



# Improving land cover classification through contextual-based optimum-path forest



D. Osaku<sup>a</sup>, R.Y.M. Nakamura<sup>b</sup>, L.A.M. Pereira<sup>c</sup>, R.J. Pisani<sup>d</sup>, A.L.M. Levada<sup>a</sup>,  
F.A.M. Cappabianco<sup>e</sup>, A.X. Falcão<sup>c</sup>, João P. Papa<sup>f,\*</sup>

<sup>a</sup> Department of Computing, UFSCar—Univ Federal of São Carlos, São Carlos, Brazil

<sup>b</sup> Big Data Brasil, São Paulo, Brazil

<sup>c</sup> Institute of Computing, University of Campinas, Campinas, Brazil

<sup>d</sup> Western University of São Paulo, Presidente Prudente, SP, Brazil

<sup>e</sup> Federal University of São Paulo, São José dos Campos, Brazil

<sup>f</sup> Department of Computing, São Paulo State University, Bauru, Brazil

## ARTICLE INFO

### Article history:

Received 14 January 2015

Revised 14 May 2015

Accepted 12 June 2015

Available online 2 July 2015

### Keywords:

Land cover classification

Optimum-path forest

Contextual classification

## ABSTRACT

Traditional machine learning algorithms very often assume statistically independent data samples. However, this is clearly not the case in remote sensing image applications, in which pixels present spatial and/or temporal dependencies. In this work, it has been presented an approach to improve land cover image classification using a contextual approach based on optimum-path forest (OPF) and the well-known Markov random fields (MRFs), hereinafter called OPF–MRF. In addition, it is also introduced a framework to the optimization of the amount of contextual information used by OPF–MRF. Experiments over high- and medium-resolution satellite (CBERS-2B, Landsat 5 TM, Ikonos-2 MS and Geoeye) and radar (ALOS-PALSAR) images covering the area of two Brazilian cities have shown the proposed approach can overcome several shortcomings related to standard OPF classification. In some cases, the proposed approach outperformed traditional OPF in about 9% of recognition rate, which is crucial for land cover classification.

© 2015 Elsevier Inc. All rights reserved.

## 1. Introduction

Pixel-based classification plays an important role in many remote sensing applications, such as land cover and target recognition, just to name a few. Remote sensing image classification is often performed by support vector machines—SVMs [3,19,21], artificial neural networks—ANNs [17,19], and optimum-path forest—OPF [28,35], but without taking into account the spatial and/or temporal dependencies among pixels. However, these approaches can generate undesirable artefacts (e.g., salt-and-pepper effect) as reported in some works [8,38]. Contextual classifiers have been proposed to address the problem by exploiting the spatial dependency of nearby pixels that are likely to belong to a same class [5].

Tarabalka et al. [40], for instance, proposed a contextual approach based on support vector machines and Markov random fields (MRFs) for remote sensing image classification called SVM–MRF. Moser and Serpico [27] also proposed a version of the

\* Corresponding author. Tel.: +55 14 3103 6079; fax: +55 14 3103 6079.

E-mail addresses: [danosaku@hotmail.com](mailto:danosaku@hotmail.com), [alexandre@dc.ufscar.br](mailto:alexandre@dc.ufscar.br) (D. Osaku), [rodrigo.mizobe@bigdata.inf.br](mailto:rodrigo.mizobe@bigdata.inf.br) (R.Y.M. Nakamura), [luis.pereira@ic.unicamp.br](mailto:luis.pereira@ic.unicamp.br) (L.A.M. Pereira), [pisanigeo@gmail.com](mailto:pisanigeo@gmail.com) (R.J. Pisani), [fcappabianco@gmail.com](mailto:fcappabianco@gmail.com) (F.A.M. Cappabianco), [afalcao@ic.unicamp.br](mailto:afalcao@ic.unicamp.br) (A.X. Falcão), [papa@fc.unesp.br](mailto:papa@fc.unesp.br), [papa.joaopaulo@gmail.com](mailto:papa.joaopaulo@gmail.com) (J.P. Papa).

SVM–MRF approach using a single-step formulation, rather than the two-phase strategy presented in [40]. Kittler and Föglein [20] combined a Bayesian classifier and a Gaussian MRF by using pixel spectral and spatial dependencies. Shekhar et al. [38] employed a Bayesian classifier with different models to capture spatial context. They provided comparisons among spatial autoregression and two different Markov models. Another related work was proposed by Zhang et al. [44], which used SVMs with an adaptive MRF-based approach (a-MRF) to avoid overcorrection on the boundary between classes. Other approaches for contextual classification of remote sensing images exploit hierarchical and multi-resolution feature description [7,9,41]. Additionally, Frery et al. [12] performed multispectral image classification by means of Gaussian maximum likelihood and contextual iterated conditional modes (ICM) algorithm.

In the context of land cover classification using contextual information, the reader can find some interesting works as well. Laha et al. [23], for instance, applied evidence theory to incorporate contextual information for further image classification using fuzzy rules. Ghimire et al. [15] applied random forests, and Stuckens et al. [39] employed a linkage-based clustering algorithm for land cover classification using contextual information. Sarkar et al. [37] presented a hybrid approach among MRF and Dempster–Shafer theory. It is also valuable to shed light over the seminal work of Wharton [42], which presented a contextual land cover classification approach based on the local distribution of nearby labels, as well as the work of Guo et al. [16], which proposed a contextual approach based on cascaded classifiers for remote sensing imagery classification. Recently, a comprehensive review about contextual-based classifiers has been presented by Li et al. [24]. Finally, Mahmoudi et al. [26] stated the contextual relations can overcome some challenges in urban areas recognition based on satellite imagery. Although Aghighi et al. [1,2] have also presented an approach to estimate a smoothing parameter that controls the amount of interactions between the spatial and contextual information based on dynamic blocks, SVM and class label co-occurrence matrices, their works differ from this one, since here we employed meta-heuristic techniques to address the problem of contextual-based classification concerning the OPF classifier.

In short, we propose a meta-heuristic-based optimization framework to find suitable values for the parameter that controls the amount of contextual information used in the classification process. Additionally, we introduce a contextual classifier named OPF–MRF based on MRF and optimum-path forest for supervised learning [32,33]. The traditional OPF classifier interprets training samples (feature vectors) as nodes of a complete graph, identifies key samples (prototypes) in all classes, and partitions the graph into an optimum-path forest rooted at those prototypes. The classification of a new sample is quickly performed in an incremental way by finding its most closely connected training sample and assigning its class to it. Such a strategy has demonstrated to be similar/superior to SVMs and ANNs in several applications, being much faster for training, since this version of OPF is parameter-independent.

In the OPF–MRF approach, the output of OPF classifier is interpreted as an MRF and improved by ICM algorithm [6] in order to add contextual (sample labels) to the original feature vectors. A second OPF classifier, trained with the extended feature vectors, outputs the final label map. As two independent and preliminary works, the OPF–MRF classifier was previously presented for magnetic resonance image classification [29], and the proposed meta-heuristic-based framework was previously evaluated in the same context as well [31]. In this paper, it is the first time we combine both approaches for contextual classification and evaluate them for land cover classification. The work have shown one can improve vanilla OPF classification in the context of land cover recognition using images obtained by four satellites with different spatial resolutions, as well as one radar image. This is important to highlight the proposed approach can obtain better results than pixel-based OPF classification in distinct scenarios.

The remainder of this paper is organized as follows. Sections 2 and 3 describe the OPF theory background, as well as its contextual version, respectively. The experimental setup and experiments are presented in Sections 4 and 5, respectively. Section 6 states conclusions and future works.

## 2. Optimum-path forest classification

The OPF framework is a recent highlight to the development of pattern recognition techniques based on graph partitions. The nodes are the data samples, which are represented by their corresponding feature vectors, and are connected according to some predefined adjacency relation. Given some key nodes (prototypes), they will compete among themselves aiming at conquering the remaining nodes. Thus, the algorithm outputs an optimum path forest, which is a collection of optimum-path trees (OPTs) rooted at each prototype. This work employs the OPF classifier proposed by Papa et al. [32,33], which is explained in more details as follows.

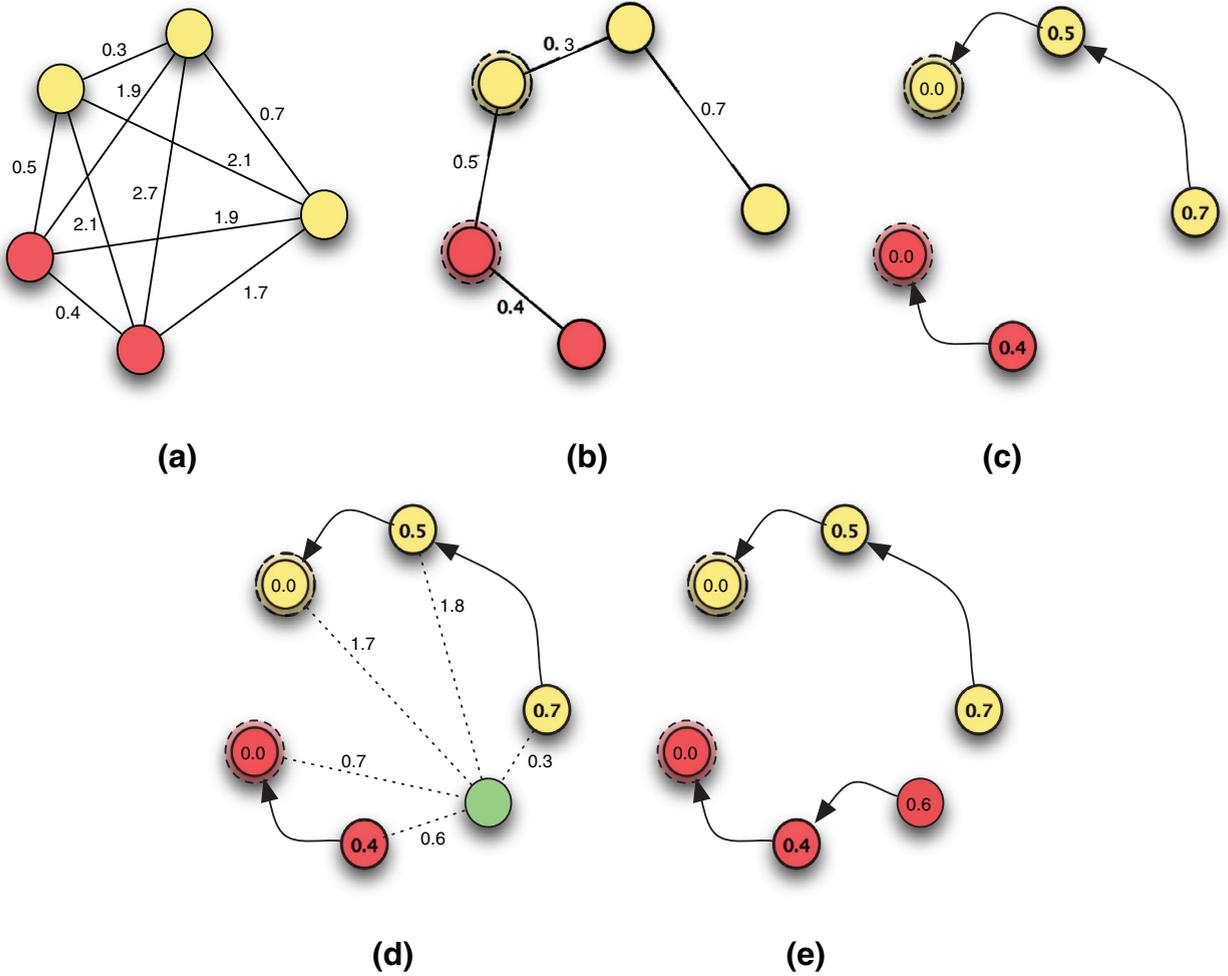
Let  $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2$  be a labeled dataset, such that  $\mathcal{D}_1$  and  $\mathcal{D}_2$  stands for the training and test sets, respectively. Let  $S \subset \mathcal{D}_1$  be a set of prototypes of all classes (i.e., key samples that best represent the classes). Let  $(\mathcal{D}_1, A)$  be a complete graph whose nodes are the samples in  $\mathcal{D}_1$ , and any pair of samples defines an arc in  $A = \mathcal{D}_1 \times \mathcal{D}_1$  (Fig. 1a).<sup>1</sup> Additionally, let  $\pi_s$  be a path in  $(\mathcal{D}_1, A)$  with terminus at sample  $s \in \mathcal{D}_1$ .

The OPF algorithm proposed by Papa et al. [32,33] employs the path-cost function  $f_{\max}$  due to its theoretical properties for estimating prototypes (Section 2.1 gives further details about this procedure):

$$f_{\max}(\langle s \rangle) = \begin{cases} 0 & \text{if } s \in S \\ +\infty & \text{otherwise,} \end{cases}$$

$$f_{\max}(\pi_s \cdot \langle s, t \rangle) = \max\{f_{\max}(\pi_s), d(s, t)\}, \quad (1)$$

<sup>1</sup> The arcs are weighted by the distance between their corresponding nodes.



**Fig. 1.** (a) Training set modeled as a complete graph, (b) a minimum spanning tree computation over the training set (prototypes are highlighted), (c) optimum-path forest over the training set, (d) classification process of a “green” sample, and (e) test sample is finally classified. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

where  $d(s, t)$  stands for a distance between nodes  $s$  and  $t$ , such that  $s, t \in \mathcal{D}_1$ . Therefore,  $f_{\max}(\pi_s)$  computes the maximum distance between adjacent samples in  $\pi_s$ , when  $\pi_s$  is not a trivial path. In short, the OPF algorithm tries to minimize  $f_{\max}(\pi_t)$ ,  $\forall t \in \mathcal{D}_1$ .

### 2.1. Training

We say that  $S^*$  is an optimum set of prototypes when *Algorithm 1* minimizes the classification errors for every  $s \in \mathcal{D}_1$ . We have that  $S^*$  can be found by exploiting the theoretical relation between the minimum-spanning tree and the optimum-path tree for  $f_{\max}$  [4]. The training essentially consists of finding  $S^*$  and an OPF classifier rooted at  $S^*$ . By computing a minimum spanning tree (MST) in the complete graph  $(\mathcal{D}_1, A)$  (Fig. 1b), one obtain a connected acyclic graph whose nodes are all samples of  $\mathcal{D}_1$  and the arcs are undirected and weighted by the distances  $d$  between adjacent samples. In the MST, every pair of samples is connected by a single path, which is optimum according to  $f_{\max}$ . Hence, the minimum-spanning tree contains one optimum-path tree for any selected root node.

The optimum prototypes are the closest elements of the MST with different labels in  $\mathcal{D}_1$  (i.e., elements that fall in the frontier of the classes, as highlighted in Fig. 1b). By removing the arcs between different classes, their adjacent samples become prototypes in  $S^*$ , and *Algorithm 1* can define an optimum-path forest with minimum classification errors in  $\mathcal{D}_1$  (Fig. 1c). *Algorithm 1* implements the above procedure for OPF training phase.

Lines 1–4 initialize maps and insert prototypes in  $Q$  (the function  $\lambda(\cdot)$  in Line 4 assigns the true label to each training sample).<sup>2</sup> The main loop computes an optimum path from  $S$  to every sample  $s$  in a non-decreasing order of cost (Lines 5–13). At each

<sup>2</sup> The cost map  $C$  stores the optimum-cost of each training sample.

**Algorithm 1** OPF training algorithm.

INPUT: A labeled training set  $\mathcal{D}_1$ .  
 OUTPUT: Optimum-path forest  $P$ , cost map  $C$ , label map  $L$ , and ordered set  $D'_1$ .  
 AUXILIARY: Priority queue  $Q$ , set  $S$  of prototypes, and cost variable  $cst$ .

1. Set  $D'_1 \leftarrow \emptyset$  and compute the prototype set by MST  $\mathcal{S} \subset \mathcal{D}_1$ .
2. For each  $s \in \mathcal{D}_1 \setminus \mathcal{S}$ , set  $C(s) \leftarrow +\infty$ .
3. For each  $s \in \mathcal{S}$ , do
  4.  $C(s) \leftarrow 0$ ,  $P(s) \leftarrow nil$ ,  $L(s) \leftarrow \lambda(s)$ , and insert  $s$  in  $Q$ .
5. While  $Q$  is not empty, do
  6. Remove from  $Q$  a sample  $s$  such that  $C(s)$  is minimum.
  7. Insert  $s$  in  $D'_1$ .
  8. For each  $t \in \mathcal{D}_1$  such that  $C(t) > C(s)$ , do
    9. Compute  $cst \leftarrow \max\{C(s), d(s, t)\}$ .
    10. If  $cst < C(t)$ , then
      11. If  $C(t) \neq +\infty$ , then remove  $t$  from  $Q$ .
      12.  $P(t) \leftarrow s$ ,  $L(t) \leftarrow L(s)$ ,  $C(t) \leftarrow cst$ .
      13. Insert  $t$  in  $Q$ .
14. Return a classifier  $[P, C, L, D'_1]$ .

iteration, a path of minimum cost  $C(s)$  is obtained in  $P$  when we remove its last node  $s$  from  $Q$  (Line 6). Ties are broken in  $Q$  using first-in-first-out policy. That is, when two optimum paths reach an ambiguous sample  $s$  with the same minimum cost,  $s$  is assigned to the first path that reached it. Note that  $C(t) > C(s)$  in Line 8 is false when  $t$  has been removed from  $Q$  and, therefore,  $C(t) \neq +\infty$  in Line 11 is true only when  $t \in Q$ . Lines 10–13 evaluate whether the path that reaches an adjacent node  $t$  through  $s$  is cheaper than the current path with terminus  $t$ , and update the position of  $t$  in  $Q$ ,  $C(t)$ ,  $L(t)$  and  $P(t)$  accordingly. The complexity for training OPF is done by  $\theta(|\mathcal{D}_1|^2)$ , due to the main (Lines 5–13) and inner loops (Lines 8–13) in Algorithm 1, which run  $\theta(|\mathcal{D}_1|)$  times each.

## 2.2. Classification

For any sample  $t \in \mathcal{D}_2$ , we consider all arcs connecting  $t$  with samples  $s \in \mathcal{D}_1$ , as though  $t$  were part of the training graph (Fig. 1d). Considering all possible paths from  $S^*$  to  $t$ , we find the optimum path  $P^*(t)$  from  $S^*$  and label  $t$  with the class  $\lambda(R(t))$  of its most strongly connected prototype  $R(t) \in S^*$  (Fig. 1e). This path can be identified incrementally, by evaluating the optimum cost  $C(t)$  as follows:

$$C(t) = \min\{\max\{C(s), d(s, t)\}, \forall s \in \mathcal{D}_1. \quad (2)$$

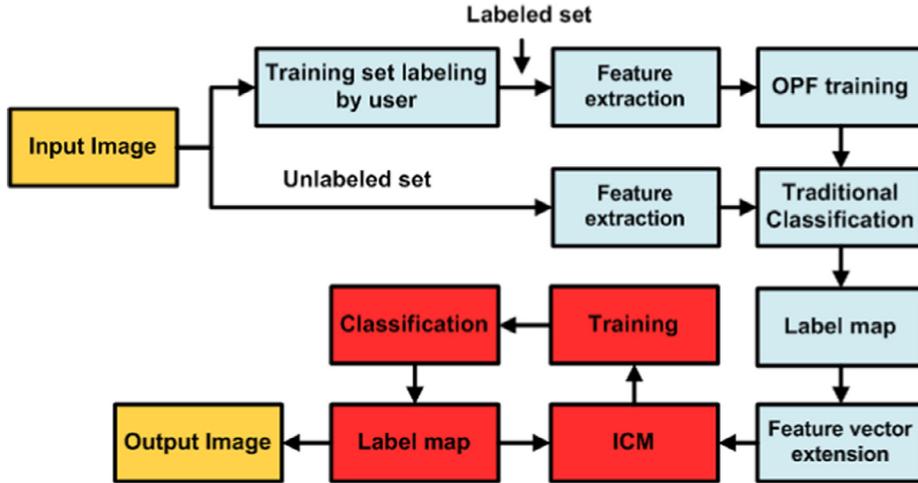
Let the node  $s^* \in \mathcal{D}_1$  be the one that satisfies Eq. (2) (i.e., the predecessor  $P(t)$  in the optimum path  $P^*(t)$ ). Given that  $L(s^*) = \lambda(R(t))$ , the classification simply assigns  $L(s^*)$  as the class of  $t$ . An error occurs when  $L(s^*) \neq \lambda(t)$ . Algorithm 2 implements this idea.

In Algorithm 2, the main loop (Lines 1–9) performs the classification of all nodes in  $\mathcal{D}_2$ . The inner loop (Lines 4–9) visits each node  $k_{i+1} \in D'_1$ ,  $i = 1, 2, \dots, |Z'_1| - 1$  until an optimum path  $\pi_{k_{i+1}} \cdot \langle k_{i+1}, t \rangle$  is found. In the worst case, the algorithm visits all

**Algorithm 2** OPF classification algorithm.

INPUT: Classifier  $[P, C, L, D'_1]$  and test set  $\mathcal{D}_2$ .  
 OUTPUT: Label  $L_o$  and predecessor  $P_o$  maps defined for  $\mathcal{D}_2$ .  
 AUXILIARY: Cost variables  $tmp$  and  $mincost$ .

1. For each  $t \in \mathcal{D}_2$ , do
  2.  $i \leftarrow 1$ ,  $mincost \leftarrow \max\{C(k_i), d(k_i, t)\}$ .
  3.  $L_o(t) \leftarrow L(k_i)$  and  $P_o(t) \leftarrow k_i$ .
  4. While  $i < |D'_1|$  and  $mincost > C(k_{i+1})$ , do
    5. Compute  $tmp \leftarrow \max\{C(k_{i+1}), d(k_{i+1}, t)\}$ .
    6. If  $tmp < mincost$ , then
      7.  $mincost \leftarrow tmp$ .
      8.  $L_o(t) \leftarrow L(k_{i+1})$  and  $P_o(t) \leftarrow k_{i+1}$ .
    9.  $i \leftarrow i + 1$ .
10. Return  $[L_o, P_o]$ .



**Fig. 2.** OPF-MRF workflow with  $\beta$  provided by the user. (For interpretation of the references to colour in the text, the reader is referred to the web version of this article.)

nodes in  $D'_1$ . Line 5 evaluates  $f_{\max}(\pi_{k_{i+1}} \cdot \langle k_{i+1}, t \rangle)$  and Lines 7–8 update the cost, label and predecessor of  $t$  whenever  $\pi_{k_{i+1}} \cdot \langle k_{i+1}, t \rangle$  is better than the current path  $\pi_t$  (Line 6). Although the reader can note the complexity of the OPF classification phase is given by  $\theta(|\mathcal{D}_1||\mathcal{D}_2|)$  (for each classification node we have to visit all training samples), Papa et al. [32] showed that, in practice, the complexity is given by  $O(p|\mathcal{D}_2|)$ , in which  $p \in O(|\mathcal{D}_1|)$ .

Another interesting point to be considered concerns with the relation between OPF and the nearest neighbor classifier (NN). Although OPF uses the distance between samples to compose the cost to be offered to them, the path-cost function encodes the power of connectivity of the samples that fall in the same path, being much more powerful than the sole distance. Therefore, this means OPF is not a distance-based classifier. Additionally, Papa et al. [33] showed that OPF is quite different than NN, being those techniques exactly the same only when all training samples become prototypes.

### 3. OPF-MRF: contextual classification using Markov random fields and optimum-path forest

Markov random fields are probabilistic models often be used to integrate both spatial and contextual information in image classification problems. Such models are based on the idea that a pixel has a high probability to belong to the same class of its neighbors.

The reader can face several Markov models that consider contextual information in the literature, being such models very appropriate to represent a prior knowledge by means of probability distributions. One of the most widely used is the well-known Potts model–Potts–Strauss model [14], which arose from the statistical physics to generalize the Ising model for multiple discrete energies [36,43]. In the context of image processing and pattern recognition, such model has been widely employed as a priori knowledge to hold the smoothness assumption, since the pixels that fall in the same neighborhood are likely to have the same label.

Given a pixel located at  $(i, j)$  position and its neighborhood system  $\eta_{ij}$ , it is possible to define the local conditional probability of Potts model as follows:

$$p(x_{ij} = m_{ij} | \eta_{ij}, \beta) = \frac{\exp\{\beta \mathcal{U}_{ij}(m_{ij})\}}{\sum_{k=1}^K \exp\{\beta \mathcal{U}_{ij}(k)\}}, \quad (3)$$

in which  $x_{ij} \in \mathcal{X}$  is a realization of the Markov label field  $\mathcal{X}$ ,  $\mathcal{U}_{ij}(k)$  stands for the number of pixels in  $\eta_{ij}$  that are from label  $k$ ,  $K$  denotes the number of labels,  $\beta$  is a parameter that models the spatial dependence among neighboring pixels, and  $m_{ij} \in \{1, 2, \dots, K\}$  means the observed label at central pixel  $(i, j)$ . It is important to shed light over that high values of  $\beta$  lead to a high spatial dependence among pixels, and  $\beta = 0$  means we have no spatial dependence, i.e., we have a traditional classification.

The idea behind OPF-MRF is to exploit the contextual information by means of a lattice-based neighborhood model, being inspired by the work of Tarabalka et al. [40]. The proposed contextual-based OPF learning algorithm can be divided in three phases: (i) a pixel-wise classification is performed by traditional OPF to generate the initial label map (classified image), (ii) further, the local probability given by Eq. (3) is maximized over the previous label map using the well-known iterated conditional modes algorithm [6], and thus generating a new label map, which is finally (iii) classified using standard OPF once again to generate an updated label map. The process is iterated from step (ii) to (iii) until a convergence criterion is met. Basically, the idea of OPF-MRF is to iteratively maximize a local probability density function followed by the minimization of the path-cost function  $f_{\max}(\cdot)$  for all dataset samples. Fig. 2 depicts the OPF-MRF workflow.

The reader can realize the first step of OPF-MRF follows a traditional classification pipeline, i.e., we have training and test sets, being the latter used to generate the label map (classified image) that will be an input for the contextual classification.

**Algorithm 3** OPF–MRF algorithm.

INPUT: A labeled training  $\mathcal{D}_1$  and test set  $\mathcal{D}_2$ , number of OPF–MRF iterations  $T$ , and the spatial dependency parameter  $\beta$ .

OUTPUT: A labeled map  $\mathcal{L}$

1. Define a Markov random field model  $\mathcal{M}$  (e.g. Potts model).
2.  $[P, C, L, \mathcal{D}'_1] \rightarrow$  Algorithm 1( $\mathcal{D}_1$ ).
3.  $[\mathcal{L}, P_0] \rightarrow$  Algorithm 2( $\mathcal{D}_1, \mathcal{D}_2$ ).
4. Create new instances  $\mathcal{D}'_1$  and  $\mathcal{D}'_2$  by extending  $\mathcal{D}_1$  and  $\mathcal{D}_2$
5. with the frequency of the neighborhood labels given by  $\mathcal{L}$ .
6.  $i \leftarrow 1, \Delta \leftarrow +\infty$ .
7. While  $i \leq T$  and  $\Delta > 0.01\%$ , do
  8.  $[\mathcal{D}^*_1, \mathcal{D}^*_2] \leftarrow$  ICM( $\mathcal{M}, \mathcal{D}'_1, \mathcal{D}'_2, \beta, \mathcal{L}$ ).
  9.  $[P, C, L, \mathcal{D}'_1] \rightarrow$  Algorithm 1( $\mathcal{D}^*_1$ ).
  10.  $[\mathcal{L}', P_0] \rightarrow$  Algorithm 2( $\mathcal{D}^*_1, \mathcal{D}^*_2$ ).
  11. Compute  $\Delta$  using  $\mathcal{L}$  and  $\mathcal{L}'$ .
  12.  $\mathcal{L} \leftarrow \mathcal{L}', \mathcal{D}'_1 \leftarrow \mathcal{D}^*_1, \mathcal{D}'_2 \leftarrow \mathcal{D}^*_2$  and  $i \leftarrow i + 1$ .
13. Return the label map  $\mathcal{L}$ .

**Algorithm 3** implements the aforementioned approach. Line 1 defines a Markov random field model based on the Potts model (Eq. (3)). Basically, such model considers the frequency of each label in a given pixel  $(i, j)$  neighborhood, defined by  $\eta_{ij}$ . Further, a standard OPF training step is performed over  $\mathcal{D}_1$  (Section 2.1) (Line 2), followed by the classification of the test set  $\mathcal{D}_2$  (Section 2.2) in order to generate an initial label map  $\mathcal{L}$  (Line 3).

Lines 4–5 are responsible for extending the feature vectors of samples from  $\mathcal{D}_1$  and  $\mathcal{D}_2$ . The *feature vector extension* is performed as follows: for each  $s \in \mathcal{D}_1 \cup \mathcal{D}_2$ , we consider the brightness of its eight-neighborhood (RGB channels), as well as its  $(x, y)$  position in the image. Therefore, the extended feature vector sample is composed by  $26 + K$  features, being  $24(3 \times 8)$  from the brightness of the *eight-neighborhood* pixels for each channel, plus two features that come from the  $(x, y)$  sample's position, and the probability of that sample to belong to class  $i$  given by Potts model (Eq. (3)),  $i = 1, 2, \dots, K$ . Note the original feature vector of each sample  $s$  (Lines 2–3) is composed by three features, that stand for the pixel's brightness for each RGB channel.

The loop in Lines 7–12 (“red” modules in Fig. 2) is responsible for refining the initial solution given by traditional OPF. This step maximizes the local probability function given by Potts model (Eq. (3)) using the ICM algorithm in Line 8. The reader can observe that ICM algorithm has five inputs: the Markov random field model  $M$  (Potts model in this paper), extended training ( $\mathcal{D}'_1$ ) and test sets ( $\mathcal{D}'_2$ ), spatial dependency parameter  $\beta$ , and the label map  $\mathcal{L}$  generated by standard OPF. After that, the outputs of ICM are the new training ( $\mathcal{D}^*_1$ ) and test sets ( $\mathcal{D}^*_2$ ), in which their samples' labels have been updated by ICM. In Lines 9–10, OPF is trained over  $\mathcal{D}^*_1$  for further classification of  $\mathcal{D}^*_2$ , which generates the updated label map  $\mathcal{L}'$ . Finally, Line 12 updates the variables for the next OPF–MRF iteration. The reader can observe that Line 7 has two stopping criteria: the maximum number of iterations  $T$ , and the number of testing samples whose labels are different from the previous iteration and the current one. That is, the variable  $\Delta$  can be seen as the amount of changing in the label map between iterations. In this work, we have used  $T = 10$  and  $\Delta = 0.01\%$  (empirically set).

Additionally, one can observe OPF–MRF has two parameters: number of iterations  $T$  and the spatial dependency parameter  $\beta$ . Another main contribution of this paper is to introduce a meta-heuristic-based optimization framework for finding  $\beta$  automatically, being a suitable additional tool to provide an user-friendly version of OPF–MRF.

### 3.1. Spatial-content dependence estimation through meta-heuristics learning

As aforementioned, the parameter  $\beta$  is responsible for constraining the amount of spatial dependency used by OPF–MRF (if  $\beta = 0$  in Eq. (3), we have no contextual information considered in the classification process). Although this parameter can be provided by the user, we proposed an evolutionary-based framework for the optimization of  $\beta$ , which makes use of a validation set to find out suitable  $\beta$  values that maximize the OPF recognition rate at that set. Since  $\beta \in [0, \beta_{\max}]$ , in which  $\beta_{\max} = \ln(1 + \sqrt{K})$  [43] (*critical value*), there are infinity values for  $\beta \in \mathbb{R}$  on that range, which turns the problem of finding reasonable values for  $\beta$  using a near-exhaustive search impractical for large (high-resolution) images. It is also interesting to highlight the  $\beta_{\max}$  value is dependent on the number of classes  $K$ . Fig. 3 illustrates the OPF–MRF pipeline considering the  $\beta$  optimization.

Now, assume the dataset  $\mathcal{D}$  can be partitioned in  $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_v \cup \mathcal{D}_2$ , which stand for the training, validating and test sets, respectively. The idea is to use both  $\mathcal{D}_1$  and  $\mathcal{D}_v$  to guide the meta-heuristic optimization algorithms to maximize the recognition rate over  $\mathcal{D}_v$ . Each evolutionary agent has a position in the search space defined by a  $\beta \in [0, \beta_{\max}]$  randomly initialized. After that, **Algorithm 3** is executed every time an agent changes its position until the convergence criterion is met. Such procedure is encoded by the “green” module from Fig. 3. Further, after finding a suitable  $\beta$  value, i.e., the one which maximizes the recognition rate over  $\mathcal{D}_v$ , the usual refinement step is performed by the “red” modules (Fig. 3).

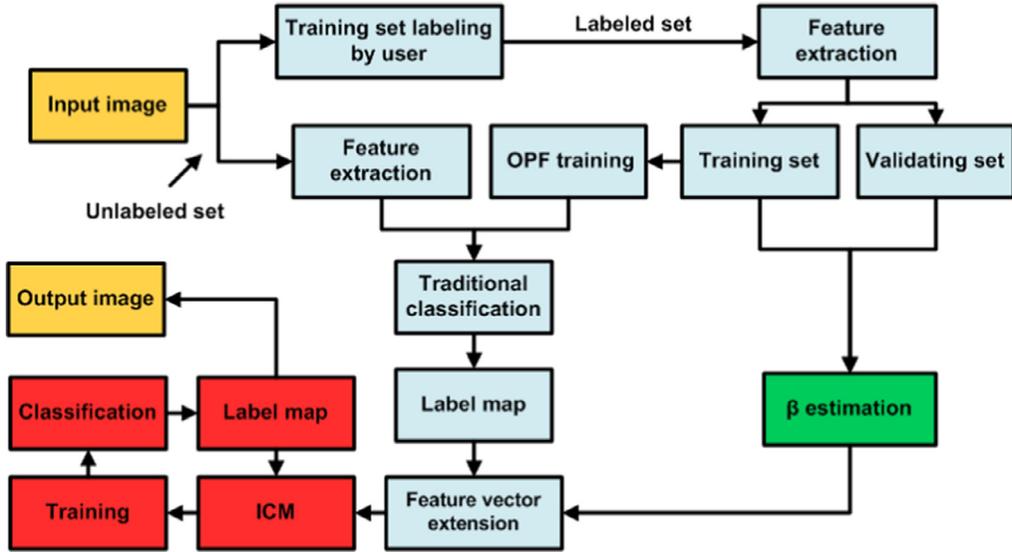


Fig. 3. OPF-MRF workflow with  $\beta$  estimated by meta-heuristic optimization. (For interpretation of the references to colour in the text, the reader is referred to the web version of this article.)

---

**Algorithm 4** Meta-heuristic-based  $\beta$  optimization algorithm.

---

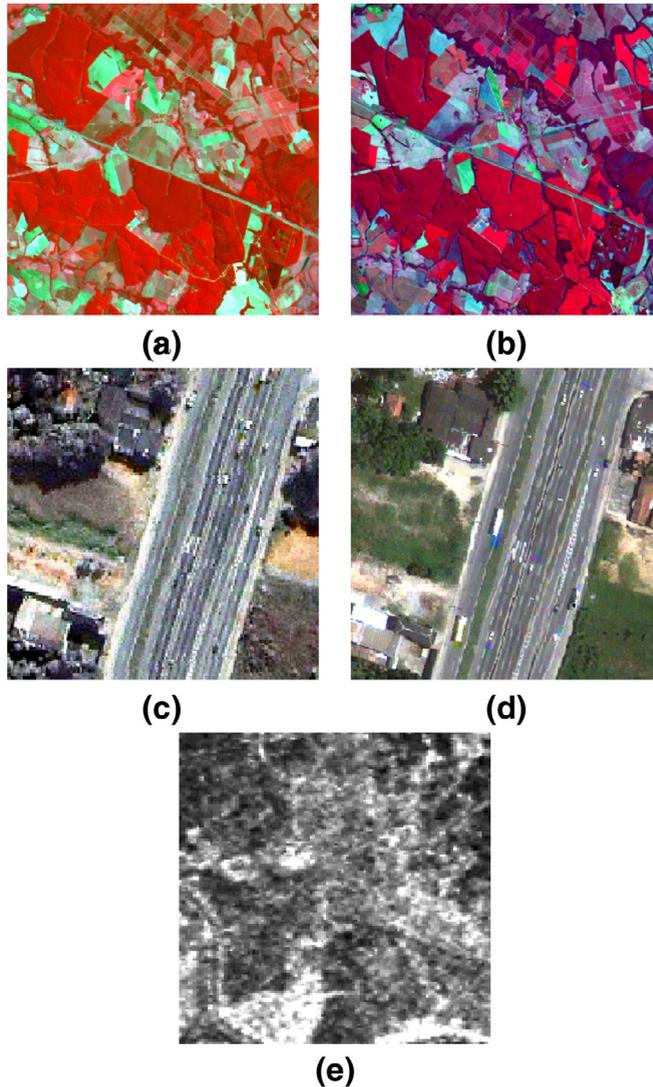
INPUT: A labeled training  $\mathcal{D}_1$ , validating  $\mathcal{D}_v$  and test set  $\mathcal{D}_2$ , the number of the optimization algorithm iterations  $T_{opt}$ , and the number of OPF-MRF iterations  $T$ .

OUTPUT: The  $\beta$  parameter that maximizes the accuracy over  $\mathcal{D}_v$ .

1. For each agent  $i$  from 1 to  $W$ , do.
  2.      $pos_i \leftarrow [0, \beta_{max}]$ .
  3.      $fit_i \leftarrow -\infty$ .
  4.  $global \leftarrow -\infty$ .
  5. For each iteration  $t$  from 1 to  $T_{opt}$ , do.
  6.     For each agent  $i$  from 1 to  $W$ , do.
  7.          $\mathcal{L} \leftarrow \text{Algorithm 3}(\mathcal{D}_1, \mathcal{D}_v, T, pos_i)$ .
  8.          $tmp \leftarrow \text{accuracy over } \mathcal{D}_v \text{ using } \mathcal{L}$ .
  9.         If  $tmp > fit_i$ , then.
  10.              $fit_i \leftarrow tmp$ .
  11.             Update the position of agent  $i$ .
  12.      $global \leftarrow$  the best fitness value among all agents.
  13.      $pos^* \leftarrow$  the position of the best agent.
  14.  $\beta \leftarrow pos^*$ .
  15. Return  $\beta$ .
- 

Although any other optimization technique can be employed to find  $\beta$ , in this paper we proposed a meta-heuristic approach for such purpose, since such sort of techniques present an elegant and user-friendly framework for optimization-oriented applications. We have used two meta-heuristic techniques: particle swarm optimization (PSO) [18] and harmony search (HS) [13]. While the former approach is based on the social dynamics of a group of individuals and the interaction among them aiming at finding a source of food, HS models the optimization problem as being a task of finding a music with the best harmony as possible. In this case, each decision variable is modeled as a musical instrument, which can come from the harmony memory (artist memory) or even from the improvisation process of the artist. Therefore, while PSO has been widely used in several applications, HS has been one of the fastest meta-heuristic optimization techniques, which make them interesting to the context of this paper. Algorithm 4 implements the basic steps of the proposed framework (“green” module in Fig. 3), which can be adapted to be used with any other meta-heuristic optimization approach, as well as any other pattern recognition technique.

Lines 1–4 are responsible for the initialization of each agent (particles for PSO and harmonies for HS). While Line 2 initializes the position of each agent with a random value within the range  $[0, \beta_{max}]$ , its fitness value is set in Line 3. In addition, the global fitness, i.e., the best fitness in swarm, is set up in Line 4. In this paper, we used  $W = 5$ , i.e., five particles/harmonies ( $W$  has been empirically chosen). In regard to the number of iterations  $T_{opt}$ , we set  $T_{opt} = 10$  for PSO and  $T_{opt} = 50$  for HS. Once again, such values have been empirically chosen. Higher values for  $T_{opt}$  tend to be inefficient, since the improvements in the classified images



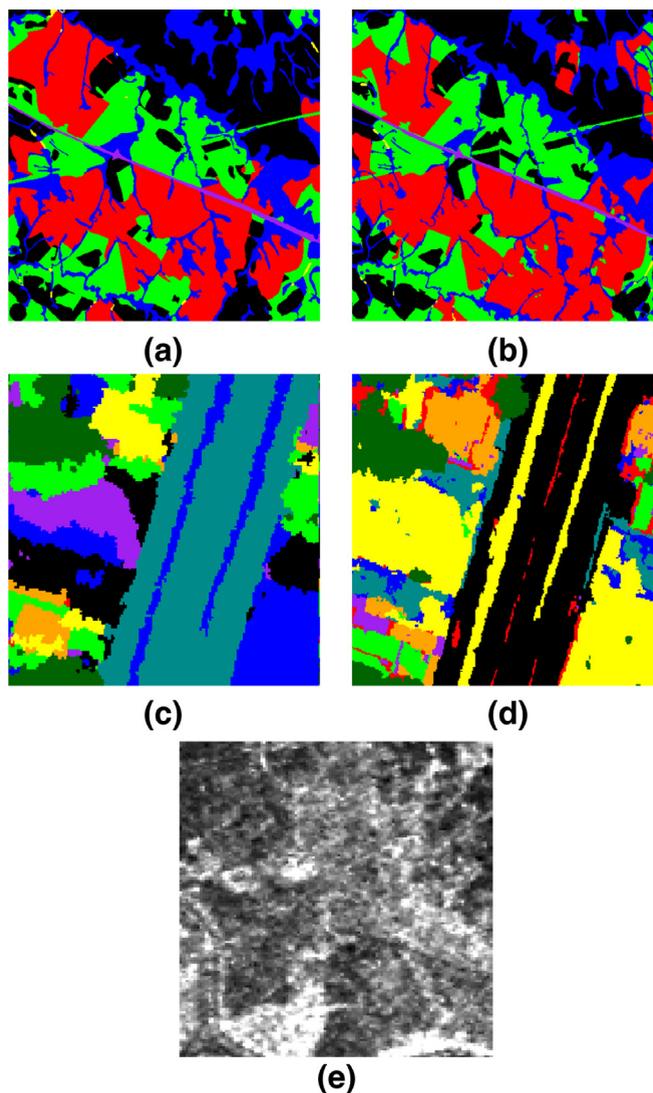
**Fig. 4.** Satellite images used in the experiments: covering the area of Itatinga, SP, Brazil by (a) CBERS-2B CCD (20 m) sensor (R2G3B4) and (b) Landsat 5 TM (30 m) sensor (R4G3B5), and covering the area of Duque de Caxias, RJ, Brazil by (c) Ikonos-2 MS sensor (R4G3B2), and (d) Geoeye sensor (R5G4B3), and (e) ALOS-PALSAR image (HH polarization). The CBERS-2B and Landsat 5 TM images have  $526 \times 492$  pixels, and Ikonos-2 MS and Geoeeye images have  $258 \times 250$  and  $268 \times 250$  pixels, respectively. Note that Ikonos-2 MS and Geoeeye images were obtained through a fusion process between the corresponding images from MS (4 m) and PAN (1 m) sensors using the pan-sharpening method. The final image has a spatial resolution of 1 m.

are no longer clear enough to require more iterations. The reason for a different number of iterations for HS is due to its slow convergence process, since just one agent changes its position per iteration. In regard to PSO, all agents update their positions in the same iteration.

The main loop in Lines 5–13 is the core of the proposed framework for  $\beta$  optimization. The quality of each agent (particle/harmony) is assessed in the inner loop (Lines 6–11), being its current position (solution) evaluated through the Lines 7–8: in Line 7, a call to [Algorithm 3](#) outputs the classified image (label map  $\mathcal{L}$ ), which is used to compute the OPF accuracy (Line 8). After that, the agent's fitness value is updated whether its value is better than the current one (Lines 9–11). Lines 12 and 13 retrieve the best fitness value in the swarm and its position, respectively. Finally, the position (solution) of the best agent corresponds to the algorithm's output (Lines 14–15).

#### 4. Experimental setup

In this section, it is described the methodology employed to validate the proposed OPF–MRF against with traditional OPF and SVM, as well as we have implemented an SVM–MRF classifier based on the idea of this work. In regard to the experiments, we used images obtained from CBERS-2B and Landsat 5 TM satellites covering the area of Itatinga, SP, Brazil, and images are obtained



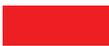
**Fig. 5.** Labeled images used in the experiments: (a) and (b) refer to the images displayed in Fig. 4a and b, respectively, (c) and (d) stand for images displayed in Fig. 4c and d, respectively, and (e) refers to the image displayed in Fig. 4e.

from Ikonos-2 MS and Geosyde satellites covering the area of Duque de Caxias, RJ, Brazil. Additionally, a dual-polarized (HH and HV) image obtained from ALOS-PALSAR radar covering the area of Duque de Caxias, RJ, Brazil is also employed. Fig. 4 displays these images, being their respective ground truth versions illustrated in Fig. 5. Table 1 displays the color associated with each land cover class considering the images used in this work.

In regard to OPF and SVM implementations, we employed the LibOPF [34] and LibSVM [10] packages, respectively. As aforementioned, the standard SVM has been modified to handle SVM-MRF similarly to OPF-MRF, but the basis of the algorithm still relies on LibSVM. In addition, we have used the radial basis function (RBF) kernel for both SVM and SVM-MRF classifiers. Note that the parameters of RBF kernel and SVM-based classifiers have been optimized through a cross-validation procedure.

The experiments have been conducted as follows: for each image we employed 5% of the samples (pixels) to compose the training set  $\mathcal{D}_1$ , 15% to be part of the validating set  $\mathcal{D}_v$ , and the remaining 80% samples to compose the test set  $\mathcal{D}_2$ . Note such percentages have been empirically set. As aforementioned in Section 3, we have executed OPF-MRF and SVM-MRF with  $T = 10$  iterations, being the mean accuracy for each iteration computed over a cross-validation procedure with five runnings. In addition, for each technique, i.e., OPF-MRF and SVM-MRF, we estimated the  $\beta$  parameter using PSO, HS, improved harmony search (IHS) [25], global-best harmony search (GHS) [30] and a near-exhaustive search (BF-Brute Force) within the range  $[0, \beta_{\max}]$  with steps of 0.1. This latter approach has been employed to simulate an “optimal” reference (baseline) for the recognition rate using different  $\beta$  values. Table 2 presents the values of the parameters considering each meta-heuristic optimization technique employed in this work. It follows, below, a brief explanation about each parameter:

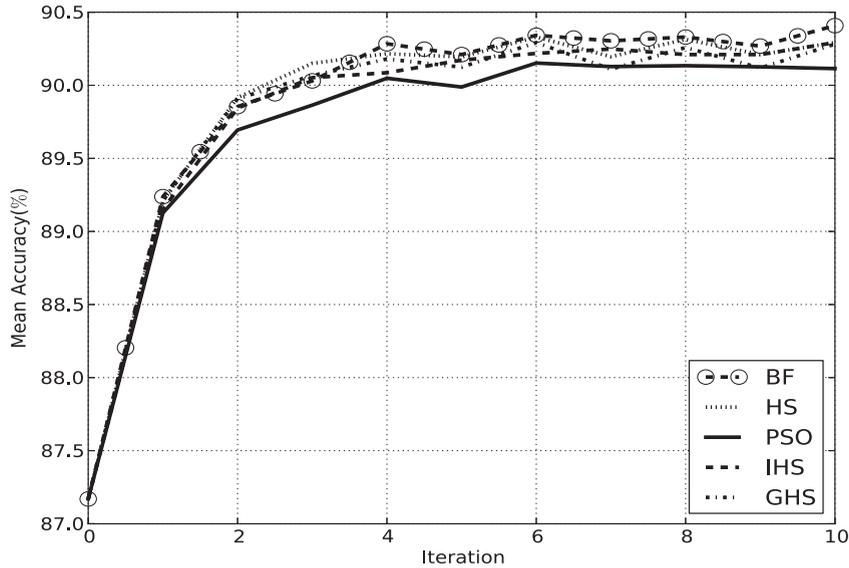
**Table 1**  
Colors associated with each land cover class.

Sensor	Color	Land cover classes
CBERS-2B and Landsat 5		Grass lands
		Bushes
		Cultures
		Reforestation
		Dams
		Roads
Ikonos-2 MS		Shadows
		Grass lands
		Bare soil clear
		Covering of average tonality
		Covering of dark tonality
		Bare soil moist
		Covering of tree
		Covering of clear tonality
Geosyde		Roads
		Covering of average tonality
		Bare soil moist
		Roads
		Shadows
		Grass lands
		Covering of clear tonality
		Covering of tree
ALOS-PALSAR		Covering of dark tonality
		Bare soil clear
		plan areas with bare soil and dry areas
		grass lands with few moist level
		grass lands with average moist level
		swampy and vegetation with very moist levels

**Table 2**

Parameters used for PSO, HS, IHS and GHS: such values have been empirically set.

Technique	Parameters
HS	HMCR = 0.9, PAR = 0.5
IHS	HMCR = 0.9, PAR <sub>min</sub> = 0.1, PAR <sub>max</sub> = 0.9 BW <sub>min</sub> = 0.1, BW <sub>max</sub> = 0.9
GHS	HMCR = 0.9, PAR = 0.5
PSO	$c_1 = c_2 = 2$ , $w = 0.5$



**Fig. 6.** Mean accuracy curve for OPF-MRF with  $\beta$  estimated through PSO, HS, IHS, GHS and BF regarding CBERS-2B image.

- PSO: it has three main parameters, being them the inertia weight  $w$ , which controls the amount of “pace” during the particle’s movement; and  $c_1$  and  $c_2$  parameters, which can control the amount of “imitation” of one particle with respect to its best position, as well as the best position of the whole swarm, respectively;
- HS: it has three main parameters, being them the “harmony memory considering rate” (HMCR), which defines the probability of taking a value from the harmony memory to compose the new harmony; and the “pitch adjusting rate” (PAR) and “bandwidth” (BW), which control the amount of noise added to a possible solution to avoid traps from local optima;
- IHS: it differs from traditional HS by updating the  $PAR \in [PAR_{min}, PAR_{max}]$  and  $BW \in [BW_{min}, BW_{max}]$  values dynamically; and
- GHS: it employs the same modification proposed by IHS with respect to dynamic PAR values. However, it does not employ the concept of bandwidth, being the values of the new harmony based on the ones from the best harmony.

It is important to highlight the above parameters are important, but they are not game-changing when compared to the number of OPF-MRF iterations  $T$ . Since OPF-MRF takes the improvement step over some few iterations, all optimization techniques tend to converge to similar solutions, independent of their initial configuration. Note a special attention has been given to BW parameter considering HS technique. Instead of using a fixed value for that, which may cause slow convergence, we developed the following rule:

$$BW = e^{-x}, \quad (4)$$

such that

$$x = \begin{cases} 2.3 + (t - 1) * 0.092 & \text{if } 1 \leq t \leq 25 \\ 4.6 & \text{otherwise,} \end{cases} \quad (5)$$

in which  $t$  stands for the iteration number, such that  $1 \leq t \leq T_{opt} = 50$  (main loop in Lines 5–13 from Algorithm 4). The value for  $T_{opt}$  has been described in Section 3.

The reason behind Eqs. (4) and (5) is due to the following: the idea is to start with high values for BW ( $BW = e^{-2.3} = 0.1$  for iteration 1) until the algorithm reaches the half of its “life time”, i.e.  $t = 25$  for a total of 50 iterations. Therefore, while  $x$  increases in steps of 0.092,  $e^{-x}$  decreases until  $t = 25$ . After that, i.e., for  $26 \leq t \leq 50$ , we limited BW as being a static parameter with value  $BW = e^{-4.6} = 0.01$ . Thus, the main idea is to allow a faster convergence in the beginning of HS execution, and then to decrease

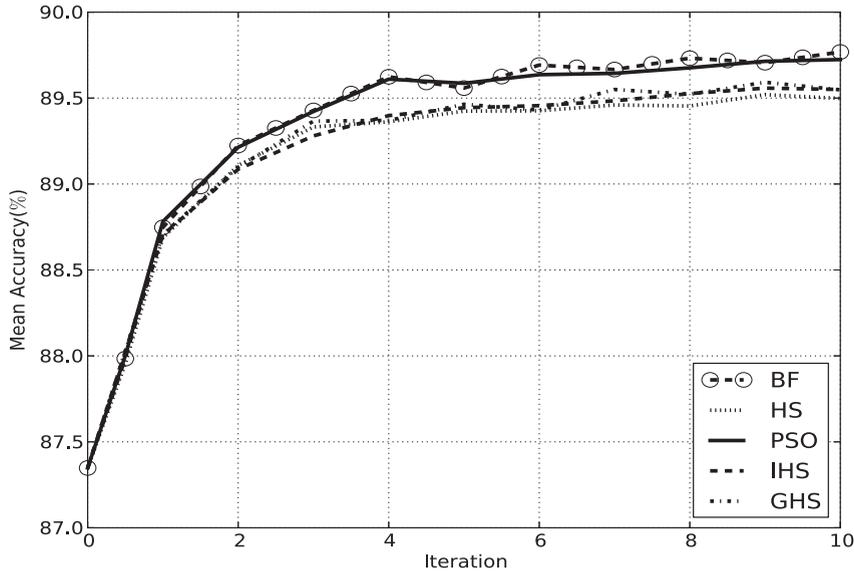


Fig. 7. Mean accuracy curve for SVM–MRF with  $\beta$  estimated through PSO, HS, IHS, GHS and BF regarding CBERS-2B image.

**Table 3**  
Mean execution times for learning  $\beta$ .

Technique		CBERS-2B	Landsat 5	Geosyde	Ikonos-2	ALOS-PALSAR
OPF–MRF	HS	5:18:19	4:48:30	0:17:30	0:15:57	0:21:35
	PSO	6:36:06	5:49:47	0:26:21	0:22:46	0:29:54
	BF	7:42:56	8:07:27	0:31:45	0:29:35	0:30:43
	IHS	4:50:15	4:37:56	0:20:24	0:19:00	0:22:52
SVM–MRF	GHS	2:01:38	01:52:04	0:07:14	0:07:23	0:09:15
	HS	12:19:53	10:52:49	0:44:43	0:51:50	0:11:29
	PSO	13:09:12	11:41:29	0:51:49	0:54:00	0:13:37
	BF	12:45:35	11:15:26	0:53:02	1:01:40	0:15:46
	IHS	12:22:13	11:10:33	0:45:23	0:52:50	0:11:26
	GHS	10:46:13	9:43:24	0:37:24	0:42:52	0:04:37

such behavior in order to avoid traps from local optima around the global solution. The specific values in Eq. (5) were chosen to restrict  $BW \in [0.1, 0.01]$ .

In regard to the recognition rate, we used an accuracy measure proposed by Papa et al. [33], which is similar to the Kappa index [11], but being more restrictive. If there are two classes, for example, with very different sizes and a classifier always assigns the label of the largest class, its accuracy will fall drastically due to the high error rate on the smallest class. The accuracy is measured by taking into account the classes may have different sizes in  $\mathcal{D}_2$  (similar definition is applied for  $\mathcal{D}_V$ ). Let us define:

$$e_{i,1} = \frac{FP_i}{|\mathcal{D}_2| - |\mathcal{D}_2^i|} \tag{6}$$

and

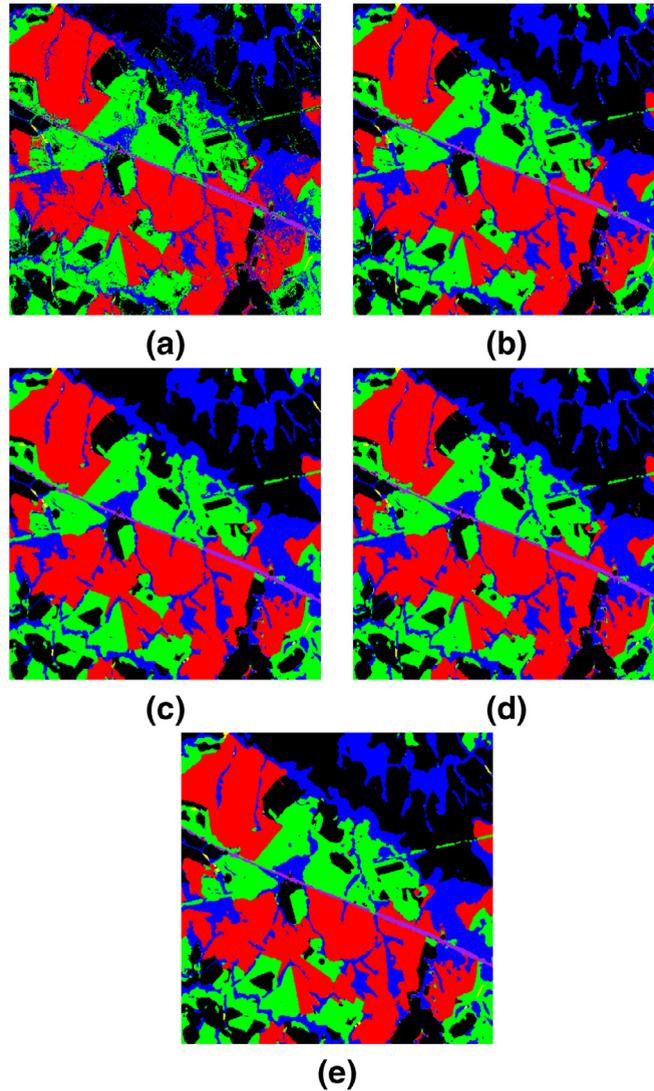
$$e_{i,2} = \frac{FN_i}{|\mathcal{D}_2^i|}, i = 1, 2, \dots, K, \tag{7}$$

where  $K$  stands for the number of classes,  $|\mathcal{D}_2^i|$  concerns with the number of samples in  $\mathcal{D}_2$  that come from class  $i$ , and  $FP_i$  and  $FN_i$  stand for the false positives and false negatives for class  $i$ , respectively. That is,  $FP_i$  is the number of samples from other classes that were classified as being from the class  $i$  in  $\mathcal{D}_2$ , and  $FN_i$  is the number of samples from the class  $i$  that were incorrectly classified as being from other classes in  $\mathcal{D}_2$ . The error terms  $e_{i,1}$  and  $e_{i,2}$  are then used to define the total error from class  $i$ :

$$E_i = e_{i,1} + e_{i,2}. \tag{8}$$

Finally, the accuracy Acc is then defined as follows:

$$Acc = 1 - \frac{\sum_{i=1}^K E_i}{2K}. \tag{9}$$



**Fig. 8.** CBERS-2B classified images: (a) traditional OPF, (b) OPF-MRF with  $\beta$  estimated by PSO, (c) OPF-MRF with  $\beta$  estimated by HS, (d) OPF-MRF with  $\beta$  estimated by IHS and (e) OPF-MRF with  $\beta$  estimated by GHS.

**Table 4**

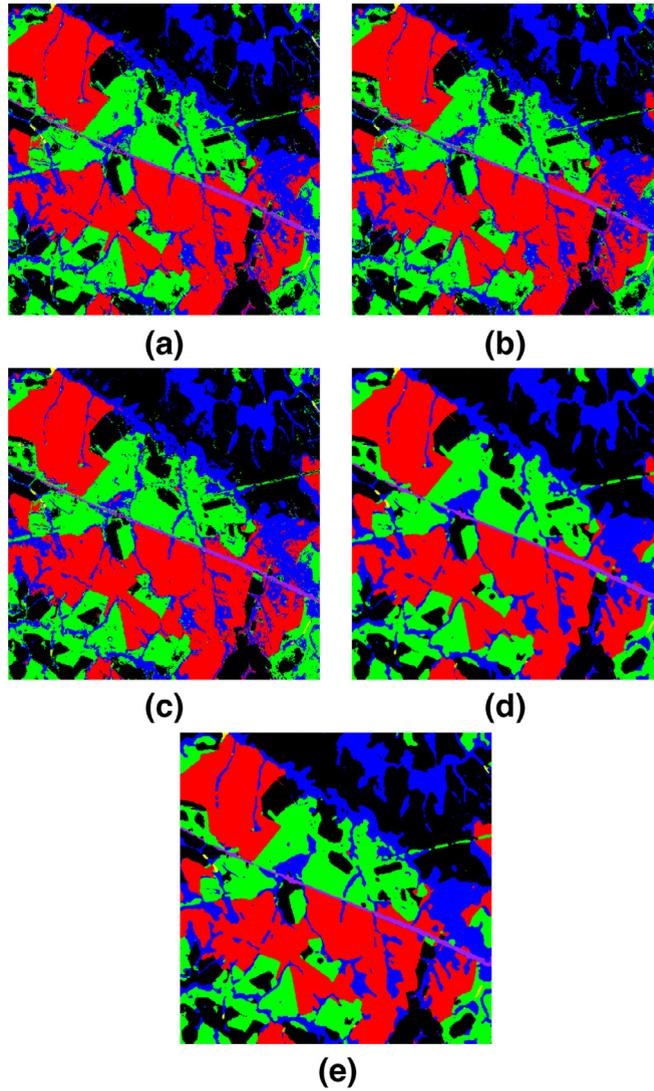
Statistical analysis regarding OPF-based classifiers over CBERS-2B image. Similar techniques are in bold.

$\chi^2$	OPF	HS	IHS	GHS	PSO	BF
OPF	–	5055.2609	5018.7463	5079.8849	4901.2751	5014.5422
HS	5055.2609	–	47.9359	<b>0.0396</b>	83.3416	<b>10.49462</b>
IHS	5018.7463	47.9359	–	50.8094	11.2752	19.1954
GHS	5079.8849	<b>0.0396</b>	50.8094	–	88.1695	11.7650
PSO	4901.2751	83.3416	11.27524	88.1695	–	47.5805
BF	5014.5422	<b>10.4946</b>	19.1954	11.7650	47.5805	–

In order to perform statistical analysis, we compare all possible pairwise combination of classifiers using McNemar test [22], which is described by Eq. (10):

$$\chi^2 = \frac{(|N_{01} - N_{10}| - 1)^2}{N_{01} + N_{10}}, \quad (10)$$

where  $N_{01}$  represents the number of times that first classifier misclassifies and the second one gives a correct classification, and  $N_{10}$  stands for the number of times the first classifier assigns the correct label to a given sample, and the second one misclassifies.



**Fig. 9.** CBERS-2B classified images: (a) traditional SVM, (b) SVM-MRF with  $\beta$  estimated by PSO, (c) SVM-MRF with  $\beta$  estimated by HS, (d) SVM-MRF with  $\beta$  estimated by IHS and (e) SVM-MRF with  $\beta$  estimated by GHS.

**Table 5**

Statistical analysis regarding SVM-based classifiers over CBERS-2B image. Similar techniques are in bold.

$\chi^2$	SVM	HS	IHS	GHS	PSO	BF
SVM	–	2426.4230	2541.7600	2541.7600	2336.4964	2381.6485
HS	2426.4230	–	<b>3.7832</b>	<b>3.7832</b>	<b>4.0742</b>	<b>1.4819</b>
IHS	2541.7600	<b>3.7832</b>	–	<b>0.0000</b>	13.0907	<b>7.9324</b>
GHS	2541.7600	<b>3.7832</b>	<b>0.0000</b>	–	13.0907	<b>7.9324</b>
PSO	2336.4964	<b>4.0742</b>	13.0907	13.0907	–	<b>0.9674</b>
BF	2381.6485	<b>1.4819</b>	<b>7.9324</b>	<b>7.9324</b>	<b>0.9674</b>	–

As the null hypothesis, we assume that both classifiers have similar performance. If the value of  $\chi^2$  is greater than 10.83 ( $p = 0.001$ ), it is possible to reject the null hypothesis and the classifiers have distinct performance.<sup>3</sup>

<sup>3</sup> Note that all experiments have been conducted on a PC equipped with an Intel CPU i3® 540 3.07 GHZ, 4 GB RAM, and Ubuntu 10.04 LTS as the operational system.

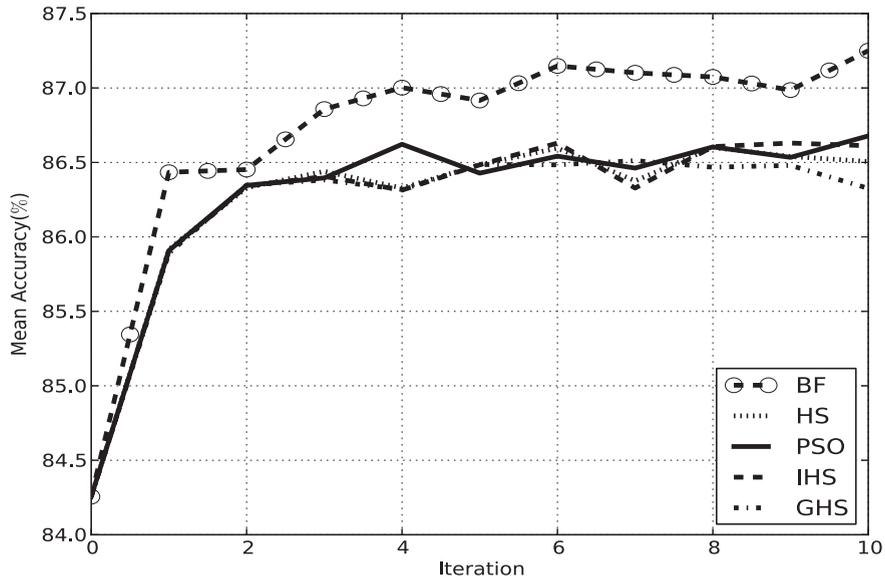


Fig. 10. Mean accuracy curve for OPF-MRF with  $\beta$  estimated through PSO, HS, IHS, GHS and BF regarding Landsat 5 TM image.

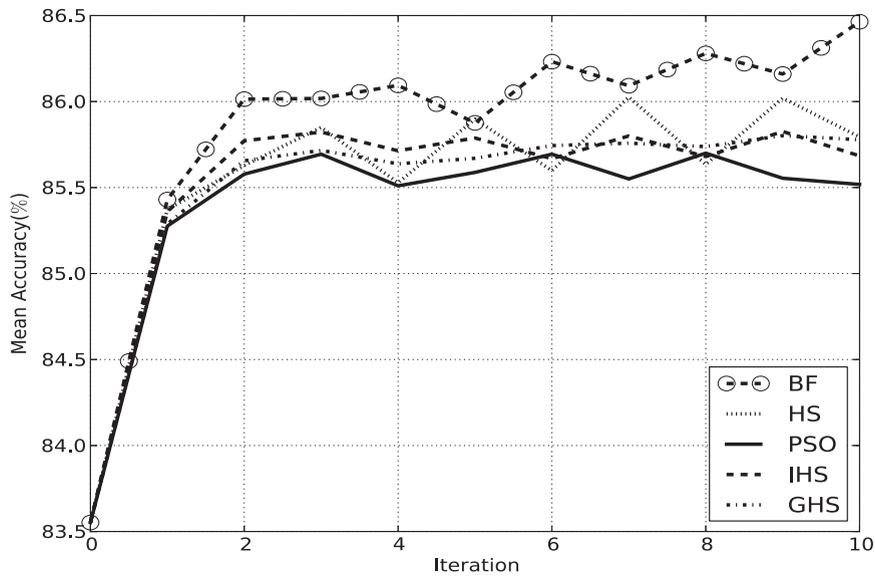
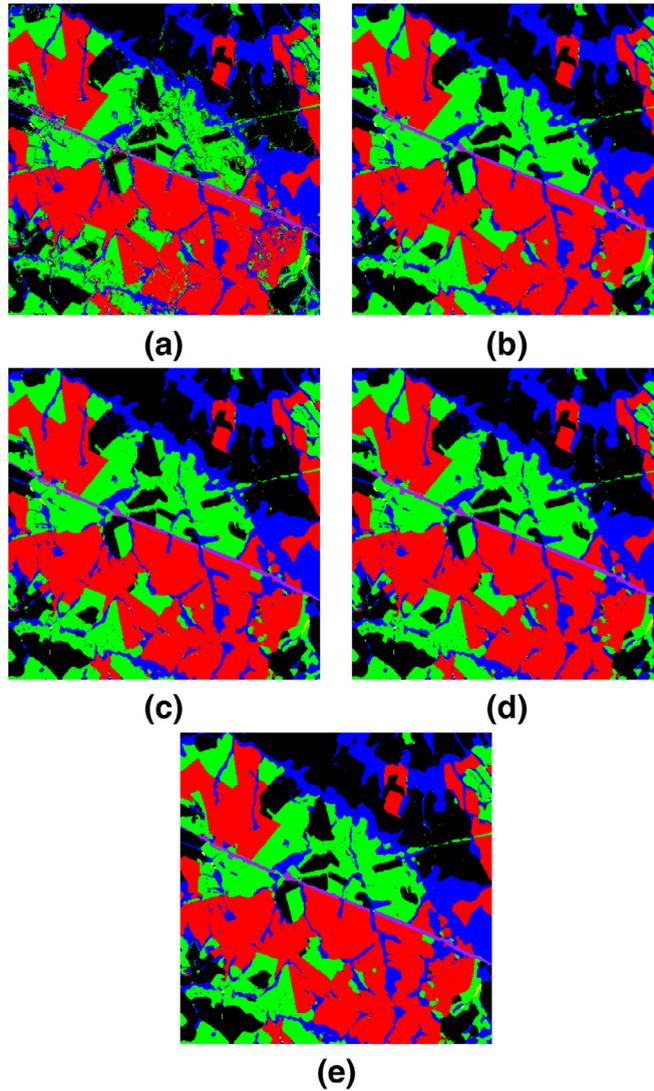


Fig. 11. Mean accuracy curve for SVM-MRF with  $\beta$  estimated through PSO, HS, IHS, GHS and BF regarding Landsat 5 TM image.

## 5. Experiments

In this section, the experimental results using the methodology described in Section 4 are discussed. As aforementioned, OPF-MRF is compared against with standard OPF and SVM classifiers, as well as SVM-MRF. In order to make the results clear, experimental section is divided according to each image employed in this work. It is worth noting to clarify the procedure used for all images: first of all, for each iteration of Algorithm 3, a cross-validation with five runnings has been used for mean accuracy computation. In addition, for each cross-validation running,  $\beta$  is estimated using BF, as well as using PSO and HS-based techniques through Algorithm 4 (the same number of iterations has been used for Algorithms 3 and 4, i.e.,  $T = 10$ , as well as the same training, validating and test sets were used for all techniques).



**Fig. 12.** Landsat 5 TM classified images: (a) traditional OPF, (b) OPF-MRF with  $\beta$  estimated by PSO, (c) OPF-MRF with  $\beta$  estimated by HS, (d) OPF-MRF with  $\beta$  estimated by IHS and (e) OPF-MRF with  $\beta$  estimated by GHS.

**Table 6**

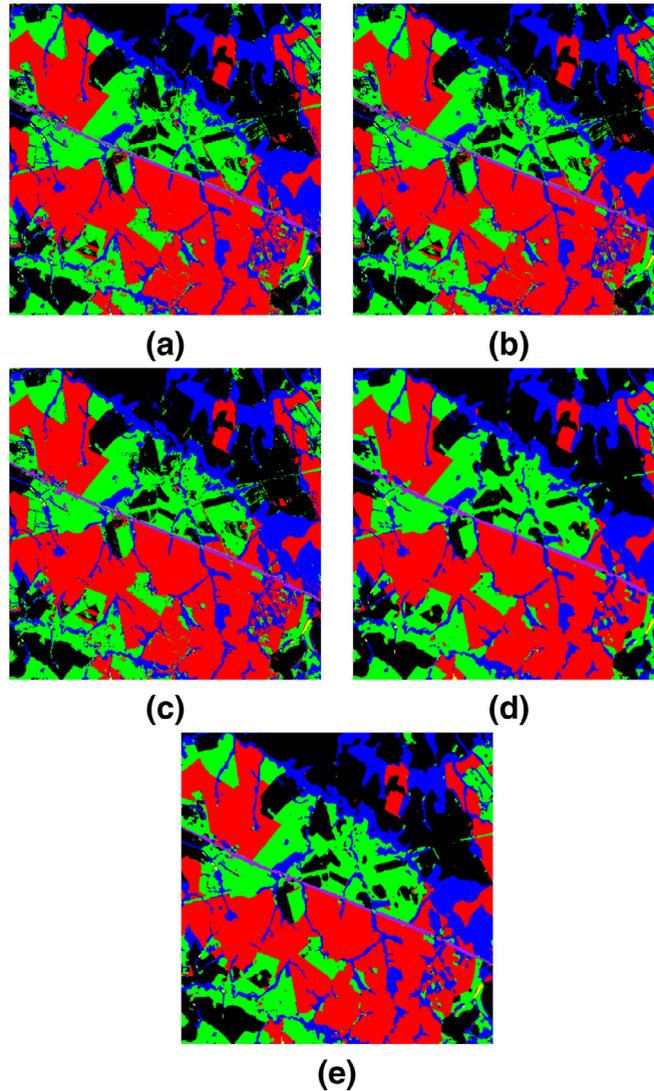
Statistical analysis regarding OPF-based classifiers over Landsat 5 image. Similar techniques are in bold.

$\chi^2$	OPF	HS	IHS	HS	PSO	BF
OPF	–	3392.9098	3406.9396	3355.1024	3418.9078	3576.0901
HS	3392.9098	–	<b>0.6946</b>	<b>1.0466</b>	<b>4.085106</b>	<b>0.0780</b>
IHS	3406.9396	<b>0.6946</b>	–	<b>3.1515</b>	<b>0.1307</b>	<b>0.6277</b>
GHS	3355.1024	<b>1.0466</b>	<b>3.1515</b>	–	<b>2.0459</b>	<b>0.1898</b>
PSO	3418.9078	<b>4.0851</b>	<b>0.1307</b>	<b>2.0459</b>	–	<b>0.3243</b>
BF	3576.0901	<b>0.0780</b>	<b>0.6277</b>	<b>0.1898</b>	<b>0.3243</b>	–

### 5.1. CBERS-2B image

This section presents the results obtained over the CBERS-2B image (Fig. 4a). Fig. 6 displays the mean accuracy curve over the iterations for OPF-MRF with  $\beta$  estimated using PSO, HS, IHS, GHS and BF. Note that the accuracy stated in iteration 0 stands for the original OPF classification, i.e., the one that does not consider the contextual information.

The reader can extract five important information from Fig. 6: (i) OPF-MRF clearly improves standard OPF classification in the first iteration, (ii) the OPF-MRF accuracy during the iterations usually increases until its convergence (iteration 10), (iii) both



**Fig. 13.** Landsat 5 TM classified images: (a) traditional SVM, (b) SVM-MRF with  $\beta$  estimated by PSO, (c) SVM-MRF with  $\beta$  estimated by HS, (d) SVM-MRF with  $\beta$  estimated by IHS and (e) SVM-MRF with  $\beta$  estimated by GHS.

**Table 7**

Statistical analysis regarding SVM-based classifiers over Landsat 5 image. Similar techniques are in bold.

$\chi^2$	SVM	HS	IHS	GHS	PSO	BF
SVM	–	1842.3732	1774.2948	1675.0094	1634.8654	1828.2435
HS	1842.3732	–	<b>3.5124</b>	18.6501	25.0703	<b>0.0274</b>
IHS	1774.2948	<b>3.5124</b>	–	<b>9.2691</b>	13.9426	<b>2.9302</b>
GHS	1675.0094	18.6501	<b>9.2691</b>	–	<b>1.1287</b>	17.4896
PSO	1634.8654	25.0703	13.9426	<b>1.1287</b>	–	23.6006
BF	1828.2435	<b>0.0274</b>	<b>2.9302</b>	17.4896	23.6006	–

PSO and HS-based approaches are able to find out similar values to the ones (“optimal”) obtained by BF (baseline), (iv) the mean accuracies obtained by HS are slightly better than the ones obtained by PSO, and finally (v) HS with the proposed schema to compute BW dynamically (Eqs. (4) and (5)) achieved slightly better results than IHS and GHS. Additionally, Fig. 7 displays the SVM-MRF accuracies over the number of iterations. Once again, the  $\beta$  values found by HS and PSO have been quite similar to the ones obtained by BF technique. From this figure, three important information can be highlighted: (i) OPF-MRF results are

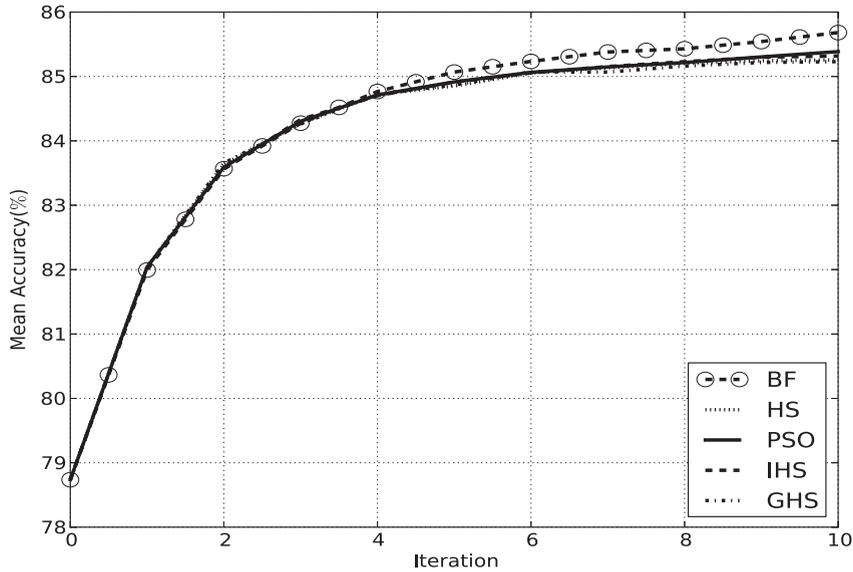


Fig. 14. Mean accuracy curve for OPF-MRF with  $\beta$  estimated through PSO, HS, IHS, GHS and BF regarding Ikonos-2 image.

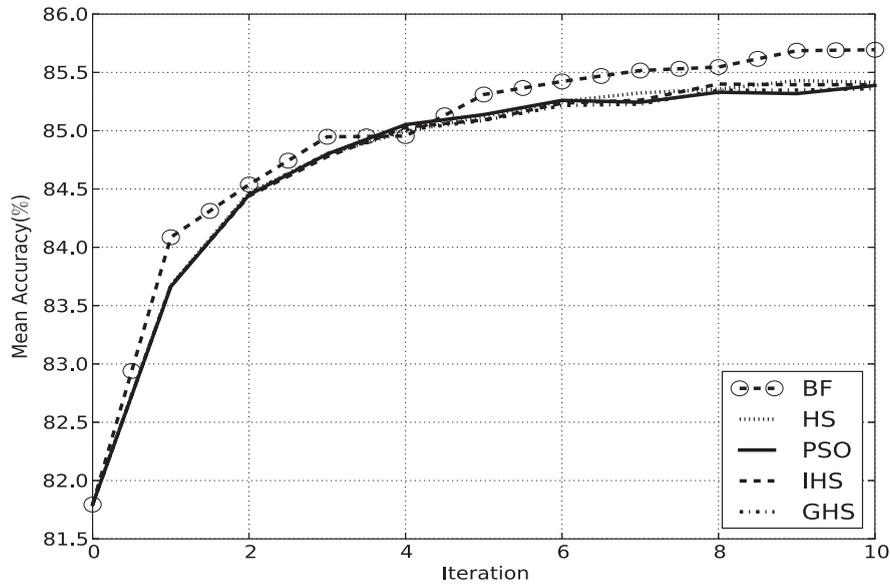
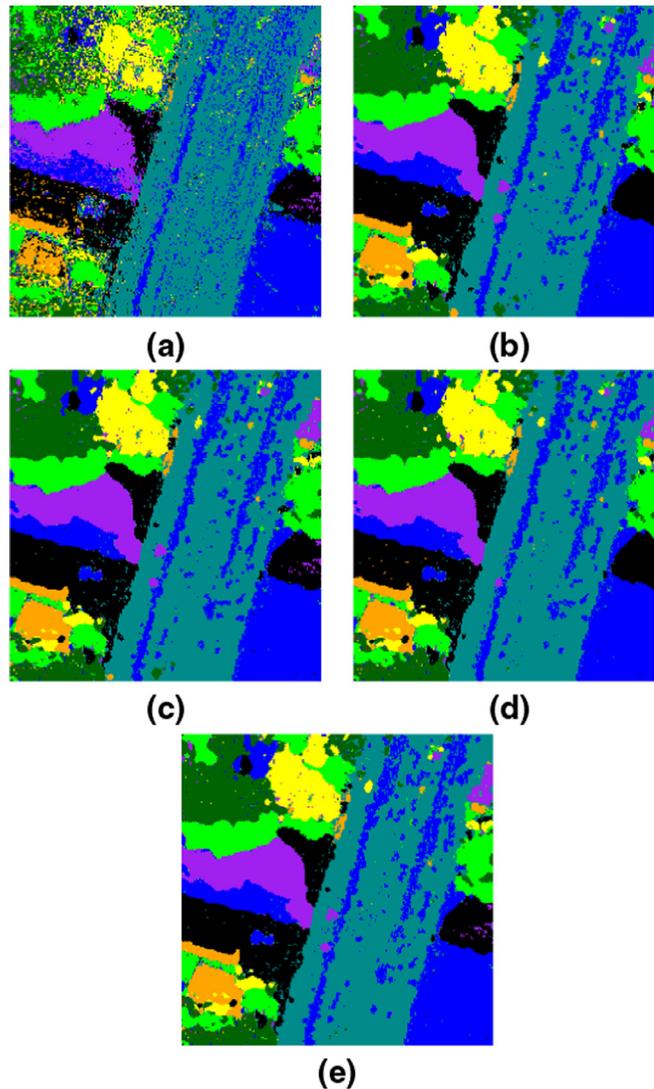


Fig. 15. Mean accuracy curve for SVM-MRF with  $\beta$  estimated through PSO, HS, IHS, GHS and BF regarding Ikonos-2 image.

slightly better than SVM-MRF, (ii) SVM-MRF slightly improved standard SVM results,<sup>4</sup> and (iii) PSO obtained better results than HS. Therefore, it is possible to draw two important aspects points: (i) the proposed contextual version of OPF has improved its standard version, (ii) as well as the proposed framework for spatial-content optimization worked well for both OPF and SVM classifiers, evidencing the second main contribution of this work.

Table 3 displays the mean execution times considering OPF-MRF and SVM-MRF for learning  $\beta$  in the following format:  $h:m:s$ , in which  $h$  and  $m$  stand for hours and minutes, respectively, and  $s$  denotes the number of seconds. It turns out, for all images, both PSO and HS-based approaches have been faster than BF considering OPF-MRF, which make them interesting approaches to avoid overload on empirical evaluations of  $\beta$ . Although HS-based techniques have a lower complexity than PSO, they need more iterations to converge, since just one agent (harmony) is changed at each iteration, while PSO changes all of them (particles). In addition, it is possible to observe SVM-MRF has been slower than OPF-MRF, since we are considering the computational load for

<sup>4</sup> Note Algorithm 3 can be also applied to build the SVM-MRF classifier: the idea is just to change OPF by SVM in Lines 2–3 and 9–10.



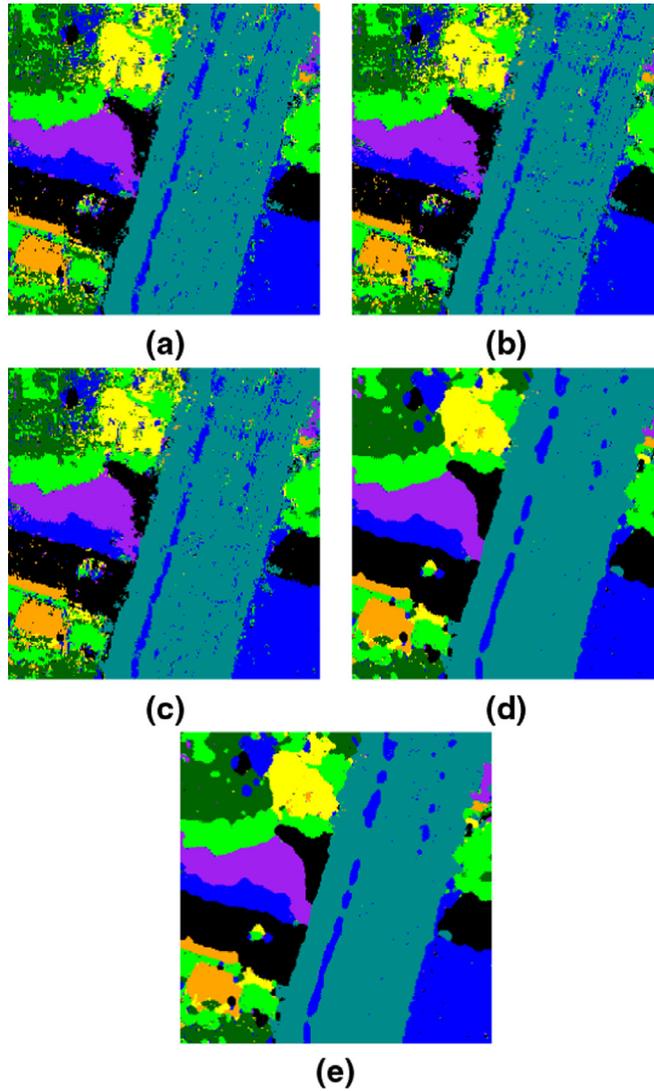
**Fig. 16.** Ikonos-2 classified images: (a) traditional OPF, (b) OPF-MRF with  $\beta$  estimated by PSO, (c) OPF-MRF with  $\beta$  estimated by HS, (d) OPF-MRF with  $\beta$  estimated by IHS and (e) OPF-MRF with  $\beta$  estimated by GHS.

optimizing SVM parameters. The only situation that meta-heuristic-based optimization needed more computational load than BF concerns with SVM-MRF optimized by PSO.

Figs. 8 and 9 display the CBERS-2B classified images using OPF and SVM classifiers, respectively, as well as their contextual versions (OPF-MRF and SVM-MRF) with  $\beta$  estimated by PSO and HS-based techniques. A good recognition result for “grasslands” and “reforesting” classes can be observed. However, it was noticed a confusion with respect to “bushes” and “cultures” to “grasslands”, and “bushes” and “roads” classes. In regard to SVM image classification, it can be verified some few errors: a confusion from “culture” to “roads” and “bushes”. The “reforesting” and “grasslands” classes showed better recognition rates. Similar results were noticed for OPF-MRF with  $\beta$  estimated by HS, but with a confusion from “culture” to other classes. SVM-MRF showed a mixture between “culture” and “reforesting” with “bushes”, “grasslands” and “roads”; “roads” were not recognized. Similar results were observed for OPF-MRF with  $\beta$  estimated by PSO and OPF-MRF with  $\beta$  estimated by HS.

Tables 4 and 5 present the McNemar statistical test for CBERS-2B image classification results regarding OPF-based and SVM-based techniques, respectively. From Table 4, it is possible to observe two main information: (i) OPF-MRF has improved standard OPF using PSO, HS and BF approaches (the values of  $\chi^2$  are greater than 10.83, as previously stated in Section 4), and (ii) the pairs of techniques that seemed to have similar performance were that of (BF, HS) and (HS, GHS) (bolded).

In regard to SVM-based results, from Table 5 it is possible to observe that SVM-MRF has outperformed standard SVM, as well as both HS-based and PSO techniques were able to find out reasonable values for  $\beta$ , since such techniques were considered statistically similar to that of BF.



**Fig. 17.** Ikonos-2 classified images: (a) traditional SVM, (b) SVM-MRF with  $\beta$  estimated by PSO, (c) SVM-MRF with  $\beta$  estimated by HS, (d) SVM-MRF with  $\beta$  estimated by IHS and (e) SVM-MRF with  $\beta$  estimated by GHS.

**Table 8**

Statistical analysis regarding OPF-based classifiers over Ikonos-2 image. Similar techniques are in bold.

$\chi^2$	OPF	HS	IHS	GHS	PSO	BF
OPF	–	2899.7008	3273.1485	2900.9332	2727.6816	4025.6730
HS	2899.7008	–	34.3607	<b>0.2874</b>	11.5361	262.8288
IHS	3273.1485	34.3607	–	30.5201	73.0151	147.5630
GHS	2900.9332	<b>0.2874</b>	30.5201	–	13.7118	260.9796
PSO	2727.6816	11.5361	73.0151	13.7118	–	321.392031
BF	4025.6730	262.828804	147.5630	260.9796	321.3920	–

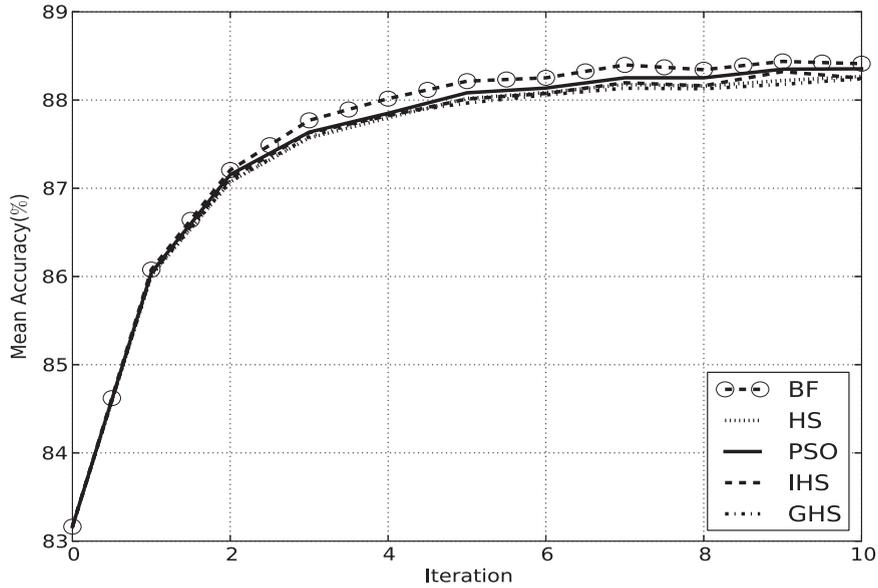
## 5.2. Landsat 5 TM image

In this section, the experimental validation over Landsat image (Fig. 4b) is presented. Fig. 10 displays the mean accuracy curve over the iterations for OPF-MRF with  $\beta$  estimated using PSO, HS, IHS, GHS and BF. Clearly, BF has outperformed the meta-heuristic techniques, being PSO slightly better than HS-based techniques. However, such improvement regarding PSO may not be suitable, since it has a higher computational burden, as shown in Table 3. The reason is that HS-based approaches execute the

**Table 9**

Statistical analysis regarding SVM-based classifiers over Ikonos-2 image. Similar techniques are in bold.

$\chi^2$	SVM	HS	IHS	GHS	PSO	BF
SVM	–	1251.1060	1288.0363	1288.0363	1319.2329	1396.8101
HS	1251.1060	–	<b>2.1485</b>	<b>2.1485</b>	<b>4.8520</b>	22.9165
IHS	1288.0363	<b>2.1485</b>	–	<b>0.0000</b>	<b>0.6677</b>	13.7734
GHS	1288.0363	<b>2.1485</b>	<b>0.0000</b>	–	<b>0.6677</b>	13.7734
PSO	1319.2329	<b>4.8520</b>	<b>0.6677</b>	<b>0.6677</b>	–	<b>10.3854</b>
BF	1396.8101	22.9165	13.7734	13.7734	<b>10.3854</b>	–

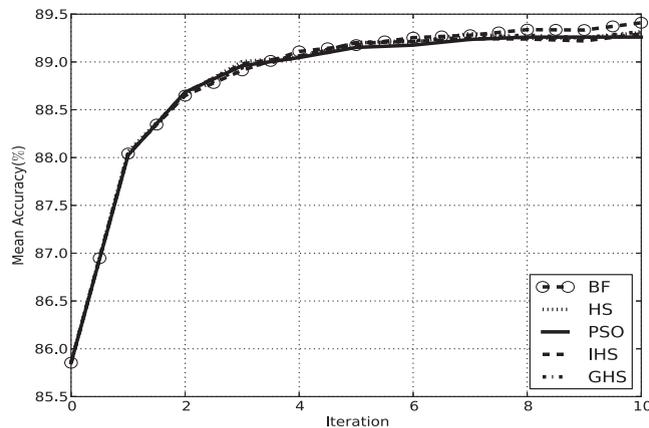


**Fig. 18.** Mean accuracy curve for OPF-MRF with  $\beta$  estimated through PSO, HS, IHS, GHS and BF regarding Geoeye image.

**Table 10**

Statistical analysis regarding OPF-based classifiers over Geoeye image. Similar techniques are in bold.

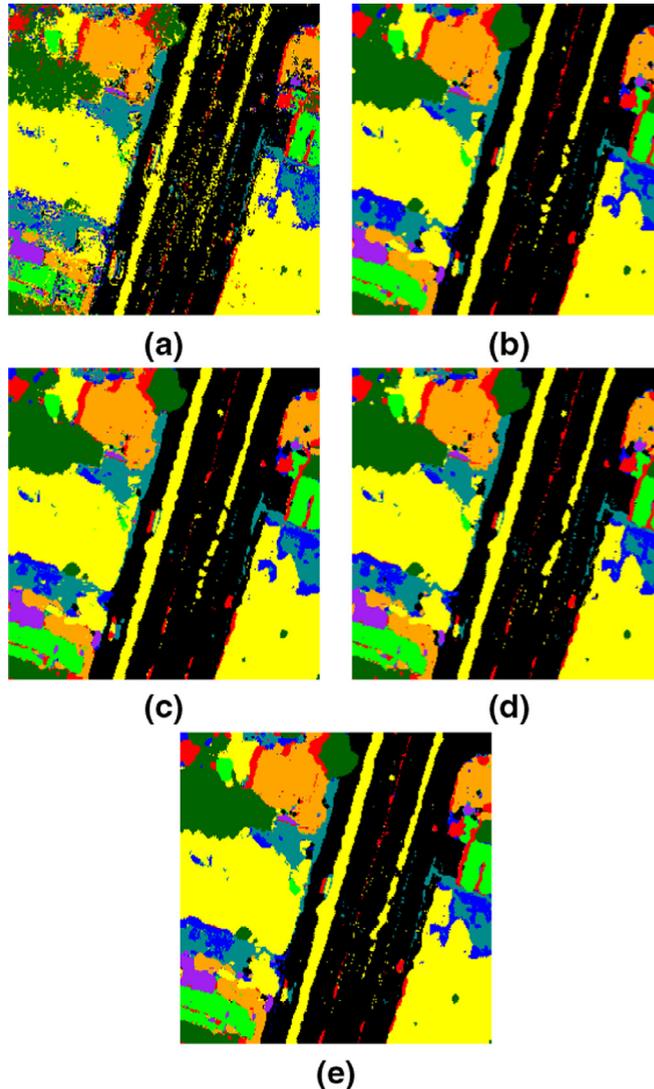
$\chi^2$	OPF	HS	IHS	GHS	PSO	BF
OPF	–	1751.3133	1780.2405	1798.6286	1825.1182	1788.4409
HS	1751.3133	–	<b>0.0216</b>	0.0079	<b>0.7689</b>	<b>0.0005</b>
IHS	1780.2405	<b>0.02166</b>	–	<b>0.0964</b>	<b>1.3823</b>	<b>0.0152</b>
GHS	1798.6286	<b>0.0079</b>	<b>0.0964</b>	–	<b>0.8444</b>	<b>0.0257</b>
PSO	1825.1182	<b>0.7689</b>	<b>1.3823</b>	<b>0.8444</b>	–	<b>2.4362</b>
BF	1788.4409	<b>0.0005</b>	<b>0.0152</b>	<b>0.0257</b>	<b>2.4362</b>	–



**Fig. 19.** Mean accuracy curve for SVM-MRF with  $\beta$  estimated through PSO, HS, IHS, GHS and BF regarding Geoeye image.

**Table 11**  
Statistical analysis regarding SVM-based classifiers over Geoeye image. Similar techniques are in bold.

$\chi^2$	SVM	HS	IHS	GHS	PSO	BF
SVM	–	886.6530	858.3420	858.3420	882.1317	872.5408
HS	886.6530	–	<b>2.5739</b>	<b>2.5739</b>	<b>0.0671</b>	<b>0.4419</b>
IHS	858.3420	<b>2.5739</b>	–	<b>0.0000</b>	<b>1.9760</b>	<b>0.3081</b>
GHS	858.3420	<b>2.5739</b>	<b>0.0000</b>	–	<b>1.9760</b>	<b>0.3081</b>
PSO	882.1317	<b>0.0671</b>	<b>1.9760</b>	<b>1.9760</b>	–	<b>0.2853</b>
BF	872.5408	<b>0.4419</b>	<b>0.3081</b>	<b>0.3081</b>	<b>0.2853</b>	–

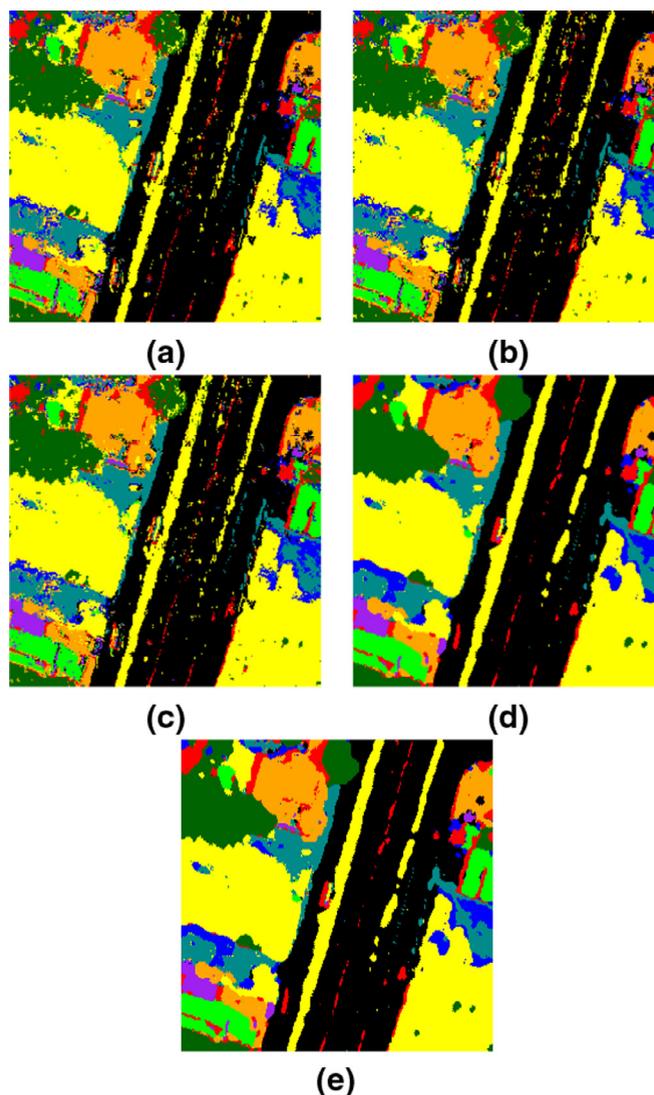


**Fig. 20.** Geoeye classified images: (a) traditional OPF, (b) OPF-MRF with  $\beta$  estimated by PSO, (c) OPF-MRF with  $\beta$  estimated by HS, (d) OPF-MRF with  $\beta$  estimated by IHS and (e) OPF-MRF with  $\beta$  estimated by GHS.

fitness function (OPF-MRF procedure) only once at each iteration, while PSO executes the fitness function as many particles it contains. Although it may converge faster, it usually requires more computational load.

Fig. 11 displays the mean accuracy curve considering SVM-MRF with  $\beta$  estimated using PSO, HS, IHS, GHS and BF. Basically, the upper results are bounded by BF, which obtained the highest accuracy rates, and the lower results are bounded by PSO, that achieved the lowest recognition rates. However, the results do not differ a lot from each other.

Figs. 12 and 13 display the classified images with  $\beta$  estimated by meta-heuristic and BF techniques considering OPF- and SVM-MRF classifiers, respectively. Once again, OPF- and SVM-MRF obtained better results than their vanilla versions, as well as



**Fig. 21.** Geoeeye classified images: (a) traditional SVM, (b) SVM-MRF with  $\beta$  estimated by PSO, (c) SVM-MRF with  $\beta$  estimated by HS, (d) SVM-MRF with  $\beta$  estimated by IHS and (e) SVM-MRF with  $\beta$  estimated by GHS.

**Table 12**

Statistical analysis regarding OPF-based classifiers over ALOS-PALSAR image. Similar techniques are in bold.

$\chi^2$	OPF	HS	IHS	GHS	PSO	BF
OPF	–	2724.9776	2641.7124	2734.4859	2592.8110	2853.7568
HS	2724.9775	–	<b>2.6495</b>	<b>0.0033</b>	<b>6.7906</b>	<b>6.9341</b>
IHS	2641.7124	<b>2.6495</b>	–	<b>2.6675</b>	<b>1.3819</b>	16.1200
GHS	2734.4859	<b>0.0033</b>	<b>2.6675</b>	–	<b>6.5583</b>	<b>6.6103</b>
PSO	2592.8110	<b>6.7906</b>	<b>1.3819</b>	<b>6.5583</b>	–	23.2619
BF	2853.7568	<b>6.9341</b>	16.1200	<b>6.6103</b>	23.2619	–

the images classified with  $\beta$  optimized by GHS presented a better visual quality, mainly with respect to the bottom-right portion of the images (Tables 6 and 7).

Considering the statistical results with OPF-MRF, the contextual version outperformed standard OPF, being all meta-heuristic techniques considered similar to each other. In regard to SVM-MRF, a similarity between the pairs (HS, BF), (HS, IHS), (IHS, GHS), (GHS, PSO), (BF, HS) and (BF, IHS) can be observed. Therefore, BF, HS and IHS have been the most accurate approaches, thus being HS and IHS a good choice considering the trade-off between effectiveness and efficiency (Table 3).

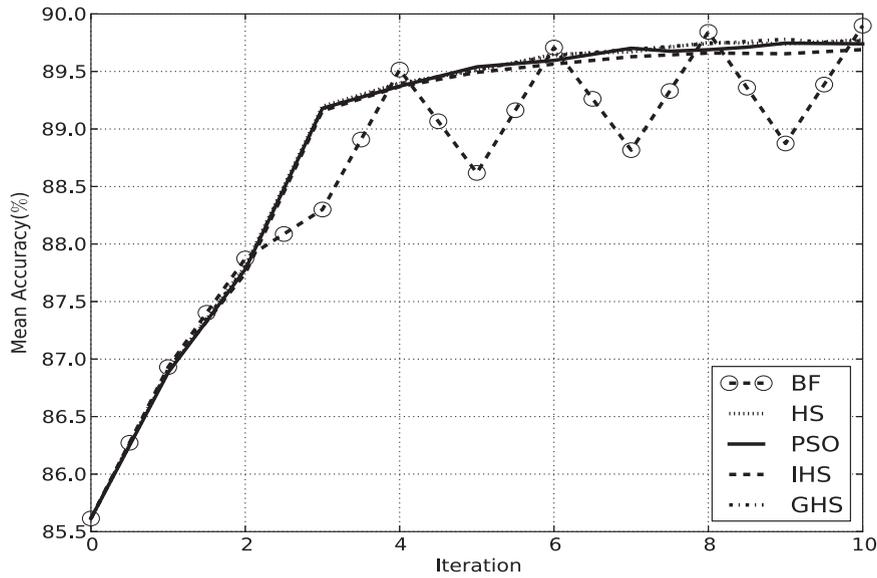


Fig. 22. Mean accuracy curve for OPF-MRF with  $\beta$  estimated through PSO, HS, IHS, GHS and BF regarding ALOS-PALSAR image.

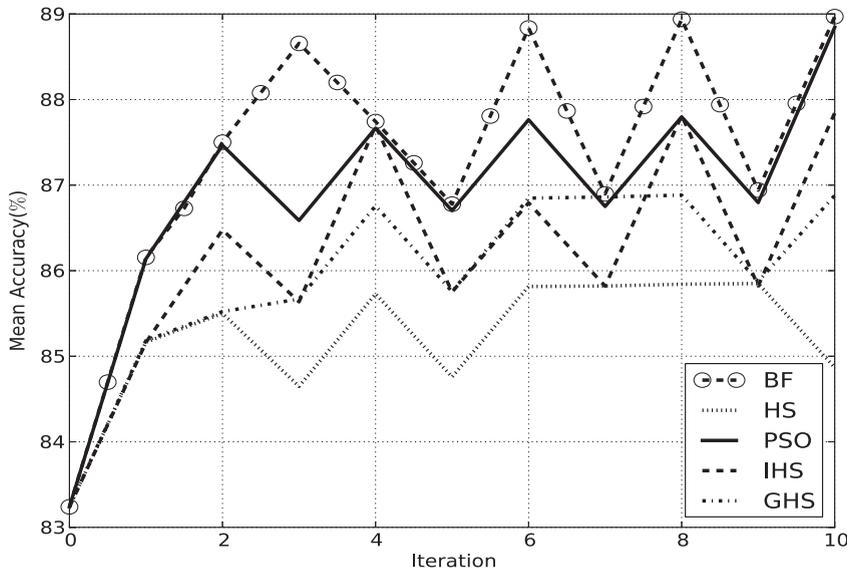


Fig. 23. Mean accuracy curve for SVM-MRF with  $\beta$  estimated through PSO, HS, IHS, GHS and BF regarding ALOS-PALSAR image.

### 5.3. Ikonos-2 image

This section presents the results obtained over Ikonos-2 image (Fig. 4c). Fig. 14 displays the mean accuracy curve over the iterations for OPF-MRF with  $\beta$  estimated using PSO, HS, IHS, GHS and BF. A similar behavior can be observed for all meta-heuristic techniques, being the results obtained by BF slightly better. The gain in the classification accuracy was higher than the previous experiments, showing the proposed OPF-MRF is suitable for learning with contextual information. Once again, the proposed meta-heuristic-based framework was able to find out similar  $\beta$  values regarding BF technique. From Table 3, it is possible to observe all meta-heuristic techniques have been faster than BF considering both OPF- and SVM-MRF.

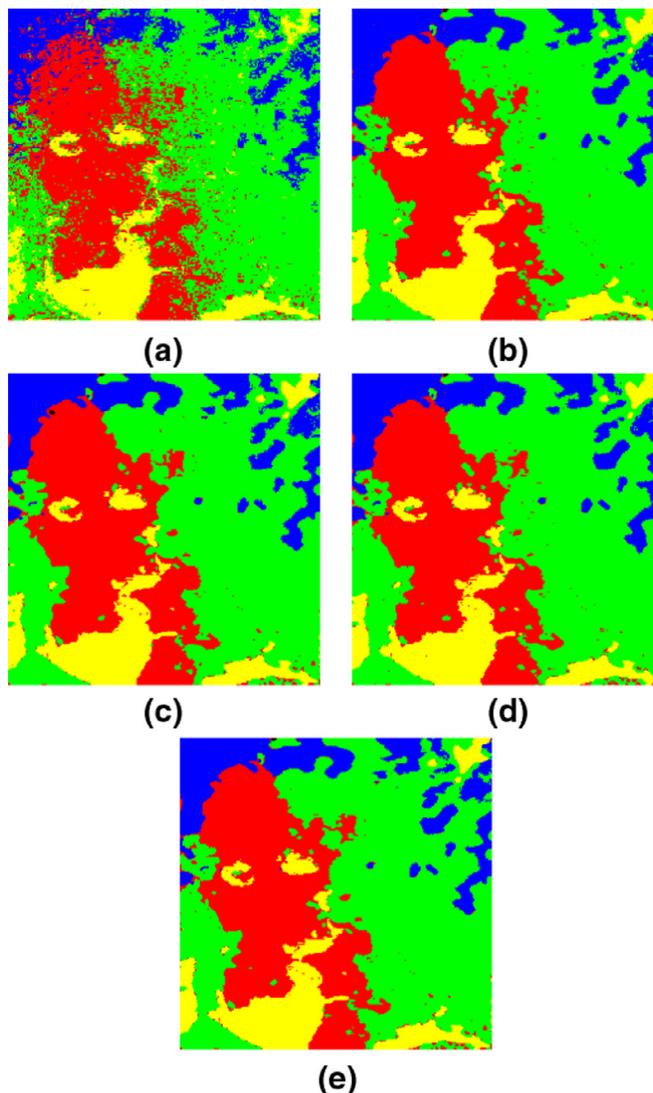
Fig. 15 depicts the SVM-MRF results over Ikonos-2 image. In this case, SVM-MRF has improved standard SVM, being the recognition rates of PSO and HS-based techniques under BF technique (except for iteration 4). Another interesting point can be highlighted here: it seems OPF adapted better to the proposed approach than SVM, since the OPF accuracy at iteration 0 was very close to 79%, while SVM accuracy at the very same iteration was close to 82%. However, the OPF accuracy at the last iteration was slightly higher than SVM recognition rate.

Figs. 16 and 17 display the Ikonos-2 classified images using OPF and SVM classifiers, respectively, as well as their contextual versions (OPF-MRF and SVM-MRF) with  $\beta$  estimated by PSO and HS-based techniques. A confusion in the great majority of classes can be observed, with exception for “covering of trees” and “grasslands” with traditional OPF. OPF-MRF with  $\beta$  estimated

**Table 13**

Statistical analysis regarding SVM-based classifiers over ALOS-PALSAR image. Similar techniques are in bold.

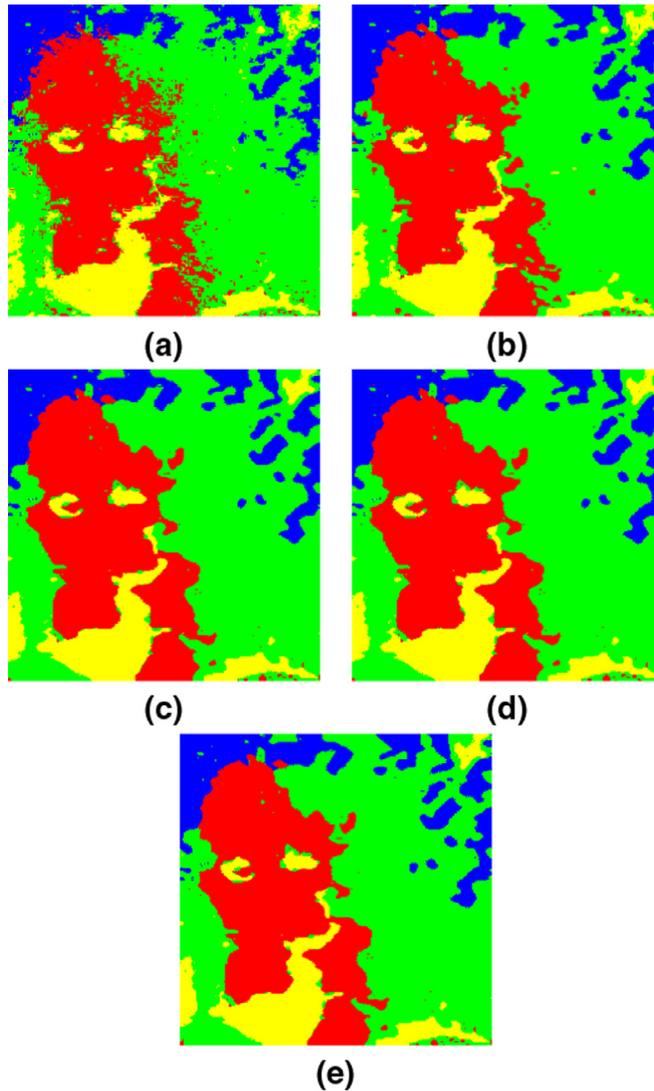
$\chi^2$	SVM	HS	IHS	GHS	PSO	BF
SVM	–	1062.3209	1062.3209	994.7229	648.8286	1048.5504
HS	1062.3209	–	<b>0.0000</b>	<b>6.9963</b>	138.7408	<b>0.9298</b>
IHS	1062.3209	<b>0.0000</b>	–	<b>6.9963</b>	138.7408	<b>0.9298</b>
GHS	994.7229	<b>6.9963</b>	<b>6.9963</b>	–	104.1773	<b>7.3518</b>
PSO	648.8286	138.7408	138.7408	104.1773	–	191.3284
BF	1048.5504	<b>0.9298</b>	<b>0.9298</b>	<b>7.3518</b>	191.3284	–



**Fig. 24.** ALOS-PALSAR classified images: (a) traditional OPF, (b) OPF-MRF with  $\beta$  estimated by PSO, (c) OPF-MRF with  $\beta$  estimated by HS, (d) OPF-MRF with  $\beta$  estimated by IHS and (e) OPF-MRF with  $\beta$  estimated by GHS.

by PSO showed good recognition rates for the majority of classes, with exception for “roads”. It was also observed a confusion from “roads” to other classes, and from “bare soil moist” and “grasslands” to “roads”, with similar results for OPF-MRF with  $\beta$  estimated by HS. In regard to SVM classifier, a mixture between “covering of dark tonality” and “covering of trees”, “bare soil clear” and “covering of clear tonality” has been observed. Similar results for SVM-MRF with  $\beta$  estimated by both PSO and HS can also be noticed.

Tables 8 and 9 present the McNemar statistical test for Ikonos-2 image classification results regarding OPF-based and SVM-based techniques, respectively. From Table 8, the reader can observe OPF-MRF has outperformed standard OPF for  $\beta$  estimation



**Fig. 25.** ALOS-PALSAR classified images: (a) traditional SVM, (b) SVM-MRF with  $\beta$  estimated by PSO, (c) SVM-MRF with  $\beta$  estimated by HS, (d) SVM-MRF with  $\beta$  estimated by IHS and (e) SVM-MRF with  $\beta$  estimated by GHS.

using BF, PSO and HS-based techniques, being only HS and GHS considered similar to each other. In regard to SVM results, the reader can also observe (Table 9) that SVM-MRF with BF and HS have outperformed standard SVM. Considering such classifier, all HS-based techniques and PSO can be considered similar to each other.

#### 5.4. Geoeye image

This section presents the results over the Geoeye satellite image (Fig. 4d). Fig. 18 displays the OPF-MRF accuracy curve with  $\beta$  estimated through BF, PSO and HS-based techniques. Considering this experiment, it is possible to observe all techniques have underlined a similar behavior, which can be noticed in the statistical validation displayed in Table 10 (all pairs of meta-heuristic techniques are in bold).

Fig. 19 displays the mean accuracy curve considering SVM-MRF optimized with BF, PSO and HS-based techniques. A similar behavior considering Fig. 18 can be observed, but with closer recognition rates considering all techniques. Once again, all techniques can be considered similar to each other (Table 11). Figs. 20 and 21 display the images classified through OPF- and SVM-MRF, respectively.

#### 5.5. ALOS-PALSAR image

This section presents the results over ALOS-PALSAR radar image. Fig. 22 displays the accuracy curves regarding OPF-MRF classification with  $\beta$  estimated through the optimization process. A similar behavior among all techniques can be observed,

except for BF, which seemed to oscillate from iterations 4 to 10. The statistical analysis showed in Table 12 evidenced OPF–MRF outperformed standard OPF, being HS similar to all compared techniques.

Fig. 23 depicts the accuracy curve considering SVM–MRF. Once again, the proposed approach outperformed standard classification (iteration 0), but with an oscillating behavior for all techniques. In this case, BF results have been much better than the ones obtained by HS, leading us to conclude the optimization landscape for SVM–MRF fine-tuning is more complex than OPF–MRF one. The statistical analysis presented in Table 13 showed HS and BF obtained similar recognition rates, which may be due to the high standard deviation obtained by HS during the optimization process. In regard to the computational load, if one considers Table 3, GHS has been the fastest technique once again, with PSO computational burden very close to BF.

Figs. 24 and 25 display the ALOS-PALSAR images classified by OPF- and SVM–MRF. Although the images classified by SVM seemed to be more accurate than OPF ones, such data were obtained from a single iteration only, and did not reflect the averaged results displayed in Figs. 22 and 23.

## 6. Conclusions

In this work, it has been presented a contextual version of the OPF classifier called OPF–MRF, which maximizes a Potts model-like energy function together with the path-cost minimization provided by standard OPF classifier. The proposed approach has been evaluated in the context of land cover image classification in four images provided by CBERS-2B, Landsat 5 TM, Ikonos-2 and Geosy satellites, plus one more image obtained by ALOS-PALSAR radar, being OPF–MRF results more accurate than the ones obtained by standard OPF. In addition, it has been proposed a meta-heuristic-based optimization framework to find out suitable values for the Potts model  $\beta$  parameter, which controls the amount of contextual information to be used during the learning process. For such purpose, the well-known particle swarm optimization and three variants of the harmony search were employed, and compared to a baseline provided by a near-exhaustive search over a reasonable range of  $\beta$  values. It is possible to conclude the proposed approach can find out similar values to the “near-optimal” ones computed by the near-exhaustive search, which can still improve OPF–MRF recognition rates. Additionally, the proposed approach was evaluated considering SVM classifier, showing it might work with any pattern recognition technique as well.

Among all optimization techniques, we consider GHS might be most suitable one, since it obtained one of the top results with the lowest computational burden. Additionally, it has less parameters than IHS and PSO, being much faster than the last one. Although PSO obtained better or similar results compared to BF, their computational load is quite similar, since PSO updates all possible solutions at each iteration, thus requiring the execution of OPF–MRF whenever a particle changes its position in the search space. In regard to future works, we aim at addressing other meta-heuristic techniques to optimize  $\beta$ , as well as to combine OPF with conditional random fields and to employ different MRF models.

## Acknowledgments

The authors are grateful to FAPESP grant nos. 2009/16206-1, 2012/06472-9, 2013/20387-7 and 2014/16250-9, as well as CNPq grant nos. 303182/2011-3, 470571/2013-6 and 306166/2014-3.

## References

- [1] H. Aghighi, J. Trinder, Y. Tarabalka, S. Lim, Dynamic block-based parameter estimation for MRF classification of high-resolution images, *IEEE Geosci. Remote Sens. Lett.* 11 (10) (2014) 1687–1691.
- [2] H. Aghighi, J. Trinder, K. Wang, Y. Tarabalka, S. Lim, Smoothing parameter estimation for Markov random field classification of non-Gaussian distribution image, in: *ISPRS TC VII Symposium, Istanbul, Turkey, 2014*.
- [3] N. Alajlan, Y. Bazi, F. Melgani, R.R. Yager, Fusion of supervised and unsupervised learning for improved classification of hyperspectral images, *Inf. Sci.* 217 (0) (2012) 39–55.
- [4] C. Allène, J.-Y. Audibert, M. Couprie, R. Keriven, Some links between extremum spanning forests, watersheds and min-cuts, *Image Vis. Comput.* 28 (10) (2010) 1460–1471.
- [5] P.M. Atkinson, P. Lewis, Geostatistical classification for remote sensing: an introduction, *Comput. Geosci.* 26 (4) (2000) 361–371.
- [6] J. Besag, On the statistical analysis of dirty pictures, *J. R. Stat. Soc., Ser. B (Methodol.)* 48 (3) (1986) 259–302.
- [7] E. Binaghi, I. Gallo, M. Pepe, A cognitive pyramid for contextual classification of remote sensing images, *IEEE Trans. Geosci. Remote Sens.* 41 (12) (2003) 2906–2922.
- [8] H. Bischof, W. Schneider, A. Pinz, Multispectral classification of Landsat-images using neural networks, *IEEE Trans. Geosci. Remote Sens.* 30 (3) (1992) 482–490, doi:10.1109/36.142926.
- [9] L. Bruzzone, L. Carlini, A multilevel context-based system for classification of very high spatial resolution images, *IEEE Trans. Geosci. Remote Sens.* 44 (9) (2006) 2587–2600.
- [10] C.C. Chang, C.J. Lin, LIBSVM: a library for support vector machines. Software available at url <http://www.csie.ntu.edu.tw/~cjlin/libsvm>, 2001.
- [11] J. Cohen, A coefficient of agreement for nominal scales, *Educ. Psychol. Meas.* 20 (1) (1960) 37–46.
- [12] A.C. Frery, S. Ferrero, O.H. Bustos, The influence of training errors, context and number of bands in the accuracy of image classification, *Int. J. Remote Sens.* 30 (6) (2009) 1425–1440.
- [13] Z. Geem, *Music-inspired Harmony Search Algorithm: Theory and Applications*, first ed., Springer Publishing Company, Incorporated, 2009.
- [14] S. Geman, D. Geman, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Trans. Pattern Anal. Mach. Intell.* 6 (6) (1984) 721–741.
- [15] B. Ghimire, J. Rogan, J. Miller, Contextual land-cover classification: incorporating spatial dependence in land-cover classification models using random forests and the Getis statistic, *Remote Sens. Lett.* 1 (1) (2010) 45–54.
- [16] J. Guo, H. Zhou, C. Zhu, Cascaded classification of high resolution remote sensing images using multiple contexts, *Inf. Sci.* 221 (0) (2013) 84–97.
- [17] C.Y. Ji, Land-use classification of remotely sensed data using Kohonen self-organizing feature map neural networks, *Photogrammetric Eng. Remote Sens.* 66 (12) (2000) 1451–1460.
- [18] J. Kennedy, R. Eberhart, *Swarm Intelligence*, M. Kaufman, 2001.

- [19] J. Keuchel, S. Naumann, M. Heiler, A. Siegmund, Automatic land cover analysis for Tenerife by supervised classification using remotely sensed data, *Remote Sens. Environ.* 86 (4) (2003) 530–541.
- [20] J. Kittler, J. Föglein, Contextual classification of multispectral pixel data, *Image Vision Comput.* 2 (1) (1984) 13–29. [http://dx.doi.org/10.1016/0262-8856\(84\)90040-4](http://dx.doi.org/10.1016/0262-8856(84)90040-4)
- [21] J. Knorn, A. Rabe, V.V. Radeloff, T. Kuemmerle, J. Kozak, P. Hostert, Land cover mapping of large areas using chains classification of neighboring Landsat satellite images, *Remote Sens. Environ.* 113 (5) (2009) 957–964.
- [22] L.I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*, Wiley-Interscience, Hoboken, NJ, USA, 2004.
- [23] A. Laha, N. Pal, J. Das, Land cover classification using fuzzy rules and aggregation of contextual information through evidence theory, *IEEE Trans. Geosci. Remote Sens.* 6 (4) (2009) 1633–1642.
- [24] M. Li, S. Zang, B. Zhang, S. Li, C. Wu, A review of remote sensing image classification techniques: the role of spatio-contextual information, *Eur. J. Remote Sens.* 47 (389–411) (2014).
- [25] M. Mahdavi, M. Fesanghary, E. Damangir, An improved harmony search algorithm for solving optimization problems, *Appl. Math. Comput.* 188 (2) (2007) 1567–1579.
- [26] F.T. Mahmoudi, F. Samadzadegan, P. Reinartz, Object oriented image analysis based on multi-agent recognition system, *Comput. Geosci.* 54 (0) (2013) 219–230.
- [27] G. Moser, S.B. Serpico, Combining support vector machines and Markov random fields in an integrated framework for contextual image classification, *IEEE Trans. Geosci. Remote Sens.* 51 (5) (2013) 2734–2752.
- [28] R.Y.M. Nakamura, L. Fonseca, J. Santos, R. Torres, X.-S. Yang, J. Papa, Nature-inspired framework for hyperspectral band selection, *IEEE Trans. Geosci. Remote Sens.* 52 (4) (2014) 2126–2137.
- [29] R.Y.M. Nakamura, D. Osaku, A.L.M. Levada, J.P. Papa, F. Cappabianco, A.X. Falcão, OPF-MRF: optimum-path forest and Markov random fields for contextual-based image classification, in: R. Wilson, E. Hancock, A. Bors, W. Smith (Eds.), *Computer Analysis of Images and Patterns, Lecture Notes in Computer Science*, 8048, Springer, 2013, pp. 233–240.
- [30] M.G.H. Omran, M. Mahdavi, Global-best harmony search, *Appl. Math. Comput.* 198 (2) (2008) 643–656.
- [31] D. Osaku, R.Y.M. Nakamura, J.P. Papa, A.L.M. Levada, F.A.M. Cappabianco, A.X. Falcão, Optimizing contextual-based optimum-forest classification through swarm intelligence, in: *Advanced Concepts for Intelligent Vision Systems*, 2013, pp. 203–214.
- [32] J.P. Papa, A.X. Falcão, V.H.C. Albuquerque, J.M.R.S. Tavares, Efficient supervised optimum-path forest classification for large datasets, *Pattern Recognit.* 45 (1) (2012) 512–520.
- [33] J.P. Papa, A.X. Falcão, C.T.N. Suzuki, Supervised pattern classification based on optimum-path forest, *Int. J. Imag. Syst. Technol.* 19 (2) (2009) 120–131.
- [34] J.P. Papa, C.T.N. Suzuki, A.X. Falcão, LibOPF: a library for the design of optimum-path forest classifiers. Software version 2.0 available at <http://www.ic.unicamp.br/~afalcao/LibOPF>, 2009.
- [35] R.J. Pisani, R.Y.M. Nakamura, P.S. Riedel, C.R.L. Zimback, A.X. Falcão, J.P. Papa, Toward satellite-based land cover classification through optimum-path forest, *IEEE Trans. Geosci. Remote Sens.* 52 (10) (2014) 6075–6085.
- [36] R. Potts, Some generalized order–disorder transformations, *Math. Proc. Cambridge Philos. Soc.* 48 (1) (1952) 106–109.
- [37] A. Sarkar, A. Banerjee, N. Banerjee, S. Brahma, B. Kartikeyan, M. Chakraborty, K.L. Majumder, Landcover classification in MRF context using Dempster–Shafer fusion for multisensor imagery, *IEEE Trans. Image Process.* 14 (2005) 634–645.
- [38] S. Shekhar, P. Schrater, R.R. Vatsavai, W. Weili, S. Chawla, Spatial contextual classification and prediction models for mining geospatial data, *IEEE Trans. Multimedia* 4 (2) (2002) 174–188.
- [39] J. Stuckens, P. Coppin, M. Bauer, Integrating contextual information with per-pixel classification for improved land cover classification, *Remote Sens. Environ.* 71 (3) (2000) 282–296.
- [40] Y. Tarabalka, M. Fauvel, J. Chanussot, J. Benediktsson, SVM- and MRF-based method for accurate classification of hyperspectral images, *IEEE Geosci. Remote Sens. Lett.* 7 (4) (2010) 736–740.
- [41] G. Thoonen, T. Spanhove, J. Vanden Borre, P. Scheunders, Classification of heathland vegetation in a hierarchical contextual framework, *Int. J. Remote Sens.* 34 (1) (2013) 96–111.
- [42] S.W. Wharton, A contextual classification method for recognizing land use patterns in high resolution remotely sensed data, *Pattern Recognit.* 15 (4) (1982) 317–324.
- [43] F. Wu, The Potts model, *Rev. Mod. Phys.* 54 (1982) 235–268.
- [44] B. Zhang, S. Li, X. Jia, L. Gao, M. Peng, Adaptive Markov random field approach for classification of hyperspectral imagery, *IEEE Geosci. Remote Sens. Lett.* 8 (5) (2011) 973–977.