amico. Three-dimensional x-ray microtomography. *Science*, 237(4821):1439–1444, 1987.

[21] Francky Fouedjio, Nicolas Desassis, and Thomas Romary. Estimation of space deformation model for non-stationary random functions. *Spatial Statistics*, 13:45–61, 2015.

[22] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. *The elements of statistical learning*, volume 1. Springer series in statistics New York, NY, USA:, 2001.

[23] Geir-Arne Fuglstad, Finn Lindgren, Daniel Simpson, and Håvard Rue. Exploring a new class of non-stationary spatial Gaussian random fields with varying local anisotropy. *Statistica Sinica*, pages 115–133, 2015.

[24] Stuart Geman and Donald Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, (6): 721–741, 1984.

[25] Ivan Gorynin, Hugo Gangloff, Emmanuel Monfrini, and Wojciech Pieczynski. Assessing the segmentation performance of pairwise and triplet Markov models. *Signal Processing*, 145: 183–192, 2018.

[26] Zoltan Kato, Josiane Zerubia, et al. Markov random fields in image segmentation. *Foundations and Trends® in Signal Processing*, 5(1–2):1–155, 2012.

[27] William Kleiber. High resolution simulation of nonstationary Gaussian random fields. *Computational Statistics & Data Analysis*, 101:277–288, 2016.

[28] Stan Z Li. *Markov random field modeling in image analysis*. Springer Science & Business Media, 2009.

[29] Jose Marroquin, Sanjoy Mitter, and Tomaso Poggio. Probabilistic solution of ill-posed problems in computational vision. *Journal of the american statistical association*, 82(397):76–89, 1987.

[30] Radford M Neal. Sampling from multimodal distributions using tempered transitions. *Statistics and computing*, 6(4):353–366, 1996.

[31] Douglas Nychka, Dorit Hammerling, Mitchell Krock, and Ashton Wiens. Modeling and emulation of nonstationary Gaussian fields. *Spatial statistics*, 28:21–38, 2018.

[32] Youngsuk Park, David Hallac, Stephen Boyd, and Jure Leskovec. Learning the network structure of heterogeneous data via pairwise exponential Markov random fields. *Proceedings of machine learning research*, 54:1302, 2017.

[33] Andrija Petrović, Mladen Nikolić, Miloš Jovanović, and Boris Delibašić. Gaussian conditional random fields for classification. *arXiv preprint arXiv:1902.00045*, 2019.

[34] Wojciech Pieczynski and Abdel-Nasser Tebbache. Pairwise Markov random fields and segmentation of textured images. *Machine graphics and vision*, 9(3):705–718, 2000.

[35] Vladan Radosavljevic, Slobodan Vucetic, and Zoran Obradovic. Continuous conditional random fields for regression in remote sensing. In *Proceedings of the 19th European Conference on Artificial Intelligence*, pages 809–814, 2010.

[36] Oleksii Rubel, Vladimir Lukin, and Karen Egiazarian. Additive spatially correlated noise suppression by robust block matching and adaptive 3D filtering. *Journal of Imaging Science and Technology*, 62(6):60401–1, 2018.

[37] Håvard Rue and Leonhard Held. *Gaussian Markov random fields: theory and applications*. CRC press, 2005.

[38] Ruslan R Salakhutdinov. Learning in markov random fields using tempered transitions. In *Advances in neural information processing systems*, pages 1598–1606, 2009.

[39] Cornelia Vacar and Jean-François Giovannelli. Unsupervised joint deconvolution and segmentation method for textured images: a Bayesian approach and an advanced sampling algorithm. *EURASIP Journal on Applied Signal Processing*, 2019:17, December 2019. doi: 10.1186/s13634-018-0597-x.

[40] Raviteja Vemulapalli, Oncel Tuzel, Ming-Yu Liu, and Rama Chellapa. Gaussian conditional random field network for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3224–3233, 2016.

[41] Hongyuan Zhu, Fanman Meng, Jianfei Cai, and Shijian Lu. Beyond pixels: A comprehensive survey from bottom-up to semantic image segmentation and cosegmentation. *Journal of Visual Communication and Image Representation*, 34:12–27, 2016.

Appendix A. Proof of the GPMF definition

Necessity: We now show that if $(\boldsymbol{X}, \boldsymbol{Y})$ is a GPMF, then its Gibbs distribution is necessarly of the form of Equation 7. If $(\boldsymbol{X}, \boldsymbol{Y})$ is a GPMF with respect to $\mathcal{N}$, it is then a PMF with repect to the same neighborhood and, thanks to the Hammersley-Clifford theorem we can write

$$-\sum_{s,s'\in\mathcal{S}^2} y_s C_{s,s'} y_{s'} - \log\sqrt{2\pi\det(\boldsymbol{C}^{-1})} = \sum_{n=1}^{|\mathcal{N}|}\Big(\sum_{\boldsymbol{c}\in\mathcal{C}_n} V_n(\boldsymbol{x_c},\boldsymbol{y_c})\Big) - \log\int_{\mathbb{R}^N} \mathrm{d}\boldsymbol{y}\exp\Big(-\sum_{n=1}^{|\mathcal{N}|}\Big(\sum_{\boldsymbol{c}\in\mathcal{C}_n} V_n(\boldsymbol{x_c},\boldsymbol{y_c})\Big)\Big). \quad (A.1)$$

$$-\left(\sum_{\boldsymbol{c}\in\mathcal{C}_1}\bar{V}_1(\boldsymbol{y_c},\boldsymbol{x_c}) + \sum_{\boldsymbol{c}\in\mathcal{C}_2}\bar{V}_2(\boldsymbol{y_c},\boldsymbol{x_c})\right) - \log\int_{\mathbb{R}^N}\mathrm{d}\boldsymbol{y}\exp\left(\sum_{\boldsymbol{c}\in\mathcal{C}_1}\bar{V}_1(\boldsymbol{y_c},\boldsymbol{x_c}) + \sum_{\boldsymbol{c}\in\mathcal{C}_2}\bar{V}_2(\boldsymbol{y_c},\boldsymbol{x_c})\right) = -\sum_{n=1}^{|\mathcal{N}|}\Big(\sum_{\boldsymbol{c}\in\mathcal{C}_n} V_n(\boldsymbol{x_c},\boldsymbol{y_c})\Big) - \log\int_{\mathbb{R}^N}\mathrm{d}\boldsymbol{y}\exp\Big(-\sum_{n=1}^{|\mathcal{N}|}\Big(\sum_{\boldsymbol{c}\in\mathcal{C}_n} V_n(\boldsymbol{x_c},\boldsymbol{y_c})\Big)\Big). \quad (A.2)$$

$$\underbrace{-\left(\sum_{\boldsymbol{c}\in\mathcal{C}_1}\bar{V}_1(\boldsymbol{y_c},\boldsymbol{x_c}) + \sum_{\boldsymbol{c}\in\mathcal{C}_2}\bar{V}_2(\boldsymbol{y_c},\boldsymbol{x_c})\right)}_{A} + \underbrace{\sum_{n=1}^{|\mathcal{N}|}\Big(\sum_{\boldsymbol{c}\in\mathcal{C}_n} V_n(\boldsymbol{x_c},\boldsymbol{y_c})\Big)}_{B} = -\log\int_{\mathbb{R}^N}\mathrm{d}\boldsymbol{y}\exp\Big(-\sum_{n=1}^{|\mathcal{N}|}\Big(\sum_{\boldsymbol{c}\in\mathcal{C}_n} V_n(\boldsymbol{x_c},\boldsymbol{y_c})\Big)\Big) + \log\int_{\mathbb{R}^N}\mathrm{d}\boldsymbol{y}\exp\left(\sum_{\boldsymbol{c}\in\mathcal{C}_1}\bar{V}_1(\boldsymbol{y_c},\boldsymbol{x_c}) + \sum_{\boldsymbol{c}\in\mathcal{C}_2}\bar{V}_2(\boldsymbol{y_c},\boldsymbol{x_c})\right). \quad (A.3)$$

that:

$$p(\boldsymbol{x},\boldsymbol{y}) = \frac{1}{Z}\exp\Big(-\sum_{n=1}^{|\mathcal{N}|}\Big(\sum_{\boldsymbol{c}\in\mathcal{C}_n} V_n(\boldsymbol{x_c},\boldsymbol{y_c})\Big)\Big), \quad (A.4)$$

On the other hand, if we want to meet the second condition of the GPMF definition, we need to ensure that $p(\boldsymbol{y}|\boldsymbol{x})$ is the density of a multivariate Gaussian function. Thus, there exists a SPD matrix $\boldsymbol{C}$ such that:

$$p(\boldsymbol{y}|\boldsymbol{x}) = \frac{1}{\sqrt{2\pi\det(\boldsymbol{C}^{-1})}}\exp\left(-\sum_{s,s'\in\mathcal{S}^2} y_s C_{s,s'} y_{s'}\right). \quad (A.5)$$

Then, using:

$$p(\boldsymbol{y}|\boldsymbol{x}) = \frac{p(\boldsymbol{x},\boldsymbol{y})}{\int_{\mathbb{R}^N}\mathrm{d}\boldsymbol{y}p(\boldsymbol{x},\boldsymbol{y})}, \quad (A.6)$$

we get Equation A.1 by Equations A.4 and A.5. Now note that since $\boldsymbol{C}$ is SPD, $\sum_{(s,s')\in\mathcal{S}^2} y_s C_{s,s'} y_{s'}$ can be written as a positive semidefinite quadratic form in the variables $\boldsymbol{y}$, such that:

$$\sum_{(s,s')\in\mathcal{S}^2} y_s C_{s,s'} y_{s'} = \sum_{\boldsymbol{c}\in\mathcal{C}_1}\bar{V}_1(\boldsymbol{y_c},\boldsymbol{x_c}) + \sum_{\boldsymbol{c}\in\mathcal{C}_2}\bar{V}_2(\boldsymbol{y_c},\boldsymbol{x_c}). \quad (A.7)$$

$\bar{V}_1$ and $\bar{V}_2$ are polynomial function of $\boldsymbol{y}$ variables where terms involving a mix of $\boldsymbol{y}$ and $\boldsymbol{x}$ variables can be found, but terms with $\boldsymbol{x}$ variables alone cannot be found. Moreover there is the constraint that the RHS of Equation A.7 is a positive semidefinite quadratic form in the $\boldsymbol{y}$ variables.

Using Equation A.7 as well as the result of the multivariate Gaussian integral "backwards", Equation A.1 becomes Equation A.2. We finally rearrange Equation A.2 into Equation A.3. In Equation A.3 it is clear that the RHS does not depend on $\boldsymbol{y}$ then so does the LHS. Thus, in the LHS, the terms containg $\boldsymbol{y}$ variables in A must simplify with terms in B. Therefore, we are able to deduce the constraints on the terms of B that we are looking for:

- $\sum_{\boldsymbol{c}\in\mathcal{C}_1} V_1(\boldsymbol{x_c},\boldsymbol{y_c}) = \sum_{\boldsymbol{c}\in\mathcal{C}_1}\Big(\bar{V}_1(\boldsymbol{y_c},\boldsymbol{x_c}) + \tilde{V}_1(\boldsymbol{x_c})\Big)$,
- $\sum_{\boldsymbol{c}\in\mathcal{C}_2} V_2(\boldsymbol{x_c},\boldsymbol{y_c}) = \sum_{\boldsymbol{c}\in\mathcal{C}_2}\Big(\bar{V}_2(\boldsymbol{y_c},\boldsymbol{x_c}) + \tilde{V}_2(\boldsymbol{x_c})\Big)$,
- all the other terms in B does not involve $\boldsymbol{y}$ variables.

We introduced $\tilde{V}_1$ and $\tilde{V}_2$ to be potential functions which only involve $\boldsymbol{x}$ variables.

This concludes the necessity part of the demonstration.

Sufficiency: Let us first show that a factorization of the form of Equation 2 using Equation 8 or, equivalently, that a Gibbs distribution whose energy is given by Equation 7 is a PMF. It is straightforward to see that $p(\boldsymbol{x},\boldsymbol{y}) > 0, \forall\boldsymbol{x}\in\Omega^N, \forall\boldsymbol{y}\in\mathbb{R}^N$. Moreover, from the conditional formulation we have clearly that, $\forall s\in\mathcal{S}$, $p(x_s,y_s|\boldsymbol{x}_{\mathcal{S}\setminus s},\boldsymbol{y}_{\mathcal{S}\setminus s}) = p(x_s,y_s|\boldsymbol{x}_{\mathcal{N}_s},\boldsymbol{y}_{\mathcal{N}_s})$. These first two points show that $(\boldsymbol{X},\boldsymbol{Y})$ is a PMF with respect to $\mathcal{N}$.

We now need to show that $p(\boldsymbol{y}|\boldsymbol{x})$ is the density of a GMRF. Using the energy formulation, we show using Bayes theorem that:

$$p(\boldsymbol{y}|\boldsymbol{x}) = \frac{\exp\Big(-\Big(\sum_{n=1}^{2}\big(\sum_{\boldsymbol{c}\in\mathcal{C}_n}\bar{V}_n(\boldsymbol{x_c},\boldsymbol{y_c})\big)\Big)\Big)}{\int_{\mathbb{R}^N}\mathrm{d}\boldsymbol{y}\exp\Big(-\Big(\sum_{n=1}^{2}\big(\sum_{\boldsymbol{c}\in\mathcal{C}_n}\bar{V}_n(\boldsymbol{x_c},\boldsymbol{y_c})\big)\Big)\Big)}. \quad (A.8)$$

In the previous equations, we remove all the V_n where the $\boldsymbol{y_c}$ variables do not play a role. The remaining terms are $\bar{V}_1$ and $\bar{V}_2$, which are assumed to be positive semidefinite forms in $\boldsymbol{y_c}$. Thus, by definition, there exists $\boldsymbol{A}$ and $\boldsymbol{B}$ SPD matrices of size $N\times N$ such that:

$$\sum_{n=1}^{2}\sum_{\boldsymbol{c}\in\mathcal{C}_n}\bar{V}_n(\boldsymbol{x_c},\boldsymbol{y_c}) = \sum_{s\in\mathcal{S}} A_{s,s}y_s^2 + \sum_{s\in\mathcal{S}}\sum_{s'\in\mathcal{N}_s} B_{s,s'}y_s y_{s'}, = \sum_{s\in\mathcal{S}}\sum_{s'\in\mathcal{N}_s\cup\{s\}} C_{s,s'}y_s y_{s'}. \quad (A.9)$$

Note that $\boldsymbol{A}$ and $\boldsymbol{B}$ can depend on $\boldsymbol{x}$ but we omit it in notations for clarity. Note also that $\boldsymbol{A}$ is a diagonal matrix. Since $\boldsymbol{C} = \boldsymbol{A} + \boldsymbol{B}$, $\boldsymbol{C}$ is a SPD matrix since it is the sum of two SPD matrices. Finally, using the result of the multivariate Gaussian integral we find Equation 13. The latter equation shows that $\boldsymbol{Y}$ given $\boldsymbol{X} = \boldsymbol{x}$ is a GMRF, which is the second element of the GPMF definition and concludes the sufficiency part of the demonstration.

Appendix B. Proof that the model is a GPMF

In this section we show that the model defined in Section 3.1, by the energy of Equation 9, or equivalently, the unnormalized local conditional probabilities of Equation 11, is a GPMF.

First, it is clear that $\forall \boldsymbol{x} \in \Omega^N, \forall \boldsymbol{y} \in \mathbb{R}^N, p(\boldsymbol{x}, \boldsymbol{y}) > 0$. Along with Equation 11 wihch attests the Markovian property, we can conclude that (X, Y) is a PMF.

We now need to show that $\boldsymbol{Y}$ given a realization $\boldsymbol{x}$ of $\boldsymbol{X}$ is a GMRF. The model energy can be written:

$$E(\boldsymbol{x}, \boldsymbol{y}) = \sum_{s \in \mathcal{S}} \sum_{\substack{s' \in \\ \mathcal{N}_s \cup \{s\}}} \left[- \mathbb{I}_{[s' \in \mathcal{N}_s^1]} \delta_{x_s}^{x_{s'}} \beta \Big(1 - \frac{1}{2}(\bar{y}_s - \bar{y}_{s'})^2\Big)\right] + \sum_{s \in \mathcal{S}} \sum_{\substack{s' \in \\ \mathcal{N}_s \cup \{s\}}} \left[\frac{1}{2} \mathbb{I}_{[s' \in \mathcal{N}_s^2]} Q_{s,s'} \bar{y}_s \bar{y}_{s'}\right] - \beta \Big(1 - \frac{1}{2}(\bar{y}_s - \bar{y}_s)^2\Big), \tag{B.1}$$

which gives:

$$E(\boldsymbol{x}, \boldsymbol{y}) = \sum_{s \in \mathcal{S}} \sum_{\substack{s' \in \\ \mathcal{N}_s^1 \cup \{s\}}} -\delta_{x_s}^{x_{s'}} \beta + \sum_{s \in \mathcal{S}} \sum_{\substack{s' \in \\ \mathcal{N}_s^2 \cup \{s\}}} \left[\frac{1}{2} Q_{s,s'} \bar{y}_s \bar{y}_{s'}\right] + \sum_{s \in \mathcal{S}} \sum_{\substack{s' \in \\ \mathcal{N}_s^1 \cup \{s\}}} \left[\frac{1}{2} \delta_{x_s}^{x_{s'}} \beta (\bar{y}_s - \bar{y}_{s'})^2\right] - \beta. \tag{B.2}$$

A detailed derivation can show that:

$$\sum_{s \in \mathcal{S}} \sum_{\substack{s' \in \\ \mathcal{N}_s^1 \cup \{s\}}} \frac{1}{2} \delta_{x_s}^{x_{s'}} \beta (\bar{y}_s - \bar{y}_{s'})^2 = \frac{1}{2} \bar{\boldsymbol{y}}^T P \bar{\boldsymbol{y}}, \tag{B.3}$$

where $\boldsymbol{P}$ is a matrix with elements:

$$P_{s,s'} = \begin{cases} 2 \sum_{s' \in \mathcal{N}_s^1} \delta_{x_s}^{x_{s'}} \beta, \text{ if } s = s', \\ -2 \delta_{x_s}^{x_{s'}} \beta, \text{ if } s' \in \mathcal{N}_s^1, \\ 0 \text{ otherwise.} \end{cases} \tag{B.4}$$

Then note that, since with restrict $\beta \in \mathbb{R}_+^*$, $\forall s \in \mathcal{S}$:

$$|P_{s,s}| \geq \sum_{\substack{s' \in \mathcal{S} \\ s' \neq s}} |P_{s,s'}|. \tag{B.5}$$

Thus $\boldsymbol{P}$ has the diagonal dominance property, which makes $\boldsymbol{P}$ a SPD matrix. In the model definition, we also restrict $\boldsymbol{Q}$ to be a SPD matrix with the constraint of Equation 12, then, $\boldsymbol{R} = \boldsymbol{P} + \boldsymbol{Q}$ is a SPD matrix. The energy can then be written:

$$\begin{aligned} E(\boldsymbol{x}, \boldsymbol{y}) &= \sum_{(s,s') \in \mathcal{S}^2} \left[\frac{1}{2} P_{s,s'} \bar{y}_s \bar{y}_{s'}\right] + \sum_{(s,s') \in \mathcal{S}^2} \left[\frac{1}{2} Q_{s,s'} \bar{y}_s \bar{y}_{s'}\right] - \sum_{s \in \mathcal{S}} \sum_{\substack{s' \in \\ \mathcal{N}_s^1 \cup \{s\}}} \delta_{x_s}^{x_{s'}} \beta \\ &= \sum_{(s,s') \in \mathcal{S}^2} \left[\frac{1}{2} R_{s,s'} \bar{y} \bar{y}_{s'}\right] - \sum_{s \in \mathcal{S}} \sum_{\substack{s' \in \\ \mathcal{N}_s^1 \cup \{s\}}} \delta_{x_s}^{x_{s'}} \beta. \end{aligned} \tag{B.6}$$

Using the result of the integral of the multivariate Gaussian we finally have:

$$p(\boldsymbol{y}|\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^N \det(\boldsymbol{R}^{-1})}} \exp\left(-\frac{1}{2} \bar{\boldsymbol{y}}^T \boldsymbol{R} \bar{\boldsymbol{y}}\right), \tag{B.7}$$

which is the expression of a Gaussian Markov Random Field (GMRF) with non stationary mean and precision matrix $\boldsymbol{R} = \boldsymbol{P} + \boldsymbol{Q}$.

This concludes the proof that the model defined in Section 3.1 is a GPMF.

Appendix C. Least-Square estimator for β

In this section we show how to adapt the Linear Least-Square estimator [17] for the β parameter, using the a completed pair of realizations of $(\boldsymbol{x}, \boldsymbol{y})$. The derivation is done for the GPMF model and is similar for the other models.

First note that, $\forall s \in \mathcal{S}$:

$$\begin{aligned} \frac{p(x_s, \boldsymbol{x}_{\mathcal{N}_s}|\boldsymbol{y})}{p(\boldsymbol{x}_{\mathcal{N}_s}|\boldsymbol{y})} &= p(x_s|\boldsymbol{x}_{\mathcal{N}_s}, \boldsymbol{y}), \\ &= \frac{p(x_s, y_s|\boldsymbol{x}_{\mathcal{N}_s}, \boldsymbol{y}_{\mathcal{N}_s})}{\sum_{x'_s \in \Omega} p(x'_s, y_s|\boldsymbol{x}_{\mathcal{N}_s}, \boldsymbol{y}_{\mathcal{N}_s})}, \\ &= \frac{\tilde{p}(x_s, y_s|\boldsymbol{x}_{\mathcal{N}_s}, \boldsymbol{y}_{\mathcal{N}_s})}{\sum_{x'_s \in \Omega} \tilde{p}(x'_s, y_s|\boldsymbol{x}_{\mathcal{N}_s}, \boldsymbol{y}_{\mathcal{N}_s})}. \end{aligned} \tag{C.1}$$

Here the second equality has been seen in Equation 5. Then we have, $\forall s \in \mathcal{S}$:

$$\frac{\tilde{p}(x_s, y_s|\boldsymbol{x}_{\mathcal{N}_s}, \boldsymbol{y}_{\mathcal{N}_s})}{p(x_s, \boldsymbol{x}_{\mathcal{N}_s}|\boldsymbol{y})} = \frac{\sum_{x'_s} \tilde{p}(x'_s, y_s|\boldsymbol{x}_{\mathcal{N}_s}, \boldsymbol{y}_{\mathcal{N}_s})}{p(\boldsymbol{x}_{\mathcal{N}_s}|\boldsymbol{y})}, \tag{C.2}$$

where we can make the same key observation as in [17]: the right-hand side of the last equation is *independent of the realization* $x_s \in \Omega$. Then so is the left-hand side. Then, $\forall s \in \mathcal{S}$, $\forall (x_s, x'_s) \in \Omega^2$, we can write:

$$\begin{aligned} &\frac{\tilde{p}(x_s, y_s|\boldsymbol{x}_{\mathcal{N}_s}, \boldsymbol{y}_{\mathcal{N}_s})}{p(x_s, \boldsymbol{x}_{\mathcal{N}_s}|\boldsymbol{y})} = \frac{\tilde{p}(x'_s, y_s|\boldsymbol{x}_{\mathcal{N}_s}, \boldsymbol{y}_{\mathcal{N}_s})}{p(x'_s, \boldsymbol{x}_{\mathcal{N}_s}|\boldsymbol{y})}, \\ \iff &\frac{\tilde{p}(x_s, y_s|\boldsymbol{x}_{\mathcal{N}_s}, \boldsymbol{y}_{\mathcal{N}_s})}{\tilde{p}(x'_s, y_s|\boldsymbol{x}_{\mathcal{N}_s}, \boldsymbol{y}_{\mathcal{N}_s})} = \frac{p(x_s, \boldsymbol{x}_{\mathcal{N}_s}|\boldsymbol{y})}{p(x'_s, \boldsymbol{x}_{\mathcal{N}_s}|\boldsymbol{y})}. \end{aligned} \tag{C.3}$$

Now, taking the exponential on each side, and using the expression of Equation 11, $\forall s \in \mathcal{S}$, $\forall (x_s, x'_s) \in \Omega^2$:

$$\ln\left(\frac{p(x_s, \boldsymbol{x}_{\mathcal{N}_s}|\boldsymbol{y})}{p(x'_s, \boldsymbol{x}_{\mathcal{N}_s}|\boldsymbol{y})}\right) - \frac{1}{2}Q_{s,s}(\bar{y}'_s)^2 + \frac{1}{2}Q_{s,s}\bar{y}_s^2 + \sum_{t\in\mathcal{N}_s}\mathbb{I}_{[t\in\mathcal{N}_s^2]}Q_{s,t}\bar{y}_s\bar{y}_t - \sum_{t\in\mathcal{N}_s}\mathbb{I}_{[t\in\mathcal{N}_s^2]}Q_{t,s}\bar{y}'_s\bar{y}_t = 2\sum_{t\in\mathcal{N}_s}V(x_s, y_s, x_t, y_t) - 2\sum_{t\in\mathcal{N}_s}V(x'_s, y_s, x_t, y_t), \tag{C.4}$$

where we define $\bar{y}'_s = (y_s - \mu_{x'_s})$ which gives:

$$\ln\left(\frac{p(x_s, \boldsymbol{x}_{\mathcal{N}_s}|\boldsymbol{y})}{p(x'_s, \boldsymbol{x}_{\mathcal{N}_s}|\boldsymbol{y})}\right) + \frac{1}{2}Q_{s,s}\bar{y}_s^2 - \frac{1}{2}Q_{s,s}(\bar{y}'_s)^2 + \sum_{t\in\mathcal{N}_s}\mathbb{I}_{[t\in\mathcal{N}_s^2]}Q_{s,t}\bar{y}_s\bar{y}_t - \sum_{t\in\mathcal{N}_s}\mathbb{I}_{[t\in\mathcal{N}_s^2]}Q_{t,s}\bar{y}'_s\bar{y}_t = \beta\left(2\sum_{t\in\mathcal{N}_s}\mathbb{I}_{[t\in\mathcal{N}_s^1]}\delta_{x_s}^{x_t}\left(1 - \frac{1}{2}(\bar{y}_s - \bar{y}_t)^2\right) - 2\sum_{t\in\mathcal{N}_s}\mathbb{I}_{[t\in\mathcal{N}_s^1]}\delta_{x'_s}^{x_t}\left(1 - \frac{1}{2}(\bar{y}'_s - \bar{y}_t)^2\right)\right). \tag{C.5}$$

The last equation can be written for every each site s, $\forall (x_s, x'_s) \in \Omega^2$. All these equations can be put in the form:

$$\boldsymbol{A} = \boldsymbol{B}\beta, \tag{C.6}$$

where $\beta \in \mathbb{R}_+^*$ and $\boldsymbol{A}, \boldsymbol{B}$ are real vectors with N elements with generic term A_s and B_s respectively defined in Equation 25 and Equation 26.

The probabilities $p(x_s, \boldsymbol{x}_{\mathcal{N}_s}|\boldsymbol{y})$, $\forall x_s \in \Omega$ (which are independent of s) are estimated using the frequency estimator, and β can then be estimated with the LS estimator given in Equation 24.

Appendix D. Tempered equations

We introduce tempered versions of the potentials to be used with the T-Gibbs sampler [11].

- Equations 11 and 10 become:

$$\tilde{p}_T(x_s, y_s|\boldsymbol{x}_{\mathcal{N}_s}, \boldsymbol{y}_{\mathcal{N}_s}) = \exp\left(-\Big(2\sum_{s'\in\mathcal{N}_s}V_T(x_s, y_s, x_{s'}, y_{s'}) + \frac{1}{2}Q_{s,s}\bar{y}_s^2 + \frac{1}{T}\sum_{s'\in\mathcal{N}_s}\mathbb{I}_{[t\in\mathcal{N}_s^2]}Q_{s,s'}\bar{y}_s\bar{y}_{s'}\Big)\right), \tag{D.1}$$

where

$$V_T(x_s, y_s, x_{s'}, y_{s'}) = -\mathbb{I}_{[t\in\mathcal{N}_s^1]}\delta_{x_s}^{x_{s'}}\beta\left(1 - \frac{1}{2T}(\bar{y}_s - \bar{y}_{s'})^2\right). \tag{D.2}$$

- Equation 20 becomes:

$$\tilde{p}_T(y_s|x_s, \boldsymbol{x}_{\mathcal{N}_s}, \boldsymbol{y}_{\mathcal{N}_s}) = \exp\left(-\ln\left(\sqrt{2\pi V_s}\right) - \frac{(y_s - M_{T,s})^2}{2V_s}\right), \tag{D.3}$$

with

$$M_{T,s} = \mu_{x_s} - \frac{V_s}{T}\sum_{s'\in\mathcal{N}_s}\mathbb{I}_{[t\in\mathcal{N}_s^2]}Q_{s,s'}\bar{y}_{s'} \text{ and } V_s = Q_{s,s}^{-1}. \tag{D.4}$$

Combinations of these potentials are used to sample with the T-Gibbs sampler with the P-GMRF and GPMF models. It consists in Equations 17 and D.3 for P-GMRF and Equation D.1 for GPMF.

Appendix E. Tempered-Gibbs sampler

Algorithm 2: Tempered-Gibbs sampler

Data: A set of ordered temperatures $\{T_k\}_{k=1}^{k=K}$
Initial states of the parallel Markov chains $\{\boldsymbol{x}_k^1\}_{k=1}^{k=K}$
Result: $\boldsymbol{x}_1^i$ the sample at model temperature
$i \leftarrow 1$
while *not converged* **do**
  /* Gibbs sampler for each chain */
  for $k \in \{K, \ldots, 1\}$ **do**
    $\boldsymbol{x}_k^{i+1}$ drawn by Gibbs sampler from $p_{T_k}(\boldsymbol{x})$ with initialization at $\boldsymbol{x}_k^i$
  end
  /* Test to swap chains */
  for $k \in \{K-1, \ldots, 1\}$ **do**
    $\boldsymbol{x}_k^{i+1} \leftarrow \boldsymbol{x}_{k+1}^{i+1}$ with probability $\alpha^k(\boldsymbol{x}_k^{i+1}, \boldsymbol{x}_{k+1}^{i+1})$
  end
  $i \leftarrow i + 1$
end

Algorithm 2 details the Gibbs sampler with tempered distributions which we call Tempered Gibbs sampler (T-Gibbs sampler). In this algorithm, let K be the number of Gibbs samplers that we run in parallel. We then have K associated and ordered temperatures starting from the temperature at which the probabilistic model is defined; which is 1 in our case. Two chains with successive temperatures can swap with probability:

$$\alpha^k(\boldsymbol{x}_k^{i+1}, \boldsymbol{x}_{k+1}^{i+1}) = \min\left(\exp\left((\frac{1}{T_k} - \frac{1}{T_{k+1}})(E(\boldsymbol{x}_k^{i+1}) - E(\boldsymbol{x}_{k+1}^{i+1}))\right), 1\right). \tag{E.1}$$

In our experiments, the set of temperatures is fixed for all the PMF models to 16 linearly spaced temperatures ranging from 1 to 20 [38].