A Bayesian Nonparametric Model Coupled with a Markov Random Field for Change Detection in Heterogeneous Remote Sensing Images

Jorge Prendes, Marie Chabert, Frédéric Pascal, Alain Giros, Jean-Yves Tourneret

A Bayesian Nonparametric Model Coupled with a Markov Random Field for Change Detection in Heterogeneous Remote Sensing Images*

Jorge Prendes†, Marie Chabert‡, Frédéric Pascal§, Alain Giros¶, and Jean-Yves Tourneret‡

Abstract. In recent years, remote sensing of the Earth surface using images acquired from aircraft or satellites has gained a lot of attention. The acquisition technology has been evolving fast and, as a consequence, many different kinds of sensors (e.g., optical, radar, multispectral, and hyperspectral) are now available to capture different features of the observed scene. One of the main objectives of remote sensing is to monitor changes on the Earth surface. Change detection has been thoroughly studied in the case of images acquired by the same sensors (mainly optical or radar sensors). However, due to the diversity and complementarity of the images, change detection between images acquired with different kinds of sensors (sometimes referred to as heterogeneous sensors) is clearly an interesting problem. A statistical model and a change detection strategy were recently introduced in [J. Prendes, M. Chabert, F. Pascal, A. Giros, and J.-Y. Tourneret, *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Florence, Italy, 2014; IEEE Trans. Image Process., 24 (2015), pp. 799–812] to deal with images captured by heterogeneous sensors. The main idea of the suggested strategy was to model the objects contained in an analysis window by mixtures of distributions. The manifold defined by these mixtures was then learned using training data belonging to unchanged areas. The changes were finally detected by thresholding an appropriate distance to the estimated manifold. This paper goes a step further by introducing a Bayesian nonparametric framework allowing us to deal with an unknown number of objects in analysis windows without specifying an upper bound for this number. A Markov random field is also introduced to account for the spatial correlation between neighboring pixels. The proposed change detector is validated using different sets of synthetic and real images (including pairs of optical images and pairs of optical and radar images) showing a significant improvement when compared to existing algorithms.

Key words. optical images, synthetic aperture radar images, change detection, Bayesian nonparametric, Markov random field, collapsed Gibbs sampler

DOI. 10.1137/15M1047908

Funding: Part of this work was supported by BNPSI ANR project ANR-13-BS-03-0006-01 and by ANR-11-LABX-0040-CIMI in particular during the program ANR-11-IDEX-0002-02 within the thematic trimester on image processing.

†TéSA Laboratory, 31500 Toulouse, France (jorge.prendes@tesa.prd.fr).

‡INP/ENSEEIHT–IRIT, University of Toulouse, 31071 Toulouse, France (marie.chabert@enseeiht.fr, jean-yves.tourneret@enseeiht.fr).

§L2S/CentraleSupelec, 91190 Gif-sur-Yvette, France (frederic.pascal@centralesupelec.fr).

¶CNES, 31401 Toulouse, France (alain.giros@cnes.fr).

NOTATION.

A summary of the notation used throughout the article can be found in the following table. Lowercase bold letters denote column vectors, while uppercase bold letters denote matrices. Subindexes are denoted using lowercase letters, and their upper limit is denoted by capital letters, e.g., the subindex k can take the values $1 \leq k \leq K$. The notation $\boldsymbol{z}_{\setminus n}$ denotes all the elements of the vector $\boldsymbol{z}$ except the nth element.

D	Number of images in a dataset
W	Analysis window
N	Number of pixels in W
S	Sensor used to acquire an image
S_d	Sensor used to acquire the dth image
$i_{n,S}$	Intensity measured by S for the nth pixel
$\boldsymbol{i}_n$	D-dimensional vector of pixel intensities acquired by all sensors for the nth pixel
$\boldsymbol{I}$	Matrix containing all the intensities $\boldsymbol{i}_n$
K	Number of different objects in W
P	Physical properties of an object
P_k	Physical properties of the kth object
w_k	Proportion of W covered by the kth object
$\boldsymbol{w}$	K-dimensional vector containing all w_k
$T_S(P)$	How S images an object with properties P
η_S	Acquisition noise of S
$\mathcal{F}(\boldsymbol{v})$	Application dependent distribution family with parameter vector $\boldsymbol{v}$
$\mathcal{V}_0$	Prior distribution for $\boldsymbol{v}$
$\boldsymbol{v}_n$	Parameter vector that identifies a distribution from $\mathcal{F}$ for the nth pixel
$\boldsymbol{V}$	Matrix containing all the $\boldsymbol{v}_n$
$\boldsymbol{v}'_k$	Parameter vector that identifies a distribution from $\mathcal{F}$ for the kth object
$\boldsymbol{V}'$	Matrix containing all the vectors $\boldsymbol{v}'_k$
$\mathcal{M}$	"No-change" manifold
$\hat{T}_{k,S_d}$	Estimated parameters for the kth component of the dth sensor
$\hat{v}_k$	Estimated parameters for the kth component
Δ_W	Similarity measure for the analysis window W
Δ_n	Similarity measure for the nth pixel
$\mathrm{DP}(\cdot,\cdot)$	Dirichlet process
$\psi(\cdot)$	Digamma function
α	Concentration parameter
z_n	Class label for the nth pixel
$\boldsymbol{z}$	Vector containing the class labels of all pixels

1. Introduction. Remote sensing images of the Earth are becoming more and more accessible. Detecting and tracking changes on the Earth surface using these images is a privileged means for urban growth tracking [57, 59], plantation monitoring, urban database updating [48], and natural disaster study management [61]. For this purpose, many different remote sensing technologies are available, such as optical multispectral and hyperspectral sensors [55] or synthetic aperture radar (SAR) [8, 17]. The specific properties of each technology make it

attractive in particular situations. For instance, optical images are easy to interpret. They are characterized by a very high spatial resolution and a high signal to noise ratio. However, their acquisition requires daylight and good weather conditions. On the other hand, SAR images can be captured during the night and even in the presence of clouds or smoke. However, they are more difficult to interpret and are characterized by lower spatial resolution and a poor signal to noise ratio [34]. In order to take advantage of the complementarity of different image modalities, multitemporal image datasets are often composed of images obtained from heterogeneous sensors. Exploiting the complementarity of these heterogeneous images is an important challenge. This paper focuses on change detection between multitemporal co-registered images, acquired by either homogeneous or heterogeneous sensors. The case of heterogeneous optical and SAR images will receive particular attention.

Detecting changes between images acquired using the same kind of sensor has received particular attention in the remote sensing community. The proposed techniques usually exploit appropriate features extracted from the images, depending on the considered kind of sensor. As a consequence, change detection strategies can be classified according to the kind of sensor they target. For instance, detecting changes between optical images can be achieved using the so-called difference image defined as the pixelwise difference between the image intensities [5, 6, 12, 22, 56]. Other approaches are based on the difference image derived in the wavelet domain [11, 13] or on artificial neural networks [27, 28, 43, 44]. Detecting changes between SAR images has also been intensively studied, usually exploiting the fact that these images are corrupted by a multiplicative speckle noise. For instance, different techniques based on the ratio of pixel intensities have been proposed [2, 7, 20, 53, 60, 63]. Other relevant methods are based on neural networks [4, 49] or on the joint distribution of the pixel intensities [15, 29, 52].

Detecting changes between heterogeneous images using multivariate distributions or copulas has also been considered in the literature [14, 33, 38]. Recently, a change detection strategy for homogeneous or possibly heterogeneous images was proposed in [50, 51]. The suggested strategy assumed that the same objects were present in the co-registered multitemporal images in the absence of change, inducing some relationships between the image intensities. A statistical model describing the joint distribution of the pixel intensities was then proposed. The parameters of this model were finally used to detect changes through a manifold learning strategy. The method showed improved performance with respect to the methods proposed in [14, 33, 38]. The image model presented in [50, 51] is a statistical model for the pixel intensities accounting for the sensor noise. However, the spatial correlation between neighboring pixels of the image was not exploited. Moreover, the algorithm required to define a maximum number of objects present in the analysis window. The present paper investigates a Bayesian nonparametric (BNP) framework which allows these two limitations to be overcome. More precisely, the proposed BNP model expresses the joint distribution of the pixels located in the analysis window as a mixture of objects whose number is not limited. This model is coupled with a Markov random field (MRF) prior [16, 41] to exploit the spatial correlation of neighboring pixels. A partially collapsed Gibbs sampler [62] is then investigated to sample the posterior distribution of the BNP model and to use these samples to build estimators of its parameters. A Jeffreys prior for the concentration parameter of the BNP model is finally derived to reduce the number of parameters required by the algorithm and improve its flexibility.

The first contribution of this paper is a new BNP model, its estimation algorithm, and its application to change detection. This algorithm is based on a label prior allowing an unbounded number of objects in each analysis window and exploiting the spatial correlations between adjacent image pixels. The second contribution of the paper is the derivation of a Jeffreys prior for the concentration parameter of the proposed nonparametric model, as well as the corresponding sampling algorithm.

The paper is structured as follows. Section 2 reviews the statistical model introduced in [50, 51] and the associated frequentist inference for change detection. Section 3 introduces the theoretical background for the BNP framework suggested to solve the model selection problem associated with this change detection. A new statistical model based on a Dirichlet process mixture is proposed to describe the joint distribution of a set of images acquired by either homogeneous or heterogeneous sensors. A Jeffreys prior for the concentration parameter of the proposed BNP model is finally derived, allowing this parameter to be estimated jointly with all the other parameters. Section 4 provides a quick introduction to MRFs and discusses their integration into the BNP framework. The change detection strategy associated with the proposed statistical model is presented in section 5. Change detection results obtained with the proposed approach for various synthetic and real images are presented in section 6. Conclusions, possible improvements and future work are reported in section 7.

2. Frequentist model for change detection.

2.1. Image model. A statistical model was introduced in [50, 51] to describe the intensities of a pixel belonging to a set of remote sensing images captured by homogeneous and heterogeneous sensors. The change detection approach proposed in this article is based on this generic model (in the sense that it is appropriate for any set of sensors), and its application to SAR and optical images will be detailed. This model considers the statistical properties of the noise affecting each sensor. For the sensor S, the observed pixel intensity (denoted as i_S) can be written

$$i_S = f_S[T_S(P), \eta_S], \tag{2.1}$$

where $T_S(P)$ expresses the ideal response of the sensor S to a particular object with physical properties P, and $f_S(\cdot,\cdot)$ indicates how the measurements are affected by the sensor noise η_S. For instance, in the case of an optical sensor, the noise can be modeled as an additive zero mean Gaussian noise [10, 37], i.e.,

$$\eta_{\text{Opt}} \sim \mathcal{N}\big(0, \sigma^2_{\text{Opt}}\big), \tag{2.2}$$

$$f_{\text{Opt}}[T_{\text{Opt}}(P), \eta_{\text{Opt}}] = T_{\text{Opt}}(P) + \eta_{\text{Opt}}, \tag{2.3}$$

where σ^2_{Opt} is the variance of the measurement noise. The distribution of i_S can be easily obtained conditionally on $T_S(P)$, i.e., conditionally on a particular object with physical properties P, leading to $\text{p}[i_S|T_S(P)] = \text{p}(i_S|P)$. For an optical sensor the following result is finally obtained:

$$i_{\text{Opt}}|P \sim \mathcal{N}\big[T_{\text{Opt}}(P), \sigma^2_{\text{Opt}}\big]; \tag{2.4}$$

it is in accordance with classical statistical properties of optical images [10, 37].

Similarly, the intensity of a multilook SAR image can be modeled as a ground truth intensity $T_{\mathrm{SAR}}(P)$ corrupted by a multiplicative speckle noise (with a gamma distribution) [21, 36, 40] η_{SAR}, i.e.,

$$\eta_{\mathrm{SAR}} \sim \Gamma\big(L_{\mathrm{SAR}}, L_{\mathrm{SAR}}^{-1}\big), \tag{2.5}$$

$$f_{\mathrm{SAR}}[T_{\mathrm{SAR}}(P), \eta_{\mathrm{SAR}}] = T_{\mathrm{SAR}}(P)\eta_{\mathrm{SAR}}, \tag{2.6}$$

where $\Gamma(x; k, \theta) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}} \mathbf{1}_{\mathbb{R}^+}(x)$ is the gamma distribution, $\Gamma(k)$ is the gamma function, L_{SAR} is the number of looks of the SAR image, and $\mathbf{1}_A(\cdot)$ is the indicator function of set A (i.e., $\mathbf{1}_A(x) = 1$ if $x \in A$ and $\mathbf{1}_A(x) = 0$ everywhere else). Thus, the conditional distribution of i_{SAR} given P is

$$i_{\mathrm{SAR}}|P \sim \Gamma\big[L_{\mathrm{SAR}}, T_{\mathrm{SAR}}(P)L_{\mathrm{SAR}}^{-1}\big]. \tag{2.7}$$

Consider now the D-dimensional vector $\boldsymbol{i} = [i_{S_1}, \ldots, i_{S_D}]^T$ containing the intensities for a particular pixel associated with D different co-registered images acquired by the sensors S_d with $1 \leq d \leq D$. Since the D images are acquired by different sensors, we can assume that the different noises η_{S_d} are independent, leading to independent intensities $i_{S_1}|P, \ldots, i_{S_D}|P$. As a consequence, an object observed jointly in the D images can be described by the following joint distribution:

$$\mathrm{p}(\boldsymbol{i}|P) = \prod_{d=1}^{D} \mathrm{p}[i_{S_d}|T_{S_d}(P)] = \prod_{d=1}^{D} \mathrm{p}(i_{S_d}|P), \tag{2.8}$$

where $\mathrm{p}(\boldsymbol{i}|P)$ defines a family $\mathcal{F}$ of distributions that depends on the kind of involved sensors. For a given object characterized by P, a vector of parameters $\boldsymbol{v} = \boldsymbol{v}(P)$ identifying a particular distribution of the family $\mathcal{F}$ can be defined as

$$\boldsymbol{v} = \boldsymbol{v}(P) = [T_{S_1}(P), \ldots, T_{S_D}(P)]^T, \tag{2.9}$$

where $T_{S_d}(P)$ is the ground truth (noiseless) intensity in (2.1), but is also the parameter of $\mathrm{p}(i_{S_d}|P)$. From this we can express $i|P$ as

$$\boldsymbol{i}|\boldsymbol{v} \sim \mathcal{F}(\boldsymbol{v}). \tag{2.10}$$

However, an analysis window W generally contains different objects. Consequently, different values of P are observed in different window regions. Since an analysis window generally contains a finite number K of objects, it is natural to model the distribution of $P|W$ for a given image as the following mixture:

$$\mathrm{p}(P|W) = \sum_{k=1}^{K} w_k \delta(P - P_k), \tag{2.11}$$

where $\delta(\cdot)$ is the Dirac delta function, K is the number of objects in the analysis window W, and w_k is related to the area of the kth object within the window W. In an unchanged window W the distribution $P|W$ will be the same for all sensors. Consequently, $\boldsymbol{i}|W$ is also a mixture distribution defined as

$$\mathrm{p}(\boldsymbol{i}|W) = \sum_{k=1}^{K} w_k \mathrm{p}(\boldsymbol{i}|P_k) = \sum_{k=1}^{K} w_k \prod_{d=1}^{D} \mathrm{p}(i_{S_d}|P_k), \tag{2.12}$$

where all the mixture components $\mathrm{p}(\boldsymbol{i}|P_k)$ belong to the same distribution family $\mathcal{F}$ with different sets of parameter vectors $\boldsymbol{v}_k = \boldsymbol{v}(P_k)$. Note that the distribution of $\boldsymbol{i}|W$ is fully characterized by the values of its parameters K, $\boldsymbol{w} = [w_1, \ldots, w_K]$ and $\boldsymbol{V} = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K]$.

2.2. Change detection strategy. Estimating the parameters of $\mathrm{p}(\boldsymbol{i}|W)$ leads to an estimator $\hat{\boldsymbol{V}} = [\hat{\boldsymbol{v}}_1, \ldots, \hat{\boldsymbol{v}}_K]$ with $\hat{\boldsymbol{v}}_k = [\hat{T}_{k,S_1}, \ldots, \hat{T}_{k,S_D}]^T$ for $k = 1, \ldots, K$ and where $\hat{T}_{k,S_d}$ is an estimator of $T_{S_d}(P_{k,d})$. For unchanged areas, $P_{k,d}$ does not depend on d, i.e., $P_{k,d} = P_k$, implying that $\hat{T}_{k,S_d}$ and $\hat{\boldsymbol{v}}_k$ are estimators of $T_{S_d}(P_k)$ and $\boldsymbol{v}(P_k)$, respectively. Conversely, the value of $\hat{P}_{k,d}$ is different from one image to another when a change has occurred in the scene between the image acquisition times. In this case, different objects are observed in the different images and $\hat{T}_{k,S_d}$ is an estimator of $T_{S_d}(P_{k,d})$, but $\hat{\boldsymbol{v}}_k$ does not define an estimator of $\boldsymbol{v}(P_k)$ for a unique P_k.

In order to build a change detection strategy, we assume that, in the absence of change, $\boldsymbol{v} = \boldsymbol{v}(P)$ defines a manifold $\mathcal{M}$ generated by the parameter P. When considering an object characterized by P_k in an unchanged area, the estimator $\hat{\boldsymbol{v}}_k$ converges to $\boldsymbol{v}(P_k)$, i.e., to a point located in the manifold $\mathcal{M}$. On the other hand, for changed areas, $\hat{\boldsymbol{v}}_k$ does not generally converge to a point located in the manifold $\mathcal{M}$. Based on these observations, a distance denoted as Δ_k between $\hat{\boldsymbol{v}}_k$ and the manifold $\mathcal{M}$ can be used as a similarity measure to detect changes between images. More precisely, we consider the binary hypothesis testing problem

$$\begin{aligned} H_0 &: \text{Absence of change,} \\ H_1 &: \text{Presence of change} \end{aligned}$$

and introduce the following change detection strategy:

$$\Delta_k \underset{H_0}{\overset{H_1}{\gtrless}} \tau, \tag{2.13}$$

where τ is a threshold related to the probability of false alarm (PFA) of the change detector.

The manifold $\mathcal{M}$ can be obtained analytically in simple cases, i.e., for simple sensor combinations. For instance, when two grayscale images are captured with the same optical sensor S in the same conditions, we have $S_1 = S_2 = S$, and thus $T_{S_1}(P) = T_{S_2}(P) = T_S(P)$. In this case, $\mathcal{M}$ is defined by the line $T_{S_1}(P) = T_{S_2}(P)$ in the plane $T_{S_1}(P) \times T_{S_2}(P)$, as depicted in Figure 1, red line. If the two images are captured with different contrasts, one of the images can be subjected to a saturation, leading to a slightly more complex manifold, as depicted in Figure 1, blue line. However, in most situations (e.g., different settings, different sensors, multichannel images), the manifold $\mathcal{M}$ is a priori unknown. Assuming that $\mathrm{p}(\hat{\boldsymbol{v}}|P)$

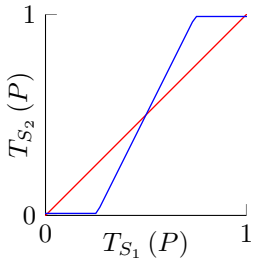


Figure 1. *Manifolds obtained with simple sensor combinations (red curve: two identical sensors; blue curve: sensors with different contrasts with a saturation).*

has a maximum at $\boldsymbol{v}(P)$ (i.e., in the manifold $\mathcal{M}$), we proposed in [50, 51] to use the estimated probability density function (pdf) of $\hat{\boldsymbol{v}}$ to define the similarity measure Δ_k, i.e.,

$$\Delta_k = \hat{\mathrm{p}}_{\hat{v}}(\hat{v}_k), \tag{2.14}$$

where the density $\hat{\mathrm{p}}_{\hat{v}}(\cdot)$ was estimated by means of the vectors $\hat{\boldsymbol{v}}$ belonging to a set of "no change" images, or by using regions of the image that are not affected by changes (supervised learning). The resulting detection strategy detects a change if the estimated pdf is less than a threshold depending on a given PFA, i.e., as follows:

$$\Delta_k \underset{H_1}{\overset{H_0}{\gtrless}} \tau^{-1}. \tag{2.15}$$

3. A BNP model. This section studies a different approach to building the distribution (2.12) based on a BNP framework [26, 58]. We begin in subsection 3.1 by presenting a simple Bayesian approach to the model described in subsection 2.1 and motivate the need for a BNP framework. Subsection 3.2 explains how the model recalled in subsection 2.1 can be extended to an infinite number of dimensions via a BNP approach. Subsection 3.3 investigates a collapsed Gibbs sampler to sample the resulting posterior distribution. The generated samples are then used to estimate the parameters of the proposed model. Finally, subsection 3.4 studies a Jeffreys prior to estimate the concentration parameter α that arises in subsection 3.2 and derives an algorithm to sample it from this prior.

3.1. Dirichlet mixture model. A Bayesian description of the model introduced in subsection 2.1 can be obtained by including a prior for the physical properties P. More precisely, we introduce the following prior for the vectors $\boldsymbol{v} = \boldsymbol{v}(P)$ for a given number of objects K in the observation window W composed of N pixels:

$$\boldsymbol{i}_n | \boldsymbol{v}_n \sim \mathcal{F}(\boldsymbol{v}_n), \tag{3.1}$$

$$p(\boldsymbol{v}_n | \boldsymbol{V}') = \sum_{k=1}^{K} w_k \delta(\boldsymbol{v}_n - \boldsymbol{v}'_k), \tag{3.2}$$

where $\boldsymbol{i}_n = [i_{n,S_1}, \ldots, i_{n,S_D}]^T$ (for $1 \leq n \leq N$) is the pixel intensity vector of the nth pixel, $\boldsymbol{v}_n$ is the parameter vector associated with the object containing the nth pixel, $\boldsymbol{V}' = [\boldsymbol{v}'_1, \ldots, \boldsymbol{v}'_K]^T$,

Algorithm 1. Generation of samples from a realization of a Dirichlet process.

Data: $\mathcal{V}_0$, α
Result: $v_1, v_2, v_3, \ldots$
for $n \geq 1$ **do**
  $u \sim \text{Uniform}(1, \alpha + n)$;
  if $u < n$ **then**
    $\boldsymbol{v}_n \leftarrow \boldsymbol{v}_{\lfloor u \rfloor}$;
  else
    $\boldsymbol{v}_n \sim \mathcal{V}_0$;

where $\boldsymbol{v}'_k$ is the parameter vector associated with the kth cluster or object, and $\mathcal{F}(\boldsymbol{v}_n)$ is a distribution on the family $\mathcal{F}$ identified by the parameter vector $\boldsymbol{v}_n$. Introducing a cluster label for each pixel in the observation window $\boldsymbol{z} = [z_1, \ldots z_N]^T$ we obtain an equivalent model given by

$$\boldsymbol{i}_n | z_n \sim \mathcal{F}(\boldsymbol{v}'_{z_n}), \tag{3.3}$$

$$z_n \sim \text{Cat}_K(\boldsymbol{w}), \tag{3.4}$$

where $\boldsymbol{w} = [w_1, \ldots, w_K]^T$ and $\text{Cat}_K(\cdot)$ is the Kth dimensional categorical distribution.

After defining a prior for $\boldsymbol{v}'_k$ and $\boldsymbol{w}$, we can build a Bayesian framework [32] to estimate the mixture parameters. In this paper, we have considered the following prior information:

$$\boldsymbol{v}'_k \sim \mathcal{V}_0, \tag{3.5}$$

$$\boldsymbol{w} \sim \text{Dir}_K(\alpha), \tag{3.6}$$

where $\mathcal{V}_0$ is a prior for the parameter vector $\boldsymbol{v}'_k$ that depends on the application, and $\text{Dir}_K(\alpha)$ denotes the classical conjugate prior for categorical distributions, i.e., the symmetric Dirichlet distribution of dimension K and concentration parameter α. Applying these priors to a mixture model results in the so-called Dirichlet mixture model. However, this model requires knowing the parameter K a priori, which can be a problem. This problem was heuristically solved in [50, 51] by testing different values of K within a predefined range, and by analyzing the sensitivity of the change detector to the number of clusters K.

The BNP framework investigated in this paper allows this limitation to be removed by making K unknown and unbounded. When K goes to infinity in the previous model, it yields an infinite dimensional parameter vector $\boldsymbol{w}$ and a matrix $\boldsymbol{V}'$ with an infinite number of columns. Estimating an infinite number of parameters is clearly untractable. The model presented in the next section considers these parameters as intermediate and does not require their estimation. The model leads to a finite number of vectors $\boldsymbol{v}_n$ through the estimation of a finite number of parameters, namely, the concentration parameter α and the distribution $\mathcal{V}_0$.

3.2. Dirichlet process mixture model. Define as $\boldsymbol{V} = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N]$ the matrix containing the N random vectors $\boldsymbol{v}_n$, $n = 1, \ldots, N$, where $\boldsymbol{v}_n$ is associated with the nth pixel of the observation window. Algorithm 1 shows an approach to generate these random vectors by using an iterative algorithm (see [3, 19, 23, 39, 45]). However, since some of these vectors

are possibly repeated, we finally obtain $K \leq N$ different vectors that are denoted as $\boldsymbol{v}'_k$, $k = 1, \ldots, K$, associated with each object of the observation window. We can compute the joint distribution of the N vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N$ by using the chain rule as explained in [26],

$$\mathrm{p}(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N) = \prod_{m=1}^{N} \mathrm{p}(\boldsymbol{v}_m | \boldsymbol{v}_1, \ldots, \boldsymbol{v}_{m-1}) = \frac{\prod_{k=1}^{K} \alpha\, \Gamma(N_k)\, \mathrm{p}_{\mathcal{V}_0}(\boldsymbol{v}'_k)}{\Gamma(\alpha + N)/\Gamma(\alpha)}, \tag{3.7}$$

where $p_{\mathcal{V}_0}$ is the pdf associated with the distribution $\mathcal{V}_0$, and N_k is the number of vectors $\boldsymbol{v}_n$ taking the value $\boldsymbol{v}'_k$. This distribution can be factorized into two terms: the first one is related to the particular values of $\boldsymbol{v}'_k$ and the second one is related to the data partitioning. The resulting conditional distribution of $\boldsymbol{V}$ given $\boldsymbol{V}'$ can be written

$$\mathrm{p}\big(\boldsymbol{V} \big| \boldsymbol{V}'\big) = \frac{\prod_{k=1}^{K} \alpha\, \Gamma(N_k)}{\Gamma(\alpha + N)/\Gamma(\alpha)}. \tag{3.8}$$

It can be observed that the distribution of $\boldsymbol{V}|\boldsymbol{V}'$ only depends on the cardinal of each partition set and not on the order the vectors have been drawn. Thus, any random vector can be thought of as if it was the last drawn vector, meaning that these random vectors are exchangeable. Using the "de Finetti's theorem" [42], one can show that the vectors $\boldsymbol{v}_n$ are conditionally independent given a latent distribution $\mathcal{V}$. In this case, $\mathcal{V}$ is defined by the pdf

$$\mathrm{p}_{\mathcal{V}}(\boldsymbol{v}_n) = \sum_{k=0}^{\infty} w_k\, \delta\big(\boldsymbol{v}_n - \boldsymbol{v}'_k\big) \tag{3.9}$$

with

$$\boldsymbol{v}'_k \sim \mathcal{V}_0, \tag{3.10}$$

$$w_k = w'_k \prod_{j=1}^{k-1} (1 - w'_j), \tag{3.11}$$

$$w'_k \sim \mathrm{Beta}(1, \alpha), \tag{3.12}$$

where (3.11) and (3.12) are known as a stick breaking process denoted as $\mathrm{SBP}(\alpha)$, which can be thought of as a generalization of $\mathrm{Dir}_K(\alpha)$ for $K \to \infty$. The parameter α controls the concentration of the variables w_k. Small values of α provide few values for the discrete distribution $\mathcal{V}$ containing most of the probability, while high values of α provide reduced sparsity in the distribution $\mathcal{V}$, leading to a uniform distribution. The vectors $\boldsymbol{v}_n$ defined by (3.9) tend to be grouped into clusters. As will be discussed in subsection 3.4, the expected number K of different clusters given that the windows contains N pixels is

$$\mathrm{E}(K|\alpha, N) = \alpha[\psi(\alpha + N) - \psi(\alpha)], \tag{3.13}$$

where $\psi(\cdot)$ denotes the digamma function. Note that $\lim_{\alpha \to \infty} \mathrm{E}(K|\alpha, N) = N$, since N samples can generate at most N clusters.

A Dirichlet process $\mathrm{DP}(\mathcal{V}_0, \alpha)$ is a stochastic process whose realizations are probability distributions $\mathcal{V}$, so that the vectors $\boldsymbol{v}_n$ in Algorithm 1 can be described as

$$\boldsymbol{v}_n \sim \mathcal{V}, \tag{3.14}$$

$$\mathcal{V} \sim \mathrm{DP}(\mathcal{V}_0, \alpha). \tag{3.15}$$

Note that the Dirichlet process is often chosen as the conjugate prior for infinite discrete distributions. The relevance of Algorithm 1 is that it provides a method to generate samples $\boldsymbol{v}_n$ from a distribution $\mathcal{V}$ with an infinite number of parameters only from its Bayesian priors without requiring the direct computation of its parameters. The finite mixture model in subsection 2.1 can thus be extended through a BNP framework into a Dirichlet process mixture model (DPMM),

$$\boldsymbol{i}_n|\boldsymbol{v}_n \sim \mathcal{F}(\boldsymbol{v}_n), \tag{3.16}$$

$$\boldsymbol{v}_n \sim \mathcal{V}, \tag{3.17}$$

$$\mathcal{V} \sim \mathrm{DP}(\mathcal{V}_0, \alpha), \tag{3.18}$$

where $\mathcal{V}_0$ is the base distribution and α is the concentration parameter.

3.3. Parameter estimation. To estimate the parameters $\boldsymbol{v}_n$ of a DPMM, we suggest use of a Markov chain Monte Carlo (MCMC) algorithm based on a collapsed Gibbs sampler [9]. Introducing the cluster labels for each pixel of the observation window $\boldsymbol{z} = [z_1, \ldots, z_N]$, we obtain the equivalent model

$$\boldsymbol{i}_n|z_n \sim \mathcal{F}\big(\boldsymbol{v}'_{z_n}\big), \tag{3.19}$$

$$z_n \sim \mathrm{Cat}_\infty(\boldsymbol{w}), \tag{3.20}$$

$$\boldsymbol{w} \sim \mathrm{SBP}(\alpha), \tag{3.21}$$

$$\boldsymbol{v}'_k \sim \mathcal{V}_0, \tag{3.22}$$

where $\mathrm{Cat}_\infty(\boldsymbol{w})$ is an infinite dimensional generalization of the categorical distribution such that $\mathrm{p}(z_n = k) = w_k$ for any $k \in \mathbb{N}^*$. This parametrization is equivalent to the one defined by (3.16)–(3.18) but makes explicit that the pixels $\boldsymbol{i}_n$, for $n = 1, \ldots, N$, are partitioned into different clusters. Moreover, (3.20) and (3.21) define the so-called Chinese restaurant process $\mathrm{CRP}(\alpha)$. Thus, the model can be reduced to

$$\boldsymbol{i}_n|z_n \sim \mathcal{F}\big(\boldsymbol{v}'_{z_n}\big), \tag{3.23}$$

$$\boldsymbol{z} \sim \mathrm{CRP}(\alpha), \tag{3.24}$$

$$\boldsymbol{v}'_k \sim \mathcal{V}_0. \tag{3.25}$$

The advantage of this last parametrization is that it allows the parameters $\boldsymbol{v}'_k$ to be integrated out, leading to a partially collapsed Gibbs sampler [62]. More precisely, to estimate the latent variables $\boldsymbol{z}$, we can sample from $\mathrm{p}(\boldsymbol{z}|\boldsymbol{I}, \mathcal{V}_0)$, where $\boldsymbol{I} = [\boldsymbol{i}_1, \ldots, \boldsymbol{i}_N]$. The Gibbs sampler is an iterative algorithm that samples sequentially the conditional probabilities of each variable with respect to (w.r.t.) the other variables. For the proposed problem, samples

from $\mathrm{p}\big(z_n\big|\boldsymbol{z}_{\backslash n},\boldsymbol{I},\mathcal{V}_0\big)$ are generated, where $\boldsymbol{z}_{\backslash n}=[z_1,\ldots,z_{n-1},z_{n+1},\ldots,z_N]$. This conditional probability can be obtained as follows:

$$\mathrm{p}\big(z_n\big|\boldsymbol{z}_{\backslash n},\boldsymbol{I},\mathcal{V}_0\big)\propto \mathrm{p}(\boldsymbol{I}|\boldsymbol{z},\mathcal{V}_0)\mathrm{p}\big(z_n\big|\boldsymbol{z}_{\backslash n}\big), \tag{3.26}$$

where $\propto$ means "proportional to," and $\mathrm{p}\big(z_n\big|\boldsymbol{z}_{\backslash n}\big)$ can be obtained using $\mathrm{p}(\boldsymbol{z})=\mathrm{p}(\boldsymbol{V}|\boldsymbol{V}')$ and (3.8). More precisely, the following result can be obtained:

$$\mathrm{p}\big(z_n\big|\boldsymbol{z}_{\backslash n}\big)=\mathrm{p}(\boldsymbol{z})\times\mathrm{p}\big(\boldsymbol{z}_{\backslash n}\big)^{-1}=\frac{\prod_{k=0}^{K^*}\alpha\,\Gamma(N_k)}{\Gamma(\alpha+N)/\Gamma(\alpha)}\times\frac{\Gamma(\alpha+N-1)/\Gamma(\alpha)}{\prod_{k=1}^{K^*}\alpha\,\Gamma[N_k-\mathbf{1}_k(z_n)]} \tag{3.27}$$

$$=\begin{cases}\frac{\alpha}{\alpha+N-1} & \text{if } z_n=0,\\ \frac{N'_{z_n}}{\alpha+N-1} & \text{if } 1\leq z_n\leq K^*,\end{cases} \tag{3.28}$$

where $\mathbf{1}_k(z_n)$ is the indicator function, K^* is the number of different values in $\boldsymbol{z}$ at the nth iteration, N'_{z_n} is the number of pixels in the cluster indicated by z_n (excluding z_n), and $z_n=0$ when a new cluster is created. Note that (3.27) incurs in an abuse of notation requiring to define $\Gamma(0)=\frac{1}{\alpha}$, so that empty clusters do not affect the equation (e.g., for $k=0$ and $z_n\neq 0$). The probability $\mathrm{p}(\boldsymbol{I}|\boldsymbol{z},\mathcal{V}_0)$ is obtained by marginalizing out $\boldsymbol{V}'$ from the likelihood $\mathrm{p}(\boldsymbol{I}|\boldsymbol{z},\boldsymbol{V}')$ as follows:

$$\mathrm{p}(\boldsymbol{I}|\boldsymbol{z},\mathcal{V}_0)=\int\mathrm{p}(\boldsymbol{I}|\boldsymbol{z},\boldsymbol{V}')\mathrm{p}(\boldsymbol{V}'|\mathcal{V}_0)d\boldsymbol{V}'=\prod_{k=1}^{K}\int\mathrm{p}(\boldsymbol{I}_{\{k\}}|\boldsymbol{v}'_k)\mathrm{p}(\boldsymbol{v}'_k|\mathcal{V}_0)d\boldsymbol{v}'_k \tag{3.29}$$

$$=\prod_{k=1}^{K}\mathrm{p}\big(\boldsymbol{I}_{\{k\}}\big|\mathcal{V}_0\big), \tag{3.30}$$

where $\boldsymbol{I}_{\{k\}}=\{\boldsymbol{i}_n:z_n=k\}$. Note that $\boldsymbol{V}'$ is not required to estimate the vector of latent variables $\boldsymbol{z}$. Straightforward computations lead to

$$\mathrm{p}\big(z_n\big|\boldsymbol{z}_{\backslash n},\boldsymbol{I},\mathcal{V}_0\big)=\frac{N_{z_n}-1}{\alpha+N-1}\prod_{k=1}^{K}\mathrm{p}\big(\boldsymbol{I}_{\{k\}}\big|\mathcal{V}_0\big). \tag{3.31}$$

Removing all factors that do not depend on z_n, we obtain

$$\mathrm{p}\big(z_n\big|\boldsymbol{z}_{\backslash n},\boldsymbol{I},\mathcal{V}_0\big)\propto\begin{cases}\alpha\,\mathrm{p}(\boldsymbol{i}_n|\mathcal{V}_0) & \text{if } z_n=0,\\ N'_{z_n}\frac{\mathrm{p}(\boldsymbol{I}_{\{z_n\}}|\mathcal{V}_0)}{\mathrm{p}(\boldsymbol{I}_{\{z_n\}\backslash n}|\mathcal{V}_0)} & \text{if } 1\leq z_n\leq K,\end{cases} \tag{3.32}$$

where $\boldsymbol{I}_{\{z_n\}\backslash n}=\{\boldsymbol{i}_m:z_m=z_n,m\neq n\}$. Moreover, when $\boldsymbol{I}_{\{k\}}\big|\mathcal{V}_0$ belongs to a distribution family that can be described using a sufficient statistic $\boldsymbol{T}\big(\boldsymbol{I}_{\{k\}}\big)$ such that $\mathrm{p}_{\boldsymbol{I}}\big(\boldsymbol{I}_{\{k\}}\big|\mathcal{V}_0\big)=\mathrm{p}_{\boldsymbol{T}}\big(\boldsymbol{T}\big(\boldsymbol{I}_{\{k\}}\big)\big|\mathcal{V}_0\big)$ (e.g., for distributions belonging to an exponential family), we have $\boldsymbol{T}\big(\boldsymbol{I}_{\{k\}}\big)=\boldsymbol{T}\big(\boldsymbol{I}_{\{k\}\backslash n}\big)+\boldsymbol{T}(\boldsymbol{i}_n)$. This means that (3.32) can be easily computed just by keeping track of $\boldsymbol{T}\big(\boldsymbol{I}_{\{k\}}\big)$ for each cluster.

Algorithm 2 shows the implementation of the described approach, where $j_{\max}$ is the maximum number of samples to generate and $\boldsymbol{T}_k$ is the sufficient statistics for the kth cluster.

Algorithm 2. A collapsed Gibbs sampler to sample from the partition distribution.

Data: $\boldsymbol{I} = [\boldsymbol{i}_1, \ldots, \boldsymbol{i}_N]$, $\mathcal{V}_0$, α, $j_{\max}$

Result: $\boldsymbol{Z} = \left[\left[z_1^{(1)}, \ldots, z_N^{(1)}\right]^T, \ldots, \left[z_1^{(j_{\max})}, \ldots, z_N^{(j_{\max})}\right]^T\right]$

$z_n^{(0)} \leftarrow 0 \; \forall 1 \leq n \leq N$;
$K \leftarrow 0$;
for $j : 1 \leq j \leq j_{max}$ **do**
  $z_n^{(j)} \leftarrow z_n^{(j-1)} \; \forall 1 \leq n \leq N$;
  for $n : 1 \leq n \leq N$, *in random order* **do**
    `/* Remove the n-th pixel from its current class */`
    if $z_n^{(j)} \neq 0$ **then**
      $\boldsymbol{T}_{z_n^{(j)}} \leftarrow \boldsymbol{T}_{z_n^{(j)}} - T(\boldsymbol{i}_n)$; $N_{z_n^{(j)}} \leftarrow N_{z_n^{(j)}} - 1$;
    `/* Sample a new class for the n-th pixel */`
    $p_0 \leftarrow \alpha \, \mathrm{p}_{\boldsymbol{T}}(\boldsymbol{T}(\boldsymbol{i}_n)|\mathcal{V}_0)$;
    $p_k \leftarrow N_k \frac{\mathrm{p}_{\boldsymbol{T}}(\boldsymbol{T}_k + \boldsymbol{T}(\boldsymbol{i}_n)|\mathcal{V}_0)}{\mathrm{p}_{\boldsymbol{T}}(\boldsymbol{T}_k|\mathcal{V}_0)} \; \forall 1 \leq k \leq K$;
    $z_n^{(j)} \sim \mathrm{Cat}\left(\frac{p_0}{\sum_{i=0}^{K} p_i}, \frac{p_1}{\sum_{i=0}^{K} p_i}, \ldots, \frac{p_K}{\sum_{i=0}^{K} p_i}\right)$;
    `/* Place the n-th pixel in its new class */`
    if $z_n^{(j)} = 0$ **then**
      $K \leftarrow K + 1$; $z_n^{(j)} \leftarrow K$;
      $\boldsymbol{T}_K \leftarrow \boldsymbol{T}(\boldsymbol{i}_n)$; $N_K \leftarrow 1$;
    else
      $\boldsymbol{T}_{z_n^{(j)}} \leftarrow \boldsymbol{T}_{z_n^{(j)}} + \boldsymbol{T}(\boldsymbol{i}_n)$; $N_{z_n^{(j)}} \leftarrow N_{z_n^{(j)}} + 1$;

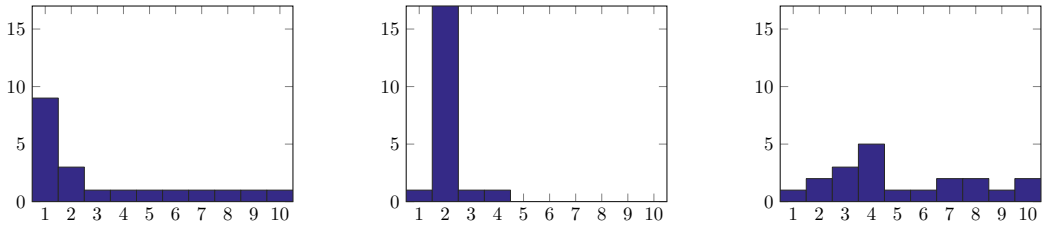


Figure 2. *Label histograms for z_n generated by the proposed algorithm for three different pixels.*

The proposed Gibbs sampler produces $j_{\max}$ samples for each z_{n}. The first $j_{\min}$ iterations are considered as a burn-in period, and the corresponding samples are discarded. For $j_{\min}$ big enough the samples $z_n^{(j_{\min+1})}, \ldots, z_n^{(j_{\max})}$ are distributed according to the joint distribution of the sampled variables and can be used to estimate z_n. Figure 2 shows typical histograms of z_n generated by the algorithm for three z_n (note that the actual numerical value of the variables were replaced by categorical values from 1 to 10 for scale reasons). It can be observed that

the first two cases correspond to pixels that can be easily assigned to a particular class, while the third case corresponds to a pixel whose class is not clearly defined. A point estimation of z_n can be obtained by considering the mode of $z_n^{(j_{\min+1})}, \ldots, z_n^{(j_{\max})}$. However, confidence intervals can also be obtained for z_n and subsequent variables derived from $\hat{z}_n$, such as $\hat{\boldsymbol{v}}_n$ and Δ_n (see section 5 for these derivations).

In order to estimate $\boldsymbol{v}_n$ we consider a maximum a posteriori (MAP) estimator defined as

$$\boldsymbol{v}'_k = \arg\max_{\boldsymbol{v}'_k} \prod_j \prod_{\substack{n \\ z_n^{(j)}=k}} \mathrm{p}(\boldsymbol{i}_n|\boldsymbol{v}'_k)\mathrm{p}(\boldsymbol{v}'_k), \tag{3.33}$$

so that the density of $\boldsymbol{v}_n$ can be similarly estimated using samples $\boldsymbol{v}_n^{(j)} = \boldsymbol{v}_{z_n^{(j)}}$.

3.4. Prior for the concentration parameter. The selection of the concentration parameter α in the model presented in subsection 3.2 has a direct influence on the number of objects detected within an image. The probability $\mathrm{p}(K|\alpha, N)$ has been studied in [1, 18] in the context of a gamma prior for α, leading to

$$\mathrm{p}(K|\alpha, N) = S_N(K)\alpha^K \frac{\Gamma(\alpha)}{\Gamma(N+\alpha)}, \tag{3.34}$$

where $S_N(K)$ is the unsigned Stirling number. We recall that the unsigned Stirling numbers follow the recursive relationship

$$S_N(K) = (N-1)S_{N-1}(K) + S_{N-1}(K-1) \tag{3.35}$$

with the boundary conditions $S_0(0) = 1$ and $S_N(K) = 0$ for $K > N$. Using (3.34), the expected value of K given by (3.13) can be derived. This implies that the choice of α provides information about the resulting number of objects. However, the number of objects within an analysis window depends on several factors, including the image resolution, the window size, and the particular scene being imaged (e.g., a rural area, an urban area). These conditions can be a priori unknown or might exhibit strong variations within the image. To avoid problems arising from an informative prior we propose to define and use a non-informative Jeffreys prior for α. The Jeffreys prior for α associated with (3.34) is [35]

$$\mathrm{p}(\alpha|N) \propto \sqrt{\mathrm{E}_K\left[\left(\frac{d}{d\alpha}\log \mathrm{p}(K|\alpha, N)\right)^2\right]}. \tag{3.36}$$

To compute this expression, we start by deriving $\frac{d}{d\alpha}\log \mathrm{p}(K|\alpha, N)$ using (3.34),

$$\frac{d}{d\alpha}\log \mathrm{p}(K|\alpha, N) = \frac{K}{\alpha} - \left(\Delta\psi_N^{(0)}(\alpha) + \alpha^{-1}\right), \tag{3.37}$$

where $\Delta\psi_N^{(i)}(\alpha) = \psi^{(i)}(N+\alpha) - \psi^{(i)}(1+\alpha)$, $\psi(\cdot)$ is the digamma function, and $\psi^{(i)}(\cdot)$ is its ith derivative. Denoting $m = \Delta\psi_N^{(0)}(\alpha) + \alpha^{-1}$, we obtain

$$\mathrm{E}_K\left[\left(\frac{K}{\alpha} - m\right)^2\right] = \frac{\Gamma(\alpha)}{\Gamma(N+\alpha)} \sum_{K=1}^{N} S_N(K)\alpha^K \left(\frac{K}{\alpha} - m\right)^2. \tag{3.38}$$

Defining the number $\mathcal{S}_N^t$ as the series

$$\mathcal{S}_N^t = \sum_{K=1}^{N} S_N(K)\, \alpha^K K^t, \tag{3.39}$$

the squared binomial in (3.38) can be expanded to obtain

$$\mathrm{E}_K\left[\left(\frac{K}{\alpha} - m\right)^2\right] = \frac{1}{\alpha^2}\mathcal{S}_N^2 - 2\frac{m}{\alpha}\mathcal{S}_N^1 + m^2\mathcal{S}_N^0. \tag{3.40}$$

By using (3.35), we obtain a closed form expression for $\mathcal{S}_N^0$,

$$\mathcal{S}_N^0 = \sum_{K=1}^{N} \alpha^K S_N(K) = (N-1)\mathcal{S}_{N-1}^0 + \alpha\, \mathcal{S}_{N-1}^0 = (N+\alpha-1)\mathcal{S}_{N-1}^0. \tag{3.41}$$

Solving the recursive relationship finally leads to

$$\mathcal{S}_N^0 = \frac{\Gamma(N+\alpha)}{\Gamma(\alpha)}. \tag{3.42}$$

Differentiating (3.39) and (3.42) w.r.t. α, we obtain

$$\mathcal{S}_N^1 = \mathcal{S}_N^0\, \alpha\, m, \tag{3.43}$$

$$\mathcal{S}_N^2 = \mathcal{S}_N^0\big(\alpha\, m + \alpha^2\, m^2 + \alpha^2\, m'\big), \tag{3.44}$$

where $m' = \frac{d\,m}{d\alpha} = \Delta\psi_N^{(1)}(\alpha) - \alpha^{-2}$. This leads to

$$\mathrm{E}_K\left[\left(\frac{K}{\alpha} - m\right)^2\right] = \frac{\Delta\psi_N^{(0)}(\alpha)}{\alpha} + \Delta\psi_N^{(1)}(\alpha) \tag{3.45}$$

yielding the Jeffrey's prior

$$\mathrm{p}(\alpha|N) = \mu_N \sqrt{\frac{\Delta\psi_N^{(0)}(\alpha)}{\alpha} + \Delta\psi_N^{(1)}(\alpha)}, \tag{3.46}$$

where μ_N is a finite normalization constant greater than $(N-1)^{-1}\pi^{-1}$, i.e., $\mathrm{p}(\alpha|N)$ is a proper prior (see Appendix A for a proof of this statement).

To include this prior for α in the Gibbs sampler, we need to compute its conditional distribution. The conditional distribution of α depends exclusively on the partitioning $\boldsymbol{z}$, particularly, on the number of pixels to be partitioned, and the number of clusters in the partition. The following results can be obtained:

$$\begin{aligned} \mathrm{p}(\alpha|K,N) &\propto \mathrm{p}(K|\alpha,N)\,\mathrm{p}(\alpha|N) \\ &\propto S_N(K)\alpha^K \frac{\Gamma(\alpha)}{\Gamma(N+\alpha)}\,\mathrm{p}(\alpha|N) && (3.47) \\ &\propto \alpha^K\,\mathrm{B}(\alpha,N)\,\mathrm{p}(\alpha|N), && (3.48) \end{aligned}$$

where $\mathrm{B}(\alpha, N) = \frac{\Gamma(\alpha)\Gamma(N)}{\Gamma(N+\alpha)}$ is the beta function. As described in [18], we can simplify this expression by introducing a new random variable. Since the beta function can also be expressed as $\mathrm{B}(\alpha, N) = \int_0^1 t^{\alpha-1}(1-t)^{N-1}dt$, we obtain

$$\mathrm{p}(\alpha|K, N) \propto \alpha^K \, \mathrm{p}(\alpha|N) \int_0^1 t^{\alpha-1}(1-t)^{N-1}dt. \tag{3.49}$$

We can interpret (3.49) as a marginal distribution of

$$\mathrm{p}(\alpha, t|K, N) \propto \alpha^K \, t^{\alpha-1}(1-t)^{N-1} \, \mathrm{p}(\alpha|N) \, \mathbf{1}_{[0,1]}(t). \tag{3.50}$$

In the MCMC scheme, we can sequentially sample t and α from

$$\mathrm{p}(t|\alpha, K, N) \propto t^{\alpha-1}(1-t)^{N-1} \, \mathbf{1}_{[0,1]}(t), \tag{3.51}$$

$$\mathrm{p}(\alpha|t, K, N) \propto \alpha^K \, t^{\alpha-1} \, \mathrm{p}(\alpha|N), \tag{3.52}$$

where (3.51) is a beta distribution, and (3.52) is defined as

$$\mathrm{p}(\alpha|t, K, N) \propto \alpha^K \, t^{\alpha-1} \sqrt{\frac{\Delta\psi_N^{(0)}(\alpha)}{\alpha} + \Delta\psi_N^{(1)}(\alpha)}. \tag{3.53}$$

It can be easily shown that the marginal distribution $\mathrm{p}(\alpha|t, K, N)$ behaves as a power law distribution for $\alpha \to 0$ and as a power law distribution with an exponential cutoff for $\alpha \to \infty$, i.e.,

$$\mathrm{p}(t|\alpha, K, N) = L_\infty(\alpha|t, K, N)\alpha^{K-\frac{3}{2}} e^{\alpha \log t} \, \mathbf{1}_{[0,1]}(t), \tag{3.54}$$

$$\mathrm{p}(\alpha|t, K, N) = L_0(\alpha|t, K, N)\alpha^{K-\frac{1}{2}} \, \mathbf{1}_{\mathbb{R}^+}(\alpha), \tag{3.55}$$

where $L_\infty(\alpha|N)$ and $L_0(\alpha|N)$ are slowly varying functions in ∞ and 0, respectively. For the positive scale factor d, we obtain

$$\lim_{\alpha\to\infty} \frac{L_\infty(\alpha|N)}{L_\infty(d\,\alpha|N)} = 1, \tag{3.56}$$

$$\lim_{\alpha\to 0} \frac{L_0(d\,\alpha|N)}{L_0(\alpha|N)} = 1. \tag{3.57}$$

This behavior is difficult to replicate with most well-known distributions, especially due to the mixed power law and exponential cutoff for $\alpha \to \infty$. However, this distribution can be upperbounded by removing the power law behavior for $\alpha \to \infty$. Indeed

$$\mathrm{p}(\alpha|t, K, N) \leq L(\alpha|t, K, N)\alpha^{K-\frac{1}{2}} \, e^{\alpha \log t}, \tag{3.58}$$

where $L(\alpha|t, K, N)$ is a bounded slowly varying function of α in ∞ and 0 and $\alpha^{K-\frac{1}{2}} \, e^{\alpha \log t}$ is a gamma distribution with shape parameter $K + \frac{1}{2}$ and scale parameter $-\frac{1}{\log t}$. A rejection sampling approach [54] can be easily implemented as shown in Algorithm 3 and is included at the end of each iteration in Algorithm 2.

Algorithm 3. Sampling the posterior distribution of α.

Data: α_0, K, N
Result: α

$t \sim \mathrm{B}(\alpha_0, N)$; $M \leftarrow M(N, K, t)$;
$\alpha \sim \Gamma\Big(K + \frac{1}{2}, -\frac{1}{\log t}\Big)$;
while $u \sim \mathrm{U}(0,1)$, $u\,\mathrm{p}_\Gamma(\alpha) < M\,\mathrm{p}(\alpha|N,K,t)$ **do**
    $\alpha \sim \Gamma\Big(K + \frac{1}{2}, -\frac{1}{\log t}\Big)$;

4. A DPMM-based MRF. This section explains how to account for spatial correlation in the parameter estimation for the image model studied in section 3.2. An MRF is introduced as a prior for the latent variables $\boldsymbol{z} = [z_1, \ldots, z_N]$. At this point, it is interesting to mention the work of [16, 41] that defined an MRF for the cluster distribution parameters $\boldsymbol{v}'_k$. However, since we want to prevent sampling from $\boldsymbol{v}'_k$ by using a collapsed Gibbs sampler, we have to apply the MRF to the labels $\boldsymbol{z}$. Subsection 4.1 reviews some MRF concepts needed to build an appropriate prior for the latent variables. Subsection 4.2 explains how to couple this prior with the DPMM model introduced in section 3 and presents the resulting algorithm.

4.1. Markov random field. A classical tool to capture spatial correlation between adjacent pixels of an image is the MRF, which allows a joint distribution to be defined using a neighborhood graph. Let $\boldsymbol{z} = \{z_1, \ldots, z_N\}$ be a group of random variables and G be a simple weighted graph [31]. The vertices of G represent the different random variables z_n, while the weighted edges represent some affinity between the connected random variables. The random vector $\boldsymbol{z}$ is an MRF if the distribution of one variable z_n of this vector conditionally to the other variables $\boldsymbol{z}_{\setminus n}$ is only dependent on the variables belonging to its neighborhood, i.e.,

$$\mathrm{p}\big(z_n \big| \boldsymbol{z}_{\setminus n}\big) = \mathrm{p}\big(z_n \big| \boldsymbol{z}_{\mathrm{ne}(n)}\big), \tag{4.1}$$

where $\boldsymbol{z}_{\mathrm{ne}(n)}$ is the group of random variables that belong to the neighborhood of z_n, i.e., that are connected by an edge to z_n. However, constructing a joint distribution such that its conditional distribution verifies (4.1) is not trivial. In particular, defining the conditional distributions independently can result in an improper joint distribution. The Hammersley–Clifford theorem [30] gives a necessary and sufficient condition ensuring the existence of the joint distribution. This condition states that $\mathrm{p}(\boldsymbol{z})$ should factorize over the cliques $\mathcal{C}$ of G (we recall that a clique of an undirected graph is a subset of vertices where any two vertices are connected by an edge). This is equivalent to defining $\mathrm{p}(\boldsymbol{z}) = \exp\left[H(\boldsymbol{z})\right]$ and requiring that

$$H(\boldsymbol{z}) = \sum_{C \in \mathcal{C}} H_C(\boldsymbol{z}_C), \tag{4.2}$$

where $H(\cdot)$ is the so-called graph cost function, $H_C(\cdot)$ is a local cost function for the clique C, and $\boldsymbol{z}_C = \{z_n : n \in C\}$. Define $H_n(\boldsymbol{z})$ as all the terms in $H(\boldsymbol{z})$ involving z_n, i.e.,

$$H_n(\boldsymbol{z}) = \sum_{C \in \mathcal{C}, n \in C} H_C(\boldsymbol{z}_C) = h_{\mathrm{ne}(n)}\big(\boldsymbol{z}_{\mathrm{ne}(n)}\big) + h_n(z_n), \tag{4.3}$$

where $h_n(z_n)$ is a cost function for the 1-vertex clique associated with the nth vertex and $h_{\text{ne}(n)}(\boldsymbol{z}_{\text{ne}(n)})$ is a cost function associated with its neighborhood. The resulting conditional distribution $\text{p}(z_n|\boldsymbol{z}_{\setminus n})$ can be written

$$\text{p}(z_n|\boldsymbol{z}_{\setminus n}) \propto \exp\left[H_n(\boldsymbol{z})\right]. \tag{4.4}$$

4.2. MRF prior. The random variables $\boldsymbol{z}$ define a partition of the data. They are discrete random variables taking a categorical value associated with the pixel clusters. Since they take categorical values, the sole interaction between z_n and z_m should consist of evaluating whether they have the same value. This is obtained by a cost function that follows a Potts model [47],

$$H(z_n|\boldsymbol{z}_{\setminus n}) = H_n(z_n) + \sum_{m\in\text{ne}(n)} \omega_{nm}\,\mathbf{1}_{z_n}(z_m) = H_n(z_n) + \sum_{\substack{m\in\text{ne}(n)\\ z_n=z_m}} \omega_{nm}, \tag{4.5}$$

where $\mathbf{1}_{z_n}(\cdot)$ is the indicator function, and ω_{nm} is the weight of the edge connecting vertices n and m. However, an arbitrary cost function $H_n(z_n)$ can be chosen. In this work, we have considered $H_n(z_n) = \log \text{p}(z_n|I_n, \boldsymbol{V})$. Since this term does not depend on any z_m for $m \neq n$, this cost function defines a valid MRF with

$$\text{p}(z_n|\boldsymbol{z}_{\setminus n}, I_n, \boldsymbol{V}) \propto \text{p}(z_n|I_n, \boldsymbol{V}) \prod_{\substack{m\in\text{ne}(n)\\ z_n=z_m}} e^{\omega_{nm}}. \tag{4.6}$$

It is interesting to note that by setting all the weights to $\omega_{nm} = 0$ (i.e., by removing spatial correlation), the suggested model reduces to the DPMM model described in section 3.

The next step is to integrate out $\boldsymbol{V}$ from (4.6). Using the properties of a DPMM, the following result is obtained:

$$\int \text{p}(z_n|I_n, \boldsymbol{V})\text{p}(\boldsymbol{V}|\boldsymbol{z}_{\setminus n}, I_{\setminus n}, \mathcal{V}_0)d\boldsymbol{V} \propto \begin{cases} \alpha\,\text{p}(I_n|\mathcal{V}_0) & \text{if } z_n = 0,\\ N'_{z_n}\dfrac{\text{p}(\boldsymbol{I}_{\{z_n\}}|\mathcal{V}_0)}{\text{p}(\boldsymbol{I}_{\{z_n\}\setminus n}|\mathcal{V}_0)} & \text{if } 1 \leq z_n \leq K. \end{cases} \tag{4.7}$$

As a consequence, integrating out $\boldsymbol{V}$ from $\text{p}(z_n|\boldsymbol{z}_{\setminus n}, I_n, \boldsymbol{V})$ leads to

(4.8)

$$\int \text{p}(z_n|\boldsymbol{z}_{\setminus n}, I_n, \boldsymbol{V})\text{p}(\boldsymbol{V}|\boldsymbol{z}_{\setminus n}, I_{\setminus n}, \mathcal{V}_0)d\boldsymbol{V} = \int \text{p}(z_n|I_n, \boldsymbol{V}) \prod_{\substack{m\in\text{ne}(n)\\ z_n=z_m}} e^{\omega_{nm}}\ \text{p}(\boldsymbol{V}|\boldsymbol{z}_{\setminus n}, I_{\setminus n}, \mathcal{V}_0)d\boldsymbol{V}$$

$$\propto \begin{cases} \alpha\,\text{p}(I_n|\mathcal{V}_0) & \text{if } z_n = 0,\\ N'_{z_n}\dfrac{\text{p}(\boldsymbol{I}_{\{z_n\}}|\mathcal{V}_0)}{\text{p}(\boldsymbol{I}_{\{z_n\}\setminus n}|\mathcal{V}_0)} \prod_{\substack{m\in\text{ne}(n)\\ z_n=z_m}} e^{\omega_{nm}} & \text{if } 1 \leq z_n \leq K, \end{cases} \tag{4.9}$$

which is the conditional probability of the DPM-MRF model that has to be included in the partially collapsed Gibbs sampler. It should be noted that the only difference between (4.9) and (3.32) is that the DPM-MRF conditional distribution requires defining a set of weights

relating a pixel with its neighbors and keeping track of which pixels in the neighborhood belong to the same class.

In order to guarantee the MRF homogeneity, the weight relating a pixel to its neighborhood is defined by an isotropic function of the spatial distance between the pixels m and n (denoted as d_{mn}) such that

$$\lim_{d_{mn}\to\infty} \omega(d_{mn}) = 0, \tag{4.10}$$

which means "the more distant two pixels, the less correlated their classes." More precisely, the function we have chosen is a Gaussian pdf that verifies (4.10) and has a maximum for $d_{mn} = 0$, which is expressed as

$$\omega(d_{mn}|\lambda,\sigma) = \lambda \exp\left(-\frac{d_{mn}^2}{\sigma^2}\right), \tag{4.11}$$

where λ controls the influence of the MRF in the model, and σ controls the neighborhood size. For computational simplicity, we have chosen to set $\omega(d_{mn}) = 0$ for $d_{mn} > 5\sigma$.

Algorithm 4 summarizes the resulting parameter estimation algorithm for the DPM-MRF based on a collapsed Gibbs sampler. Note that each cluster keeps track of the pixels it contains though the sufficient statistics $\boldsymbol{T}_k$ and that each pixel $\boldsymbol{i}_n$ keeps track of the influence of the surrounding pixels through $h_{n,k}$. Note also that the DPMM parameter α is estimated from its posterior distribution at the end of the algorithm.

5. Similarity measure. As described in subsection 2.2, the similarity measure obtained in (2.14) is based on the distance between the vectors $\hat{\boldsymbol{v}}'_k$ and the "no-change" manifold $\mathcal{M}$ (the "no-change" manifold is the manifold spanned by vectors $\boldsymbol{v}'_k$ obtained when there is no change). We recall that the variable $\boldsymbol{v}'_k$ is obtained as a MAP estimation from the labels z_n, which are estimated through the BNP-MRF framework. Figure 3 shows a directed acyclic graph highlighting the relationship between the different variables, where $\boldsymbol{v}'_k$, z_n, and α parametrize the BNP description of $\boldsymbol{v}_n$, while $\boldsymbol{z}_{\backslash n}$, σ, and λ parametrize the MRF constraint on the variable z_n.

In previous works [50, 51], we suggested decomposing some images (or some parts of images) not affected by changes in analysis windows producing a set of vectors $\hat{\boldsymbol{v}}'_k$ for each class that were used to learn the "no-change" manifold. This strategy was motivated by the fact that it was not possible to compute vectors $\hat{\boldsymbol{v}}_n$ for individual pixels. More precisely, the similarity measure considered in [50, 51] was computed for an analysis window W as the weighted average of the distances between the vectors $\hat{\boldsymbol{v}}_k$ and the manifold $\mathcal{M}$ leading to the following change detection rule:

$$\Delta_W \underset{H_1}{\overset{H_0}{\gtrless}} \tau^{-1}, \tag{5.1}$$

where

$$\Delta_W = \sum_{k=1}^{K} w_k \mathrm{p}_{\hat{\boldsymbol{v}}}(\hat{\boldsymbol{v}}_k). \tag{5.2}$$

Conversely, the BNP parameter estimation algorithm investigated in this paper produces a hard classification of the image pixels, resulting in an estimated vector $\hat{\boldsymbol{v}}_n$ for each of the

Algorithm 4. A collapsed Gibbs sampler implementing a DPM-MRF parameter estimator with unknown α.

Data: $\boldsymbol{I} = [\boldsymbol{i}_1, \ldots, \boldsymbol{i}_N]$, $\mathcal{V}_0$, λ, σ, $j_{\max}$

Result: $\boldsymbol{Z} = \left[\left[z_1^{(1)}, \ldots, z_N^{(1)}\right]^T, \ldots, \left[z_1^{(j_{\max})}, \ldots, z_N^{(j_{\max})}\right]^T\right]$

$z_n^{(0)} \leftarrow 0 \; \forall \, 1 \leq n \leq N$;
$K \leftarrow 0$;
$\alpha \leftarrow 1$;
$\omega_{nm} \leftarrow \lambda \exp\left(-\frac{d_{mn}^2}{\sigma^2}\right) \; \forall \; 1 \leq n \leq N, \, m \in \delta(n)$;
for $j : 1 \leq j \leq j_{max}$ **do**
    $z_n^{(j)} \leftarrow z_n^{(j-1)} \; \forall \, 1 \leq n \leq N$;
    for $n : 1 \leq n \leq N$, *in random order* **do**
        `/* Remove the n-th pixel from its current class */`
        if $z_n^{(j)} \neq 0$ **then**
            $\boldsymbol{T}_{z_n^{(j)}} \leftarrow \boldsymbol{T}_{z_n^{(j)}} - T(\boldsymbol{i}_n)$; $N_{z_n^{(j)}} \leftarrow N_{z_n^{(j)}} - 1$;
            $h_{m,z_n^{(j)}} \leftarrow h_{m,z_n^{(j)}} - \omega_{mn} \; \forall \, m \in \delta(n)$;
        `/* Sample a new class for the n-th pixel */`
        $p_0 \leftarrow \alpha \, \mathrm{p}_{\boldsymbol{T}}(\boldsymbol{T}(\boldsymbol{i}_n)|\mathcal{V}_0)$;
        $p_k \leftarrow N_k \frac{\mathrm{p}_{\boldsymbol{T}}(\boldsymbol{T}_k + \boldsymbol{T}(\boldsymbol{i}_n)|\mathcal{V}_0)}{\mathrm{p}_{\boldsymbol{T}}(\boldsymbol{T}_k|\mathcal{V}_0)} \, e^{h_{n,k}} \; \forall \, 1 \leq k \leq K$;
        $z_n^{(j)} \sim \mathrm{Cat}\left(\frac{p_0}{\sum_{i=0}^K p_i}, \frac{p_1}{\sum_{i=0}^K p_i}, \ldots, \frac{p_K}{\sum_{i=0}^K p_i}\right)$;
        `/* Place the n-th pixel in its new class */`
        if $z_n^{(j)} = 0$ **then**
            $K \leftarrow K + 1$; $z_n^{(j)} \leftarrow K$;
            $\boldsymbol{T}_K \leftarrow \boldsymbol{T}(\boldsymbol{i}_n)$; $N_K \leftarrow 1$;
            $h_{m,K} \leftarrow \omega_{mn} \; \forall \, m \in \delta(n)$;
            $h_{m,K} \leftarrow 0 \; \forall \, m \notin \delta(n)$;
        else
            $\boldsymbol{T}_{z_n^{(j)}} \leftarrow \boldsymbol{T}_{z_n^{(j)}} + \boldsymbol{T}(\boldsymbol{i}_n)$; $N_{z_n^{(j)}} \leftarrow N_{z_n^{(j)}} + 1$;
            $h_{m,z_n^{(j)}} \leftarrow h_{m,z_n^{(j)}} + \omega_{mn} \; \forall \, m \in \delta(n)$;
    `/* Estimate the new α */`
    $t \sim \mathrm{B}(\alpha, N)$; $M \leftarrow M(N, K, t)$;
    $\alpha \sim \Gamma\left(K + \frac{1}{2}, -\frac{1}{\log t}\right)$;
    while $u \sim \mathrm{U}(0,1)$, $u \, \mathrm{p}_\Gamma(\alpha) < M \, \mathrm{p}(\alpha|N, K, t)$ **do**
        $\alpha \sim \Gamma\left(K + \frac{1}{2}, -\frac{1}{\log t}\right)$;

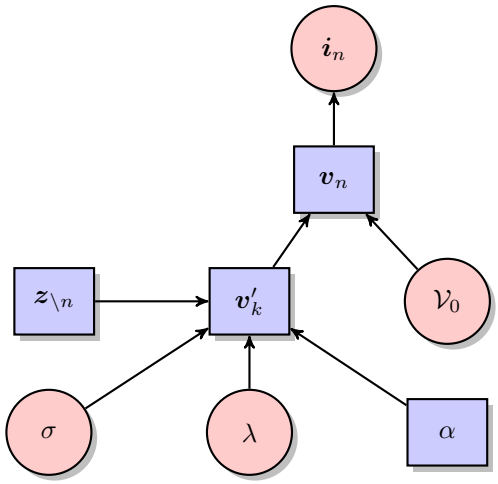


Figure 3. *Directed acyclic graph highlighting the relationship between the different variables (variables inside circles are fixed, whereas those in squares are estimated by the proposed approach).*

N pixels in W. This remark suggests that the following BNP change detection strategy can be applied on a pixel by pixel level:

$$\Delta_n \underset{H_1}{\overset{H_0}{\gtrless}} \tau^{-1}, \tag{5.3}$$

where

$$\Delta_n = \mathrm{p}_{\hat{v}}(\hat{\boldsymbol{v}}_n) \tag{5.4}$$

and where $\hat{\boldsymbol{v}}_n$ is the estimated parameter vector associated with the pixel n (which is shared by all the pixels belonging to the same class). A direct consequence of the new detector defined by (5.3) is the improved resolution of the final change detection map, as will be illustrated in the experiments presented in the next section.

In order to evaluate the hypothesis test defined by (5.3), a point estimation of the test statistic can be obtained using the last L iterations of the Gibbs sampler as $\Delta_n = \frac{1}{L}\sum_l \Delta_n^{(l)}$ (where $\Delta_n^{(l)} = \mathrm{p}_{\hat{v}}(\hat{\boldsymbol{v}}_n^{(l)})$), which is the approach adopted in this paper. However, other statistics can also be obtained. For instance, the cumulative distribution of Δ_n conditionally to each hypothesis can be obtained as

$$\mathrm{p}(\Delta_n < \tau | H_i) = \frac{1}{L}\sum_{l=1}^{L} \mathbf{1}_{\Delta<\tau}\left(\Delta_n^{(l)}\right), \tag{5.5}$$

where $\mathbf{1}_A(\cdot)$ is the indicator function. This cumulative distribution is shown in Figure 4 for three particular pixels. The distribution of $\Delta_n|H_i$ is obtained from the values of z_n sampled from the Gibbs sampler, and thus, the number of steps observed in the CDF are dependent on the number of considered iterations. Figures 4(a) and 4(b) correspond to pixels in unchanged

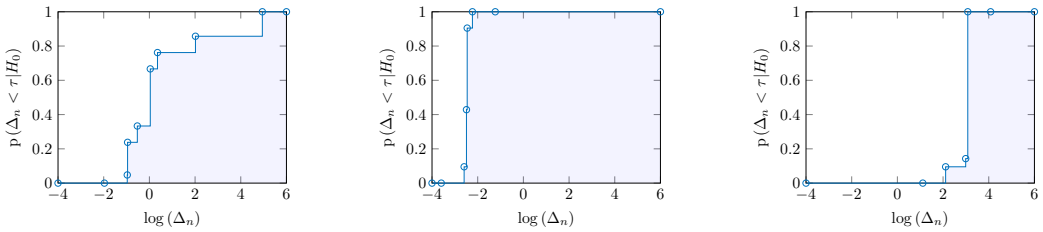


Figure 4. *Cumulative distribution $p(\Delta_n < \tau|H_0)$ for three different pixels. Figures 4(a) and 4(b) were obtained from pixels in unchanged areas, while Figure 4(c) was obtained from a pixel in a changed area.*

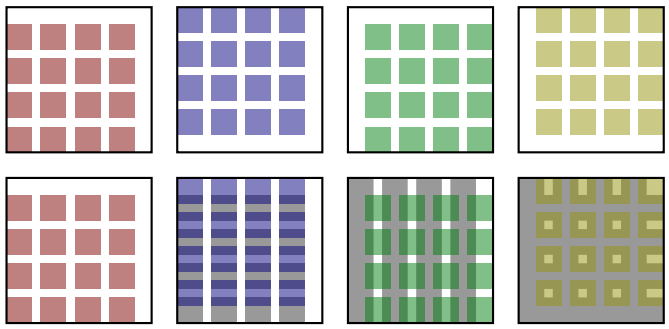


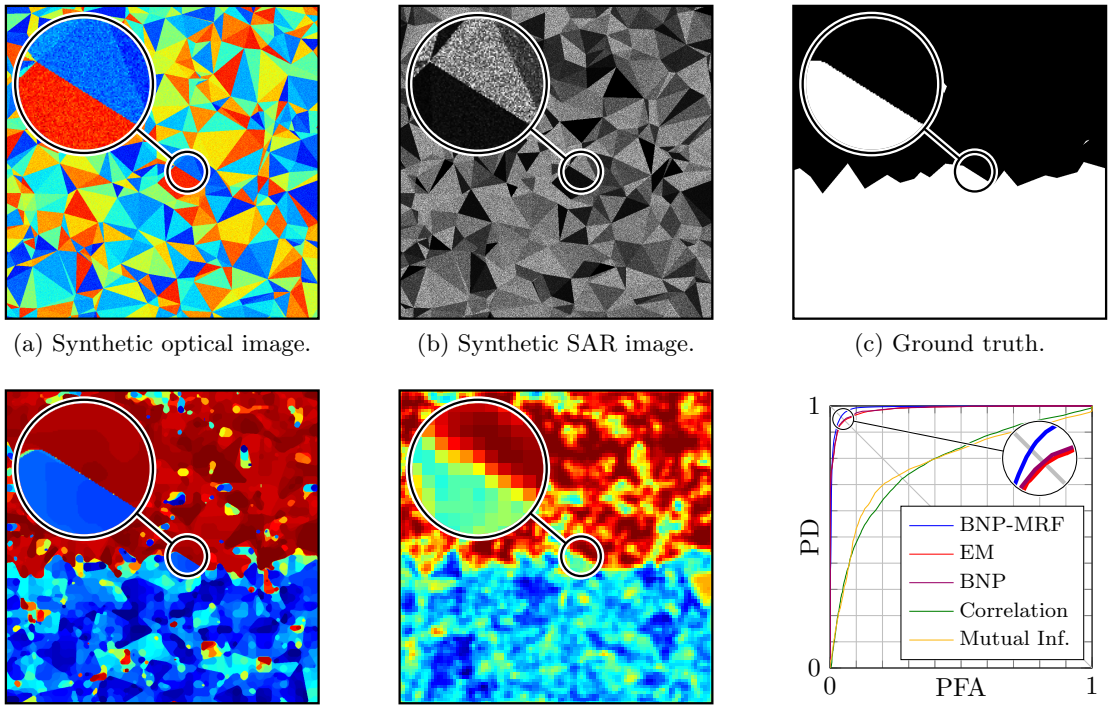
Figure 5. *Decomposition of the image into overlapping windows (first row). The windows of each color can be computed in parallel with the strategy in the second row.*

areas, while Figure 4(c) corresponds to a pixel in a changed area. This can be observed in the former providing a higher probability of H_0 (no change) than the latter for almost all values of τ. Moreover, it can also be observed that the middle plot is easier to classify as unchanged than the first one.

6. Simulation results. This section presents simulation results obtained for different datasets. The change detection results are compared with those obtained with different methods, namely, correlation coefficient [33], mutual information [33], conditional copulas [38] (when available), and a previous version of the method presented in this paper based on the EM algorithm [51].

It is important to note that the computational cost of the proposed algorithm as well as the required memory do not allow very large images to be processed (for a neighborhood size of 30×30 pixels, the maximum practical image size is about 300×300 pixels on a computer with 8 GB of memory). Instead, the image is decomposed into a set of overlapping windows and the proposed algorithm is run for each of these windows. More precisely, the adopted approach was constructed considering parallelization of the problem by assuming that the interpixel dependency decreases with the distance between the pixels. It involves the following steps:

- Dividing the image into a grid of overlapping windows. In our simulations, we have considered four sets of nonoverlapping windows as depicted in the first row of Figure 5.



(a) Synthetic optical image. (b) Synthetic SAR image. (c) Ground truth.

(d) Distance Δ_n. (e) Distance Δ_W. (f) Comparison of the detector performance.

Figure 6. *Results obtained by applying the proposed method with a dataset consisting of two synthetic heterogeneous optical and SAR images.*

- Running the BNP-MRF algorithm for all the windows in each set, as depicted in the second row of Figure 5 and explained below:
    1. Run the algorithm on all the red windows. Note that these windows can be analyzed in parallel since they do not overlap.
    2. Run the algorithm on the light blue pixels, without resampling the pixels already considered in the previous step (dark blue and gray areas). Note that this sampling requires using the pixels belonging to the dark blue areas (which have been processed in the first step).
    3. Repeat for windows in the green set.
    4. Repeat for windows in the yellow set.
- Merging the samples of z_n obtained for all windows to obtain the whole image.

It is important to note that this approach does not allow the use of overlaps $\geq$50% of the window size. Thanks to the use of overlapping windows, the resulting label map z_n is spatially smooth, as opposed to nonoverlapping windows.

6.1. Synthetic data. Figure 6 presents the simulation results obtained on a dataset consisting of two heterogeneous synthetic images. Figure 6(a) is a synthetic optical image, while Figure 6(b) is a synthetic image representing the intensity channel of a SAR image. Figure 6(c) displays the ground truth showing in black the areas of the image that have been affected by changes.

To obtain Δ_n, the labels z_n were computed using windows of size 100×100, which was chosen to optimize the processing time. The neighborhood graph defined by $\omega(d_{mn}|\lambda, \sigma)$ was obtained with $\lambda = 30$ and $\sigma = 6$, with $\omega(d_{mn}) = 0$ for $d_{mn} \geq 30$. This choice of σ is due to the fact that the scene contains mostly big objects. A second test was performed with a neighborhood size of 0 (which removes all the effects produced by the MRF) in order to analyze the incidence of the MRF in the proposed model (this test is denoted as BNP in Figure 6(f)). The windows were processed using an overlap of 30 pixels in order to ensure the continuity of the spatial correlation in z_n. The values $\hat{\boldsymbol{v}}_n$ were obtained by computing the maximum likelihood estimator of $\boldsymbol{v}_n$ using $\boldsymbol{I}_{\{z_n\}}$, while the manifold $\mathcal{M}$ was estimated using 1% of the pixels randomly selected from unchanged areas.

Figure 6(d) shows Δ_n, the computed distance to the manifold $\mathcal{M}$, while Figure 6(e) shows the corresponding distance Δ_W obtained with the change detection method studied in [50, 51], where red corresponds to a large value of Δ_n and Δ_W, respectively. It can be observed that Figure 6(d) provides a more accurate change detection compared to Figure 6(e), which is highlighted in the circular zoom area. Figure 6(f) compares the receiver operating characteristics (ROC) [46] obtained with our method and with other classical methods.[1] This comparison leads to the following conclusions:

- The proposed BNP-MRF model provides better results than the model introduced in [50, 51] (referred to as EM) for this example.
- The BNP framework without the MRF does not provide any significant improvement when compared to the EM method with a number of clusters optimized for this image. However, the BNP model does not require a careful selection of a minimum and maximum number of clusters.
- The introduction of the MRF provides a significant improvement for the change detection performance.

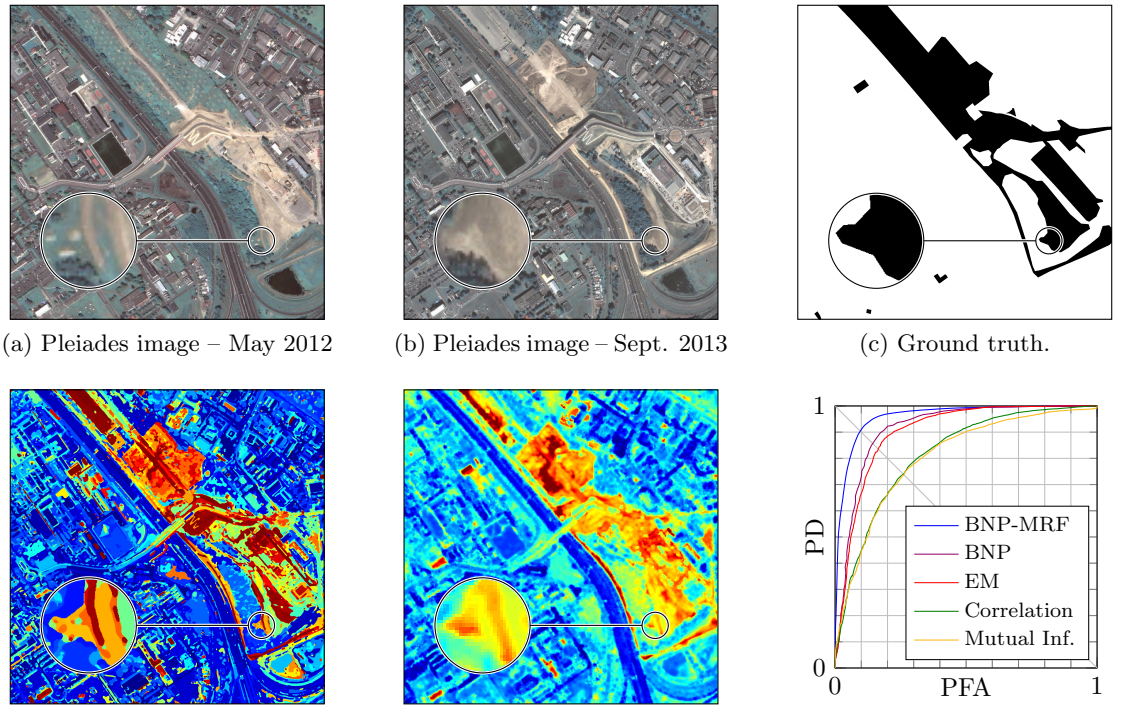
If we consider the working situation defined by equal PFA and probability of nondetection $(1 - \text{PD})$, the method in [50, 51] provides an error rate of 5.52%, while the method proposed in this paper yields an error rate of 4.18%, which represents a reduction of 24%.

6.2. Real data.

6.2.1. Real homogeneous optical images. The first experiment considers a pair of images from an urban area in the south of Toulouse, France, acquired by Pleiades satellites within a time period of 16 months and characterized by some new constructions. Figures 7(a) and 7(b) display multispectral images with four spectral bands (blue, green, red and infrared) obtained by pansharpening the 2-m resolution multispectral image using the 50-cm resolution panchromatic image obtained by the satellite. Figure 7(c) shows the ground truth provided by a photo-interpreter indicating in black the areas affected by changes.

To obtain Δ_n, the labels z_n were computed on windows of size 200×200, which was determined based on the system memory and processing power. The neighborhood graph defined by $\omega(d_{mn}|\lambda, \sigma)$ was obtained with $\lambda = 60$ and $\sigma = 1$, with $\omega(d_{mn}) = 0$ for $d_{mn} \geq 5$. The choice of a small value for σ is due to the fact that the scene corresponds to an urban area consisting

[1] These results were obtained using window sizes optimized by cross validation, to produce the best performance for each method. More precisely, we obtained window sizes of 20×20 for [51] and 40×40 for the correlation coefficient and the mutual information, in all cases with an overlap of 50%.



(a) Pleiades image – May 2012 (b) Pleiades image – Sept. 2013 (c) Ground truth.

(d) Distance Δ_n. (e) Distance Δ_W. (f) Comparison of the detector performance.

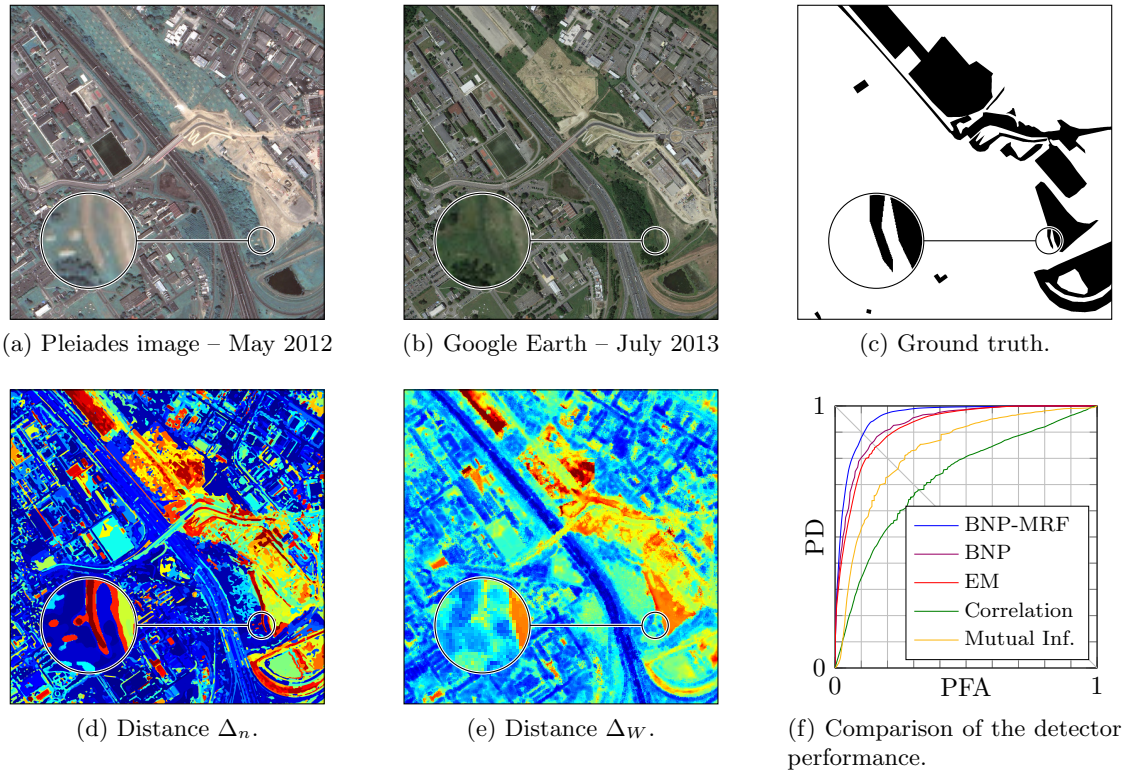
Figure 7. *Results obtained by applying the proposed method with a dataset consisting of two real homogeneous optical images.*

mostly of small objects. A second test was performed with a neighborhood size of 0 (which removes all the effects produced by the MRF) in order to analyze the incidence of the MRF in the proposed model (this test is denoted as BNP in Figure 7(f)). The windows were processed using an overlap of five pixels in order to ensure the continuity of the spatial correlation in z_n. The values $\hat{\boldsymbol{v}}_n$ were obtained by computing the maximum likelihood estimator of $\boldsymbol{v}_n$ using $\boldsymbol{I}_{\{z_n\}}$, while the manifold $\mathcal{M}$ was estimated using 1% of the pixels randomly selected from unchanged areas.

Figures 7(d) and 7(e) show the estimated distances Δ_n and Δ_W obtained with the proposed BNP-MRF model, and the model of [51] respectively.[2] It can be observed that Figure 7(d) provides a more accurate change detection compared to Figure 7(e), which is highlighted in the circular zoom area. This improved performance is confirmed in the ROCs displayed in Figure 7(f), leading to the following conclusion:

- Unlike the previous example in Figure 6(f), the introduction of the BNP framework without the MRF provides a slight improvement compared to the EM method with an optimized number of clusters. This can be explained by a higher variance in the size of the objects contained in the image.

[2]These results were obtained using window sizes optimized by cross validation, to produce the best performance for each method. More precisely, we obtained window sizes of 40×40 for [51], for the correlation coefficient and the mutual information, in all cases with an overlap of 50%.



(a) Pleiades image – May 2012 (b) Google Earth – July 2013 (c) Ground truth.

(d) Distance Δ_n. (e) Distance Δ_W. (f) Comparison of the detector performance.

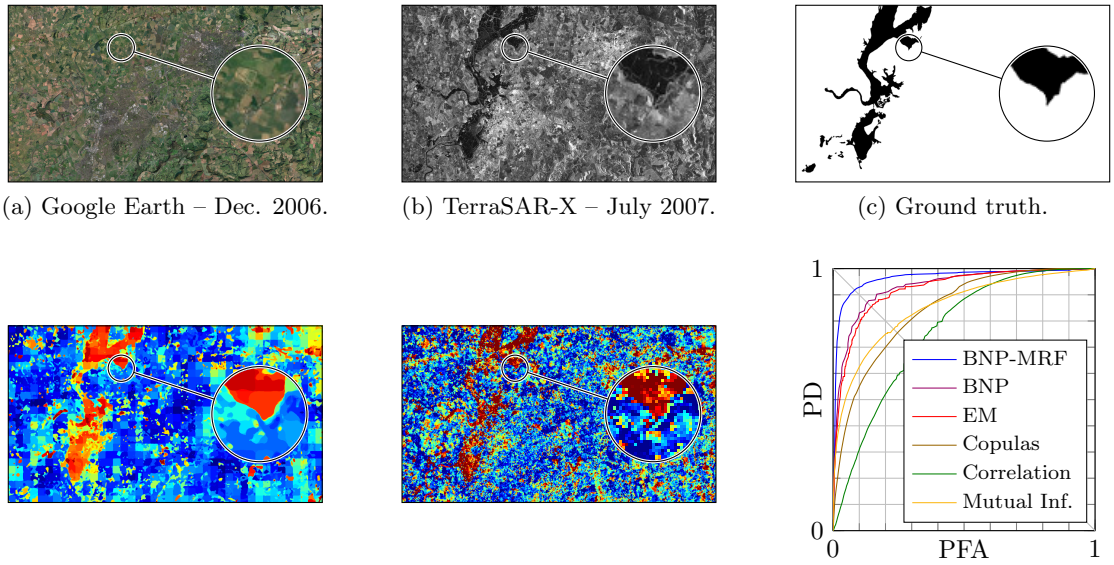
Figure 8. *Results obtained by applying the proposed method with a dataset consisting of two real heterogeneous optical images.*

- As for synthetic data, the BNP-MRF model provides a significant improvement for the change detection when compared to the BNP and EM approaches.

If we consider the working situation defined by PFA $= 1 -$ PD, the method in [50, 51] leads to an error rate of 16.5%, while the proposed method yields an error rate of 9.62%, which represents a reduction of 42%.

It is interesting to note that since the images are homogeneous, the pixel intensity of both images is linearly dependent. This remark explains why the correlation coefficient and the mutual information perform very similarly for this example.

6.2.2. Real heterogeneous optical images. The second simulation studies a pair of images from a urban area, in the same geographical position as the dataset used in subsection 6.2.1 also characterized by some construction works made within a 14-month time interval. Figure 8(a) is identical to Figure 7(a), while Figure 8(b) is a three-channel image obtained from Google Earth which has been downsampled to match the 50-cm resolution of the Pleiades image. Figure 8(c) shows the ground truth provided by a photo-interpreter indicating in black the areas affected by changes. The estimated values of Δ_n are displayed in Figure 8(d). Note that the simulation scenario considered to obtain z_n, Δ_n, and $\hat{\boldsymbol{v}}_n$ was the same as the one used in subsection 6.2.1.



(a) Google Earth – Dec. 2006. (b) TerraSAR-X – July 2007. (c) Ground truth.

(d) Distance Δ_n. (e) Distance Δ_W. (f) Comparison of the detector performance.

Figure 9. *Results obtained by applying the proposed method with a dataset consisting of two real heterogeneous optical and SAR images.*

Figures 8(d) and 8(e) show the estimated distances Δ_n and Δ_W obtained with the proposed BNP-MRF model and the model of [50, 51], respectively.[2] It can be observed that Figure 8(d) provides an improved change detection compared to Figure 8(e), which is highlighted in the circular zoom area. As in the previous datasets, a second test was performed with a neighborhood size of 0 (i.e., without MRF) in order to analyze the incidence of the MRF in the proposed model (this test is denoted as BNP in Figure 8(f)). Figure 8(f) compares the ROCs obtained with the different methods, highlighting the interest of the proposed BNP-MRF model. If we consider the working situation defined by PFA $= 1 - \text{PD}$, the method studied in [50, 51] yields an error rate of 15.6%, while the proposed method leads to an error rate of 10.7%, which represents a reduction of 31%.

It should be noted that since these images are heterogeneous, the pixel intensity in both images is not necessarily linearly dependent. This results in a significant performance reduction when the correlation coefficient is used as a similarity measure. However, the performance of the proposed change detector as well as the other two methods, which are adapted to heterogeneous images, is very similar to what was obtained in the first experiments.

6.2.3. Real heterogeneous optical and SAR images. The last experiment considers a pair of images from a mixture of urban and rural areas, near Gloucester before and during flooding. Figure 9(b) is the intensity channel of an image captured by a TerraSAR-X satellite with pixel resolution of 7.3 m. Figure 9(a) is a three-channel image obtained from Google Earth which has been downsampled to match the pixel resolution of the TerraSAR-X image. Figure 9(c) presents the ground truth provided by a photo-interpreter indicating in black the areas affected by the flooding.

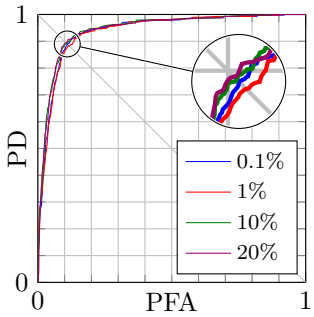


Figure 10. *Result of the change detection method on the Pleiades dataset using different percentages of the unchanged data for training.*

To obtain Δ_n, the labels z_n were computed on windows of size 200×200, which was determined based on the system memory and processing power. The neighborhood graph defined by $\omega(d_{mn}|\lambda, \sigma)$ was obtained with $\lambda = 60$ and $\sigma = 6$, with $\omega(d_{mn}) = 0$ for $d_{mn} \geq 30$. The choice of σ was motivated by the fact that the scene corresponds mostly to a rural area with big homogeneous regions. As in the previous datasets, a second test was performed with a neighborhood size of 0 (i.e., without MRF) in order to analyze the incidence of the MRF in the proposed model (this test is denoted as BNP in Figure 9(f)). The estimates $\hat{\boldsymbol{v}}_n$ were obtained by computing the maximum likelihood estimator of $\boldsymbol{v}$ using $\boldsymbol{I}_{\{z_n\}}$, while the manifold $\mathcal{M}$ was estimated using 1% of the pixels randomly selected from unchanged areas.

Figures 9(d) and 9(e) show the estimated distances Δ_n and Δ_W obtained with the proposed BNP-MRF model and the model of [50, 51], respectively.[3] It can be observed that Figure 9(d) provides a better change detection compared to Figure 9(e), which is highlighted in the circular zoom area. Figure 9(f) shows the ROCs obtained with our method and with alternative strategies showing the interest of the proposed BNP-MRF strategy. If we consider the working situation where $\text{PFA} = 1 - \text{PD}$, the method of [50, 51] yields an error rate of 14.6%, while the proposed approach leads to an error rate of 8.19%, which represents a reduction of 44%.

6.3. Influence of the training data. In the simulation results presented previously, a small percentage of the pixels from unchanged regions was used to learn the manifold. This section studies the influence of the size of the data fraction selected to learn the manifold.

For this purpose, the manifold associated with the dataset displayed in Figure 7 was learned using 0.1%, 1%, 10%, and 20% of the pixels contained in unchanged areas. Figure 10 shows that the ROCs obtained in the different cases do not change significantly and the detector performance remains stable as long as all kinds of unchanged objects within the image are well represented.

6.4. Influence of the Jeffreys prior for α. In [18] a gamma prior was assigned to the parameter α for its inference. However, the parameters of this gamma prior (usually referred

[3]These results were obtained using window sizes optimized by cross validation, to produce the best performance for each method. More precisely, we obtained window sizes of 10×10 for [51] and 20×20 for the correlation coefficient and the mutual information, in all cases with an overlap of 50%, and finally windows of size 9×9 for [38] with an overlap of eight pixels.

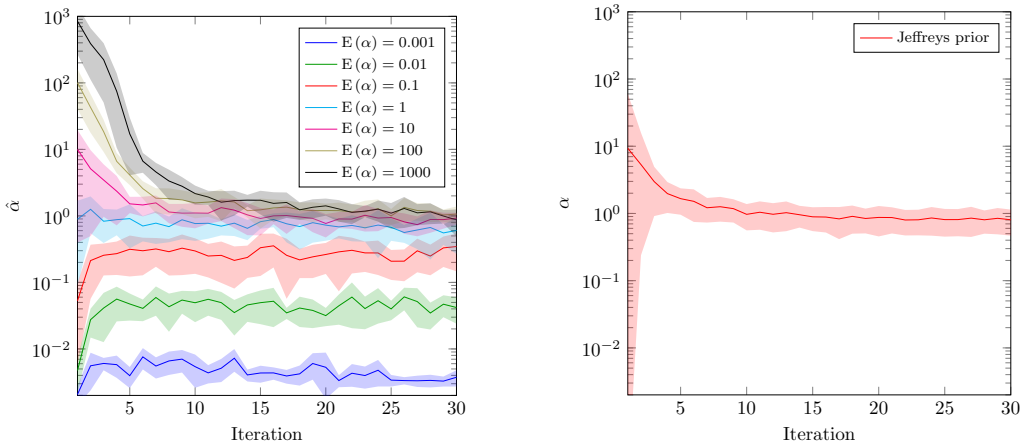


Figure 11. *Influence of the Jeffreys prior compared to the different parameters of the gamma prior. These results were obtained for a synthetic image generated with* $\alpha = 1$.

to as hyperparameters) have an influence on the resulting number of clusters K detected within the image. This section compares the quality of the estimation of parameter α when this parameter is assigned a gamma prior or the Jeffreys prior (3.46). Let us point out here that the Jeffreys prior derived in subsection 3.4 as well as the normalization proof provided in Appendix A correspond to the BNP model of section 3 without the MRF extension. For this comparison, a synthetic image was generated as in subsection 6.1 with $\alpha = 1$. The results of 10 Monte Carlo runs of the proposed BNP algorithm were used to estimate the clusters associated with this dataset. Moreover, the samples of α were used to estimate this parameter (using the minimum mean square estimator) and to compute the corresponding confidence intervals. Several runs of the proposed BNP algorithm were executed on the synthetic data to estimate its clusters, and the sampled values of α were registered. Figure 11 displays the results obtained for different values of the mean of the gamma prior (denoted as E(α)) and for the Jeffreys prior. The parameters of the gamma prior were chosen to obtain different values of E(α) while maintaining a variance of 1 (i.e., a big enough variance). The color lines show the average evolution of $\hat{\alpha}$ for each iteration (whereas the shaded areas display confidence intervals defined by mean ± 3 standard deviations). The gamma prior provides good estimates of α when E[α] is bigger than the actual value of α. However, it fails for other parameter values. Conversely, the proposed Jeffreys prior does not require any hyperparameter to be adjusted and provides an accurate estimation of α. This outlines the advantage of the noninformative Jeffreys prior w.r.t. to the gamma prior used in [1, 18]. Note again that the Jeffreys prior studied in this section does not consider the MRF part of the BNP model. However, there is empirical evidence that this prior is weakly informative and leads to reasonable parameter estimates.

The mixing of the proposed sampler parameter can be assessed through the potential scale reduction factor (PSRF) [25], where a PSRF close to 1 indicates a good mixing (a value of the PSRF below 1.2 is recommended in [24, p. 332]). Figure 12 shows the PSRF values obtained for the simulations in Figure 11 starting from the 10th iteration. In this case, the first half of

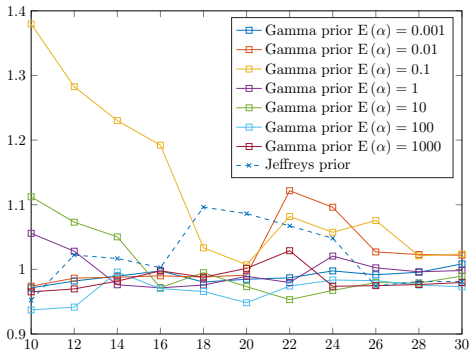


Figure 12. *PSRFs for the parameter α using different priors.*

the iterations is considered as burn-in, and the second half is used to compute the PSRF. As can be seen, all these PSRF values confirm good mixing of the proposed sampler.

A possible approach that would enable the use of a gamma prior is to systematically choose a set of parameters that overestimate α. However, bigger values of α increase the number of model classes K, producing a significant negative impact on the computation time (from a few minutes for $\mathrm{E}[\alpha] \leq 1$ to a whole day for $\mathrm{E}[\alpha] = 1000$) since K posteriors have to be evaluated for each pixel at every iteration.

7. Conclusion. A statistical model was introduced in [50, 51] for detecting changes between different remote sensing images possibly acquired by heterogeneous sensors. This model was based on a mixture distribution assuming that each estimation window contains a finite number of homogeneous areas. The relations between the pixel intensities associated with the images were defined by using a manifold whose parameters can be estimated using no-change areas. This paper improved this change detection strategy by using several modifications. First, a nonparametric approach was introduced to estimate the number of mixture components belonging to each estimation window. Second, a noninformative prior was derived for the concentration parameter of the resulting mixture model. Including this prior in the change detection model of [50, 51] yielded interesting results obtained for different window sizes and for different numbers of objects contained in these analysis windows. Finally, an MRF was coupled with the BNP model to account for the spatial correlation present in remote sensing images. These three modifications increased the robustness of the parameter estimation algorithm, allowing bigger sizes for the analysis windows to be considered and thus improving the parameter estimation accuracy. Moreover, the introduction of a Potts model led to a classification pixel map allowing a pixel-level change detection strategy. The change detection rule resulting from the proposed BNP model showed improved detection performance when compared with existing strategies.

Theoretical prospects include the derivation of the Jeffreys prior associated taking into account the MRF within the BNP model. It would also be interesting to study the effect of the smoothness parameter of the MRF on the detection results. This paper focused on detecting changes between optical and SAR images from the image intensities. However, the proposed

approach might also be interesting for other kinds of images such as hyperspectral images, or for features computed from these images (i.e., wavelet coefficients). These problems are currently under investigation. The application of the proposed methodology to heterogeneous image registration, segmentation, or fusion is also an interesting prospect.

Appendix A. Proof that the Jeffreys prior of α is proper. The posterior arising from the Jeffrey's prior is normalizable. This is a direct consequence that $\mathrm{p}(\alpha, N)$ is a proper prior. We were not able to find an analytic expression for this normalization constant, but an upper bound can be obtained as follows. Considering that

$$\mathrm{p}(\alpha|N) \propto \sqrt{\frac{\psi(N+\alpha)-\psi(\alpha)}{\alpha} + \psi_1(N+\alpha) - \psi_1(\alpha)} \tag{A.1}$$

$$\propto \sqrt{\frac{m}{\alpha} + m'}, \tag{A.2}$$

where $\psi(\cdot)$ is the digamma function, $\psi_1(\cdot)$ is the first derivative of $\psi(\cdot)$, and $m = \psi(N+\alpha) - \psi(\alpha)$, we can prove that $\int_0^\infty \mathrm{p}(\alpha|N)$ is bounded. The squared probability $[\mathrm{p}(\alpha|N)]^2$ can be expressed as

$$\frac{m}{\alpha} + m' = \sum_{K=0}^{N-1} \frac{\alpha^{-1}}{K+\alpha} - \sum_{K=0}^{N-1} \frac{1}{(K+\alpha)^2} \tag{A.3}$$

$$= \sum_{K=0}^{N-1} \frac{\alpha^{-1}K}{(K+\alpha)^2} = \sum_{K=1}^{N-1} \frac{\alpha^{-1}K}{(K+\alpha)^2}, \tag{A.4}$$

where the triangular inequality can be applied to obtain a bound on $\mathrm{p}(\alpha)$ as follows:

$$\sqrt{\frac{m}{\alpha} + m'} \leq \sum_{K=1}^{N-1} \sqrt{\frac{\alpha^{-1}K}{(K+\alpha)^2}}, \tag{A.5}$$

$$\int_0^\infty \sqrt{\frac{m}{\alpha} + m'}\, d\alpha \leq \sum_{K=1}^{N-1} \int_0^\infty \sqrt{\frac{\alpha^{-1}K}{(K+\alpha)^2}}\, d\alpha, \tag{A.6}$$

$$\int_0^\infty \sqrt{\frac{m}{\alpha} + m'}\, d\alpha \leq (N-1)\,\pi, \tag{A.7}$$

where the last equality has been obtained using $\int_0^\infty \sqrt{\frac{\alpha^{-1}K}{(K+\alpha)^2}}\, d\alpha = \pi \;\forall\; K > 0$.

REFERENCES

[1] C. E. Antoniak, *Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems*, Ann. Statist., 2 (1974), pp. 1152–1174.

[2] Y. Bazi, L. Bruzzone, and F. Melgani, *An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR images*, IEEE Trans. Geosci. Remote Sens., 43 (2005), pp. 874–887.

[3] D. Blackwell and J. MacQueen, *Ferguson distributions via Pólya urn schemes*, Ann. Statist., (1973), pp. 353–355.

[4] L. Bruzzone and D. Fernández Prieto, *An iterative technique for the detection of land-cover transitions in multitemporal remote-sensing images*, IEEE Trans. Geosci. Remote Sens., 35 (1997), pp. 858–867.
[5] L. Bruzzone and D. Fernández Prieto, *Automatic analysis of the difference image for unsupervised change detection*, IEEE Trans. Geosci. Remote Sens., 38 (2000), pp. 1171–1182.
[6] L. Bruzzone and D. Fernández Prieto, *An adaptive semi parametric and context-based approach to unsupervised change detection in multitemporal remote-sensing images*, IEEE Trans. Image Process., 11 (2002), pp. 452–466.
[7] C. Carincotte, S. Derrode, and S. Bourennane, *Unsupervised change detection on SAR images using fuzzy hidden Markov chains*, IEEE Trans. Geosci. Remote Sens., 44 (2006), pp. 432–441.
[8] W. G. Carrara, R. S. Goodman, and R. M. Majewski, *Spotlight Synthetic Aperture Radar: Signal Processing Algorithms*, Artech House, Boston, 1995.
[9] G. Casella and E. I. George, *Explaining the Gibbs sampler*, Amer. Statist., 46 (1992), pp. 167–174.
[10] K. R. Castleman, *Digital Image Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1996.
[11] T. Celik, *Multiscale change detection in multitemporal satellite images*, IEEE Geosci. Remote Sens. Lett., 6 (2009), pp. 820–824.
[12] T. Celik, *Unsupervised change detection in satellite images using principal component analysis and k-means clustering*, IEEE Geosci. Remote Sens. Lett., 6 (2009), pp. 772–776.
[13] T. Celik and M. Kai-Kuang, *Unsupervised change detection for satellite images using dual-tree complex wavelet transform*, IEEE Trans. Geosci. Remote Sens., 48 (2010), pp. 1199–1210.
[14] M. Chabert and J.-Y. Tourneret, *Bivariate Pearson distributions for remote sensing images*, in Proceedings of IGARSS, Vancouver, 2011, pp. 4038–4041.
[15] F. Chatelain, J.-Y. Tourneret, A. Ferrari, and J. Inglada, *Bivariate gamma distributions for image registration and change detection*, IEEE Trans. Image Process., 16 (2007), pp. 1796–1806.
[16] S. Chatzis and G. Tsechpenakis, *The infinite hidden Markov random field model*, IEEE Trans. Neural Netw., 21 (2010), pp. 1004–1014.
[17] J. C. Curlander and R. N. McDonough, *Synthetic Aperture Radar: Systems and Signal Processing*, Wiley Ser. Remote Sensing Image Process., Wiley, New York, 1991.
[18] M. D. Escobar and M. West, *Bayesian density estimation and inference using mixtures*, J. Amer. Statist. Assoc., 90 (1995), pp. 577–588.
[19] T. Ferguson, *A Bayesian analysis of some nonparametric problems*, Ann. Statist., (1973), pp. 209–230.
[20] R. Fjortoft, A. Lopés, P. Marthon, and E. Cubero-Castan, *An optimal multiedge detector for SAR image segmentation*, IEEE Trans. Geosci. Remote Sens., 36 (1988), pp. 793–802.
[21] V. S. Frost, J. Abbott Stiles, K. S. Shanmugan, and J. Holtzman, *A model for radar images and its application to adaptive digital filtering of multiplicative noise*, IEEE Trans. Pattern Anal. Mach. Intell., 4 (1982), pp. 157–166.
[22] T. Fung, *An assessment of TM imagery for land-cover change detection*, IEEE Trans. Geosci. Remote Sens., 28 (1990), pp. 681–684.
[23] H. Ge, Y. Chen, M. Wan, and Z. Ghahramani, *Distributed inference for Dirichlet process mixture models*, in Proceedings of the International Conference on Machine Learning, Lille, France, 2015, pp. 2276–2284.
[24] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*, Chapman & Hall/CRC Texts Stat. Sci. Ser., Chapman & Hall/CRC, Boca Raton, FL, 1995.
[25] A. Gelman and D. B. Rubin, *Inference from iterative simulation using multiple sequences*, Statist. Sci., 7 (1992), pp. 457–472.
[26] S. J. Gershman and D. M. Blei, *A tutorial on bayesian nonparametric models*, J. Math. Psychol., 56 (2012), pp. 1–12.
[27] A. Ghosh, B. N. Subudhi, and L. Bruzzone, *Integration of Gibbs Markov random field and Hopfield-type neural networks for unsupervised change detection in remotely sensed multitemporal images*, IEEE Trans. Image Process., 22 (2013), pp. 3087–3096.
[28] S. Ghosh, L. Bruzzone, S. Patra, F. Bovolo, and A. Ghosh, *A context sensitive technique for unsupervised change detection based on Hopfield type neural networks*, IEEE Trans. Geosci. Remote Sens., 45 (2007), pp. 778–789.

[29] L. Giustarini, R. Hostache, P. Matgen, G. J.-P. Schumann, P. D. Bates, and D. C. Mason, *A change detection approach to flood mapping in urban areas using TerraSAR-X*, IEEE Trans. Geosci. Remote Sens., 51 (2013), pp. 2417–2430.
[30] G. R. Grimmett, *A theorem about random fields*, Bull. Lond. Math. Soc., 5 (1973), pp. 81–84.
[31] F. Harary, *Graph Theory*, Addison-Wesley Ser. Math., Addison-Wesley, Reading, MA, 1994.
[32] H. L. Harney, *Bayesian Inference: Parameter Estimation and Decisions*, Adv. Texts Phys., Springer, New York, 2003.
[33] J. Inglada and A. Giros, *On the possibility of automatic multisensor image registration*, IEEE Trans. Geosci. Remote Sens., 42 (2004), pp. 2104–2120.
[34] J. Inglada and G. Mercier, *A new statistical similarity measure for change detection in multitemporal SAR images and its extension to multiscale change analysis*, IEEE Trans. Geosci. Remote Sens., 45 (2009), pp. 1432–1445.
[35] H. Jeffreys, *An invariant form for the prior probability in estimation problems*, Proc. R. Soc. A, 186 (1946), pp. 453–461.
[36] J.-S. Lee, *Speckle suppression and analysis for synthetic aperture radar images*, Optim. Eng., 25 (1986), 255636.
[37] L. Liu, Y. Jiang, and C. Wang, *Noise analysis and image restoration for optical sparse aperture systems*, in Proceedings of ETTand6RS, Vol. 1, Shanghai, China, 2008, pp. 353–356.
[38] G. Mercier, G. Moser, and S. B. Serpico, *Conditional copulas for change detection in heterogeneous remote sensing images*, IEEE Trans. Geosci. Remote Sens., 46 (2008), pp. 1428–1441.
[39] R. M. Neal, *Markov chain sampling methods for Dirichlet process mixture models*, J. Comput. Graph. Statist., 9 (2000), pp. 249–265.
[40] C. Oliver and S. Quegan, *Understanding Synthetic Aperture Radar Images*, Artech House, Boston, 1998.
[41] P. Orbanz and J. M. Buhmann, *Nonparametric Bayesian image segmentation*, Int. J. Comput. Vis., 77 (2008), pp. 25–45.
[42] P. Orbanz and Y. W. Teh, *Bayesian nonparametric models*, in Encyclopedia of Machine Learning, Springer, New York, 2010, pp. 81–89.
[43] F. Pacifici and F. D. Frate, *Automatic change detection in very high resolution images with pulse-coupled neural networks*, IEEE Geosci. Remote Sens. Lett., 7 (2010), pp. 58–62.
[44] F. Pacifici, F. D. Frate, C. Solimini, and W. J. Emery, *An innovative neural-net method to detect temporal changes in high-resolution optical satellite imagery*, IEEE Trans. Geosci. Remote Sens., 45 (2007), pp. 2940–2952.
[45] O. Papaspiliopoulos and G. Roberts, *Retrospective Markov chain Monte Carlo methods for Dirichlet process hierarchical models*, Biometrika, 95 (2008), pp. 169–186.
[46] W. W. Peterson, T. G. Birdsall, and W. Fox, *The theory of signal detectability*, IRE Trans. Inf. Theory, 4 (1954), pp. 171–212.
[47] R. B. Potts, *Some generalized order-disorder transformations*, Math. Proc. Cambridge Philos. Soc., 48 (1952), pp. 106–109.
[48] V. Poulain, J. Inglada, M. Spigai, J.-Y. Tourneret, and P. Marthon, *High resolution optical and SAR image fusion for building database updating*, IEEE Trans. Geosci. Remote Sens., 49 (2011), pp. 2900–2910.
[49] C. Pratola, F. D. Frate, G. Schiavon, and D. Solimini, *Toward fully automatic detection of changes in suburban areas from VHR SAR images by combining multiple neural-network models*, IEEE Trans. Geosci. Remote Sens., 51 (2013), pp. 2055–2066.
[50] J. Prendes, M. Chabert, F. Pascal, A. Giros, and J.-Y. Tourneret, *A multivariate statistical model for multiple images acquired by homogeneous or heterogeneous sensors*, in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Florence, Italy, 2014.
[51] J. Prendes, M. Chabert, F. Pascal, A. Giros, and J.-Y. Tourneret, *A new multivariate statistical model for change detection in images acquired by homogeneous and heterogeneous sensors*, IEEE Trans. Image Process., 24 (2015), pp. 799–812.
[52] G. Quin, B. Pinel-Puyssegur, J.-M. Nicolas, and P. Loreaux, *MIMOSA: An automatic change detection method for SAR time series*, IEEE Trans. Geosci. Remote Sens., 52 (2014), pp. 5349–5363.

[53] E. J. M. Rignot and J. Van Zyl, *Change detection techniques for ERS-1 SAR data*, IEEE Trans. Geosci. Remote Sens., 31 (1993), pp. 896–906.
[54] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, Texts in Statistics, Springer, New York, 2005.
[55] R. A. Schowengerdt, *Remote Sensing: Models and Methods for Image Processing*, Elsevier Science, New York, 2006.
[56] A. Singh, *Digital change detection techniques using remotely-sensed data*, Int. J. Remote Sens., 10 (1989), pp. 989–1003.
[57] C. D. Storie, J. Storie, and G. Salinas de Salmuni, *Urban boundary extraction using 2-component polarimetric SAR decomposition*, in Proceedings of IGARSS, Munich, 2012, pp. 5741–5744.
[58] Y. W. Teh and M. I. Jordan, *Hierarchical Bayesian Nonparametric Models with Applications*, Cambridge University Press, Cambridge, UK, 2010.
[59] C. Tison, J.-M. Nicolas, F. Tupin, and H. Maitre, *A new statistical model for Markovian classification of urban areas in high-resolution SAR images*, IEEE Trans. Geosci. Remote Sens., 42 (2004), pp. 2046–2057.
[60] R. Touzi, A. Lopés, and P. Bousquet, *A statistical and geometrical edge detector for SAR images*, IEEE Trans. Geosci. Remote Sens., 26 (1988), pp. 764–773.
[61] P. Uprety and F. Yamazaki, *Use of high-resolution SAR intensity images for damage detection from the* 2010 *Haiti earthquake*, in Proceedings of IGARSS, Munich, 2012, pp. 6829–6832.
[62] D. A. Van Dyk and T. Park, *Partially collapsed Gibbs samplers: Theory and methods*, J. Amer. Statist. Soc., 103 (2008), pp. 790–796.
[63] J. D. Villasenor, D. R. Fatland, and L. D. Hinzman, *Change detection on Alaska's north slope using repeat-pass ERS-1 SAR imagery*, IEEE Trans. Geosci. Remote Sens., 31 (1993), pp. 227–235.