



# An Improvement on the Clustering of High-Resolution Satellite Images Using a Hybrid Algorithm

Alireza Chehreghan<sup>1</sup> · Rahim Ali Abbaspour<sup>1</sup>

Received: 1 March 2016 / Accepted: 8 September 2016 / Published online: 17 September 2016  
© Indian Society of Remote Sensing 2016

**Abstract** The challenge of obtaining training data for supervised classifications of satellite images has led researchers to unsupervised algorithms, i.e. cluster analysis. Numerous researches have been conducted to improve quality and decrease uncertainty of results of this analysis. This study proposes a hybrid cost function as well as a hybrid clustering algorithm-Artificial Bee Colony optimization approach for the clustering of high-resolution satellite images. In order to evaluate viability of the proposed methodology, it is compared to some other classic clustering algorithms such as modified K-Means, K-Medoids, Fuzzy C-Means, and Kernel-based Fuzzy C-Means methods over three different study areas selected from a WorldView-2 satellite image. The Shannon entropy technique, Kappa coefficient, compactness, and separation criteria are used as quality and uncertainty indicators for the evaluation. The results of the study show that, compared to other methods, the hybrid algorithm obtained from the proposed cost function, Kernel-based Fuzzy C-Means method, and ABC algorithm provide clustering capabilities of higher quality and lower uncertainty levels.

**Keywords** Clustering · High-resolution satellite images · Hybrid algorithms · Quality · Uncertainty

## Introduction

There are two general approaches towards dataset classification: supervised classification and unsupervised classification (Shi 2014). One of the major and cost-intensive problems dealt with supervised classification is the acquiring of training data. As such, for many cases, unsupervised classification (also referred to as clustering) represents an alternative which attempts to solve the problem to some extent (Han et al. 2012). Clustering algorithms without a priori training data are used to identify different clusters within a dataset, with some of them working with given number of the clusters, leaving the others identifying the clusters with no knowledge of the number of the clusters (Mann and Kaur 2013).

Clustering analysis, or simply clustering, refers to the process of dividing a set of data or observations into subsets (clusters) in such a way that the members in each subset share similar properties with one another, while those are significantly different from the members in any other cluster (Han et al. 2012; Shi 2014). In other words, clustering aims to allocate a set of data into different clusters in such a way that intra-cluster member similarity is maximized, while minimizing inter-cluster member similarities (Karaboga and Ozturk 2011). Satellite images represent spatial datasets which have attracted a great deal of attention from researchers. High-spatial resolution satellite images are a potential source of information for better identification of different clusters (Niazmardi et al. 2012). However, as the images show a high level of details, existing algorithms are engaged with a sort of uncertainty which influences the clustering process.

Numerous algorithms have been proposed to solve the clustering problem. Among other algorithms, K-Means is one of the most common and simple algorithms proposed

✉ Alireza Chehreghan  
achehreghan@ut.ac.ir

Rahim Ali Abbaspour  
abaspour@ut.ac.ir

<sup>1</sup> School of Surveying and Geospatial Engineering, University of Tehran, North Kargar Street, After Ale-Ahmad Junction, Tehran, Iran

by Rosenberger and Chehdi (2000). Examples of other clustering algorithms include K-Medoids, FCM,<sup>1</sup> and KFCM.<sup>2</sup> Various researches have considered these methods to improve the clustering process.

Santos and Pedrini (2016) incorporated the K-Means algorithm into entropy technique to obtain enhanced results. Among other researches where K-Means algorithm was used for clustering, one can refer to Jia et al. (2016), Rozanda et al. (2015), and Liao and Compse (2003).

Yang and Huang (2007) compared FCM algorithm against K-Means algorithm and showed that the FCM algorithm can end up with lower cost function; he further added a term into the cost function to improve the FCM algorithm. Among other researches where FCM algorithm was employed for clustering, one can refer to Kannan et al. (2011), Bilgin et al. (2008), Bidhendi et al. (2007), and Ooi and Lim (2006).

Zhang and Chen (2003b) compared FCM and KFCM algorithms to show that the KFCM algorithm outperforms the former algorithm when it comes to the cluster identification in high-spatial resolution images. Among other researches where KFCM algorithm was utilized for clustering, one can refer to Tsai et al. (2012), Das and Sil (2010), Yang and Tsai (2008), and Zhang and Chen (2003a).

Despite their partial superiority over other algorithms, each of the above-mentioned algorithms suffers from a general problem: being highly dependent on an initial guess, the algorithms may return local optima when initial cluster centers are assigned randomly (Karaboga and Ozturk 2011). In recent years, in response to this concern, a combination of these algorithms with optimization algorithms has been proposed to solve the problem (Chen et al. 2014; Esmin and Matwin 2012; Goel et al. 2011; Youssef 2011).

Zhaoxia (2011) succeeded to improve clustering results by combining so-called genetic optimization algorithm with the K-Means algorithm. Xu and Xiao (2012), Yu and JinZhi (2010), and Ke et al. (2009) are other examples of the researches wherein genetic algorithm was used to improve the results of the above-mentioned clustering algorithm.

Among the researches where artificial bee colony (ABC) algorithm was incorporated into FCM clustering, one can refer to Karaboga and Ozturk (2010) where it was shown that, when combined with ABC algorithm, FCM provides better performance than the cases where it was combined with other optimization algorithms. Sathishkumar et al. (2013) and Karaboga and Ozturk (2011) further used a

combination of ABC and FCM algorithms to enhance clustering results.

Zhao and Zhang (2011) used a combination of KFCM and ABC to cluster images. Niazmardi et al. (2012) used KFCM combined with particle swarm optimization (PSO) algorithm to cluster hyperspectral images, demonstrating superiority of the combined approach over K-Means and FCM algorithms.

Each of the cited researches in the two groups have provided a separate combination to improve the clustering; however, the present research proposes a combination of K-Means, K-Medoids, FCM and KFCM methods with ABC optimization algorithm into which a mixed cost function is introduced, and investigates the results. In the researches referred to in the two groups, clustering uncertainty and quality are not accounted for simultaneously; however, in the present research, not only the clustering uncertainty and quality are investigated, also the best hybrid algorithm for the clustering of high-spatial resolution satellite images is selected. Moreover, the proposed solution not only makes it possible to identify optimum number of clusters within the images, it rather is able to identify non-convex clusters.

In the remainder of this paper, “Hybrid Cost Function (HCF)” section introduces a Hybrid Cost Function (HCF) before resolving the optimized local value-obtaining problem through combining the proposed HCF with K-means, K-Medoids, FCM, KFCM, and ABC methods in “Hybrid Algorithms” section. “Methods for Results Assessment” section evaluates performances of common methods in clustering of satellite images of high-spatial resolution. The final section presents a comparison of the clustering uncertainty through Shannon entropy technique, and the clustering quality through indicators such as Kappa coefficient, compactness, and separation criteria the section further provides a discussion on the results.

## Hybrid Cost Function (HCF)

For most of the clustering algorithms used for raster data (e.g. satellite images), cluster calculation is based on the minimization of a cost function presented in Eq. (1) (Chunfei and Zhiyi 2013; Ke et al. 2009; Yu and JinZhi 2010).

$$CF = \sum_{i=1}^k \sum_{p \in C_i} d(z_p, m_i) \quad (1)$$

where  $k$  is the number of clusters,  $m_i$  is the center of the  $i$ th cluster ( $C_i$ ),  $z_p$ s are the image pixels, and  $d(z_p, m_i)$  is the distance between center of a cluster and assigned pixels to that cluster (Han et al. 2012). It is worth noting that, in

<sup>1</sup> Fuzzy C-Means.

<sup>2</sup> Kernel-based Fuzzy C-Means.

Eq. (1), only the distance between the center of a cluster and assigned members to that cluster is used, with the other parameters considered for improvement of the clustering.

The cost function used in this study, as shown in Eq. (2), is defined in a different form than that of common functions, i.e. the sum of three sections (Omran et al. 2002; Salman et al. 2005).

$$f(x, z) = w_1 \bar{d}_{\max}(x, z) + w_2(z_{\min} - d_{\min}(x)) + w_3 J_e \quad (2)$$

where  $w_1$ ,  $w_2$ , and  $w_3$  are the weights of each section utilized in the calculation of the total cost function. These values are optional and where the sections are of the same level significance, the same weight can be considered for all three components. The weights  $w_1$ ,  $w_2$ , and  $w_3$  have non-negative values, with the sum of them being equal to 1.  $x$  is the vector of the cluster centers, i.e.  $x = (m_1, \dots, m_j, \dots, m_{N_c})$ , in which  $m_j$  is the center of  $j$ th cluster and  $N_c$  is the number of clusters.  $z$  is a matrix containing components of each cluster. In the first section of Eq. (2), mean distance from the center of each cluster to its components is calculated (Eq. 4), with the longest mean distance taken as  $\bar{d}_{\max}$ . Equation (3) shows how this value is calculated.

$$\bar{d}_{\max}(x, z) = \max_{j=1, \dots, N_c} \left\{ \sum_{z_p \in C_j} d(z_p, m_j) / |C_j| \right\} \quad (3)$$

where  $|C_j|$  is the cardinality of the  $j$ th cluster. The distance from each cluster center to its components can be calculated through Eq. (4).

$$d(z_p, m_j) = \sqrt{\sum_{k=1}^{N_b} (z_{pk} - m_{jk})^2} \quad (4)$$

where  $N_b$  is the number of image bands.

It is clear that the value of the first section in Eq. (2) should be minimized to increase the clustering quality (Omran et al. 2002). In the second part of Eq. (2), the distances between each pair of cluster centers (i.e.  $d(m_i, m_j)$ ) are calculated, so as to find the smallest such distance (Eq. 5).

$$d_{\min}(x) = \min_{\forall i, j, i \neq j} \{d(m_i, m_j)\} \quad (5)$$

To identify the main clusters, the distance between cluster centers should be maximized; however, this value is subtracted from the largest distance existing within the dataset, as is expressed in Eq. (6). Consequently, the second section of Eq. (2) should also be minimized.

$$z_{\max} = \sqrt{N_b} d_m \quad (6)$$

where  $d_m$  is the maximum distance within the dataset.

In the third part of Eq. (2), a general measure of clustering quality is calculated through assessing average maximum distance in each cluster, according to Eq. (7).

$$J_e = \frac{\sum_{j=1}^{N_c} \left[ \sum_{z_p \in C_j} d(z_p, m_j) \right] / |C_j|}{N_c} \quad (7)$$

In the Eq. (1), the distance between each cluster center and the members in the same cluster is considered only. This is while, in order to improve the clustering performance, Eq. (2) comes with three terms which not only investigate cluster center-to-member distance, but also account for the distance between centers of different clusters and average of the maximum distance in each cluster.

### Hybrid Algorithms

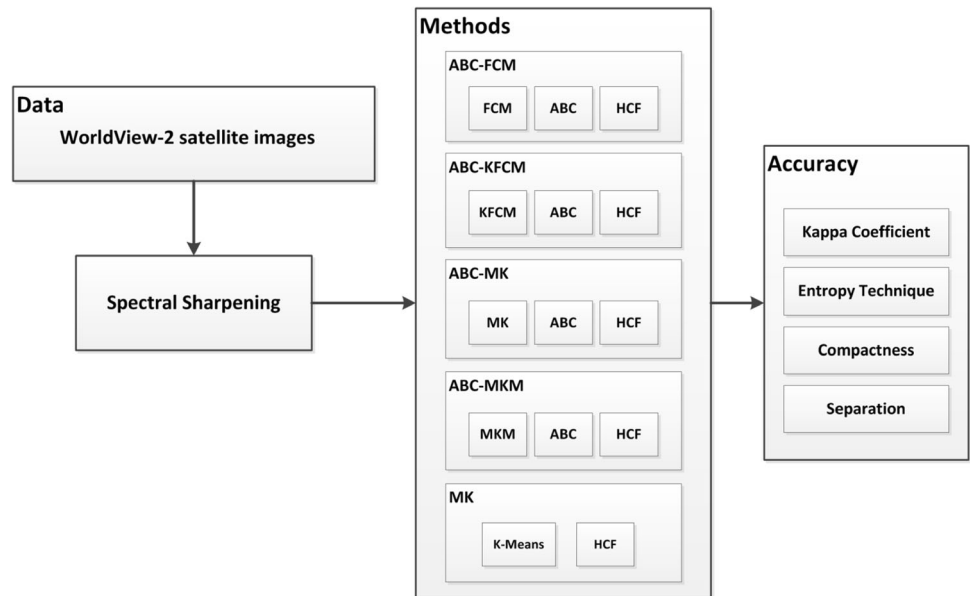
Because K-means (K), K-Medoids (KM), Fuzzy C-Means (FCM) and Kernel-based Fuzzy C-Means (KFCM) algorithms depend largely on the starting point of the algorithms (primarily random values for the cluster centers), they may yield local optimum values (Karaboga and Ozturk 2011). In order to address this critical issue, a combination of clustering methods with optimization algorithms has been generally recommended in recent years. In the present study, in order to achieve clustering algorithms of higher performance, the three components (i.e. HCF, clustering algorithms, and ABC) were combined to end up with a clustering approach of higher quality and lower uncertainty for satellite images of high-spatial resolution.

In fact, in the present study, combining the most popular clustering methods with the ABC optimization algorithm and the mixed cost function, efficiency of each of the combinations will be investigated from three perspectives. Clustering results are evaluated via Kappa coefficient, compactness–separation criteria and entropy technique, so as to select the most efficient method for clustering high-spatial resolution satellite images. Figure 1 shows a layout of the process.

### ABC-FCM Hybrid Algorithm

Being first introduced by Dunn (1973) and then improved by Bezdek (1980), FCM clustering algorithm is one of the most widely used algorithms in the field of clustering (Xiaojun et al. 2012). In contrast to K-Means method in which each pixel belongs to only one cluster, this algorithm assigns a set of membership degrees to each pixel indicating membership values of each pixel to different clusters. The FCM algorithm is an iterative clustering process that stops when the number of generated clusters reach  $N_c$ .

**Fig. 1** General process of the application of hybrid algorithms and their evaluation



These clusters are obtained by minimizing Eq. (8) (Ouadfel et al. 2012).

$$J = \sum_{k=1}^n \sum_{i=1}^{N_c} u_{ik}^f d^2(z_k, m_i) \tag{8}$$

where  $f$  is the fuzzification degree, which is usually considered as 2 (Niazmardi et al. 2012; Pour and Homayouni 2016; Xiaojun et al. 2012),  $u_{ik}$  is the degree to which pixel  $k$  belongs to the  $i$ th cluster [obtained through Eq. (9)], and  $m_i$  is the center of  $i$ th cluster which is calculated by Eq. (10).

$$u_{ik} = \frac{1}{\sum_{j=1}^{N_c} \left(\frac{d_{jk}}{d_{ik}}\right)^{2f} - 1} \tag{9}$$

$$m_i = \frac{\sum_{k=1}^n u_{ik}^f z_k}{\sum_{k=1}^n u_{ik}^f} \tag{10}$$

As this algorithm depends heavily on initial randomly-selected cluster centers, it may yield local optimized values (Karaboga and Ozturk 2011). To address this issue, the algorithm should be combined with meta-heuristics such as ABC algorithm.

The ABC algorithm was first introduced by Karaboga (2005) for optimization purposes. This algorithm uses the explorative behavior of bees during the search for food (Karaboga 2005). Karaboga and Ozturk (2011) showed that this algorithm outperforms other optimization methods when it comes to clustering; hence, this algorithm is selected among the existing meta-heuristics for optimization.

There are three types of bees in ABC algorithm: scout bees, follower bees, and employee bees. Although all bees

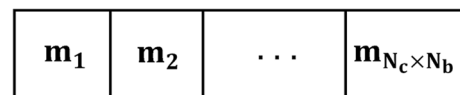
act as employee bees in the first iteration, the bee with a smaller fitness function value is considered as the scout bee, with the other bees assessed accordingly (Karaboga and Ozturk 2011). In the proposed method, the fitness function is obtained through Eq. (2) and the objective is to minimize this value. Once the scout bee is recognized, corresponding values are assigned to the follower and employee bees through Eq. (11). The values are different for these two types of bees (Karaboga and Ozturk 2011).

$$z_i = Z_i + \varphi_i(v_{i1} - v_{i2}) \tag{11}$$

where  $z_i$  is the new location of bee,  $Z_i$  is the location of scout bee,  $\varphi_i$  is the random function returning a value ranging from 1 to  $-1$ ,  $v_{i1}$  and  $v_{i2}$  are the lowest and highest limits of the search range, respectively. Location of employee bee can be obtained through  $v_{i1} = z_{max}$  and  $v_{i2} = z_{min}$ . Figure 2 shows a bee structure in the proposed method. According to Fig. 2, each bee takes, as unknown, cluster centers within each band for which it calculates an optimum value via an iterative approach. As such, each bee hosts  $N_c \times N_b$  unknown parameters.

**ABC-KFCM Hybrid Algorithm**

Developed based on the concept of FCM algorithm (Zhang and Chen 2003b), KFCM algorithm projects the pixels in kernel space before looking for similarities between points.



**Fig. 2** Structure of bees in ABC-FCM

Equation (12) shows this algorithm (Niazmardi et al. 2012).

$$J = \sum_{i=1}^{N_c} \sum_{k=1}^n u_{ik}^f \|\varphi(z_k) - \varphi(m_i)\|^2 \tag{12}$$

The distance in Eq. (12) is calculated using Eq. (13) (Niazmardi et al. 2012).

$$\|\varphi(z_k) - \varphi(m_i)\|^2 = K(z_k, z_k) + K(m_i, m_i) - 2K(z_k, m_i) \tag{13}$$

The Gaussian Kernel Function,  $K$ , is obtained through Eq. (14) (Niazmardi et al. 2012).

$$K(z_k, m_i) = \exp(-\|z_k - m_i\|^2 / \sigma^2) \tag{14}$$

According to Eq. (14),  $K(z_k, z_k) = K(m_i, m_i) = 1$ ; accordingly, Eq. (12) is changed into Eq. (15).

$$J = 2 \sum_{i=1}^{N_c} \sum_{k=1}^n u_{ik}^f (1 - K(z_k, m_i)) \tag{15}$$

In FCM method, cluster centers are the only unknown parameters, while KFCM algorithm has additional unknown parameters including the width of Gaussian Function ( $\sigma$ ).

The combination of ABC algorithm with KFCM method is identical to that described for FCM in “ABC-FCM Hybrid Algorithm” section, but bee structures are different. Figure 3 shows the bee structures in ABC-KFCM algorithm. According to Fig. 3, similar to Fig. 2, in this algorithm, each bee takes, as unknown, cluster centers within each band. However, in KFCM algorithm,  $\sigma$  parameter is also unknown for each band. As such, each bee hosts  $N_c \times N_b$  unknown cluster centers (each cluster in each band represents one unknown) along with  $N_b$  unknown  $\sigma$  parameters.

### Methods for Results Assessment

One of the important issues in clustering is results assessment. This examination helps identifying most appropriate groups for the data under study (Halkidi et al. 2001). While there are different methods for assessing clustering algorithms results in the literature, this paper utilizes three indices, namely Kappa coefficient, entropy technique, and compactness and separation criteria.

### Kappa Coefficient

Kappa coefficient is one of the most important statistical indices that can be extracted from error matrix. In fact, this index removes the effects of chance from modeling. Equation (16) shows how Kappa coefficient is calculated (Mather and Tso 2009).

$$\hat{k} = \frac{n \sum_{i=1}^k n_{ii} - \sum_{i=1}^k n_{i+} n_{+i}}{n^2 - \sum_{i=1}^k n_{i+} n_{+i}} \tag{16}$$

where  $n$  is the total number of actual earth pixels, and  $n_{i+}$  and  $n_{+i}$  show all of the components on  $i$ th row and  $i$ th column, respectively. Kappa coefficient is a value between 0 and 1, with the former indicating a completely random clustering, while the latter refers to an ideal clustering. Kappa coefficient is, in fact, a pessimistic estimation of the modeling and yields an accuracy lower than the actual value (Mather and Tso 2009).

### Entropy Technique

Entropy is a general concept in physics, social sciences, and information theory. It shows the untrustworthiness level of an expected content in a message. In other words, entropy in information theory is a criterion stated for the untrustworthiness level by a discrete probability distribution, such that this untrustworthiness will be more if the distribution is scattered rather than peaked. Equation (17) shows how this parameter is calculated (Asagharpour 2010).

$$E = - \sum_{i=1}^n [P_i \times \ln(P_i)], \quad P_i = \frac{d_i}{\sum_{i=1}^n d_i} \tag{17}$$

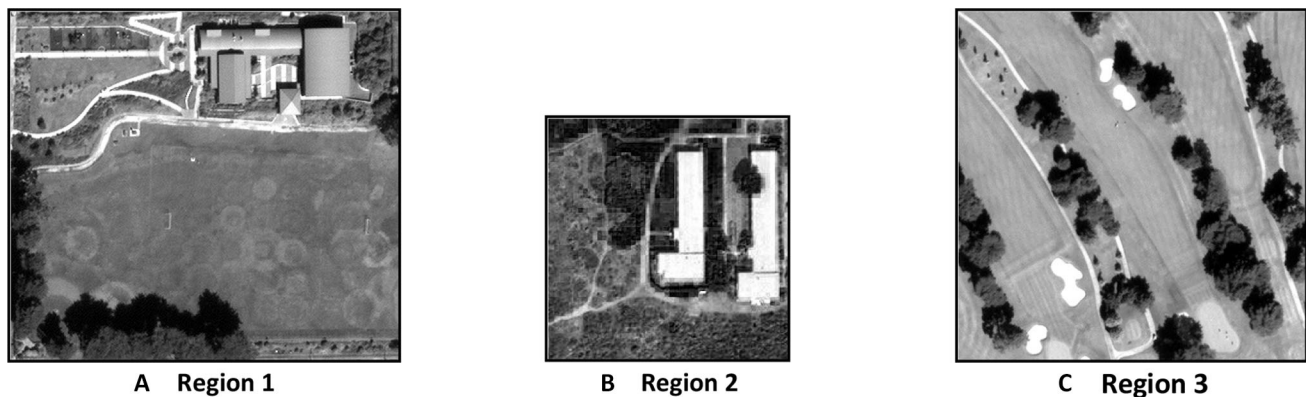
where  $d$  is the distance from each cluster center to its components.

### Compactness and Separation Criteria

Berry and Linoff (1996) introduced two criteria called compactness and separation criteria. Expressed in terms of standard deviation, compactness criterion shows how close are values of members in each cluster to one another; this value should be minimized (Halkidi et al. 2001). Separation shows an appropriate difference in clusters for which three approaches have been introduced by Berry and Linoff (1996).

Fig. 3 Bee structures in ABC-KFCM

$\sigma_1$	$\sigma_2$	...	$\sigma_{N_b}$	$m_1$	$m_2$	...	$m_{N_c \times N_b}$
------------	------------	-----	----------------	-------	-------	-----	----------------------



**Fig. 4** The study areas selected from the WorldView2 satellite images

1. Measuring the distance between the closest two components in clusters.<sup>3</sup>
2. Measuring the distance between the farthest two components in clusters.<sup>4</sup>
3. Measuring the distance between cluster centers.<sup>5</sup>

### Implementation and Evaluation

To evaluate the efficiency of the proposed methods, it was implemented and tested for the clustering of WorldView-2 satellite images. Captured on October 9, 2011, the test images were from Lawrence, Kansas, USA. For these images, a spatial resolution of 0.5 m was obtained following the spectral sharpening. All of the eight spectral bands (4 standard colors: red, blue, green, near-IR; 4 new colors: red edge, coastal, yellow, near-IR2) were used for resolving the clustering issue. Figure 4 shows the three case study areas selected for cluster analysis using the proposed method. Different sizes of images were selected to make sure image size did not affect assessment results. Region 1 is of  $750 \times 833$  pixel dimension, Region 2 is of  $484 \times 478$  pixel dimension, and Region 3 covers an area of  $734 \times 697$  pixel dimension.

To evaluate the proposed combinations, five algorithms were used for the image clustering: K-means, Modified K-means (called MK) (using hybrid cost function), combined Modified K-means-ABC (called ABC-MK), combined Modified K-Medoids-ABC algorithms (called ABC-MKM), combined ABC-FCM algorithms (called ABC-FCM), and combined ABC-KFCM algorithms (called ABC-KFCM). K-Means and MK were used with the hybrid cost function, so as to explain the effect of the proposed

cost function on clustering, and compare the results. Moreover, the proposed cost function was also used for other methods. The ABC algorithm was combined with each of these methods to obtain the most optimized clustering method for high-resolution satellite images. Figure 5 shows the clustering of the areas under study using different algorithms.

Figure 6 value of the hybrid cost function value in different iterations with the utilized algorithms.

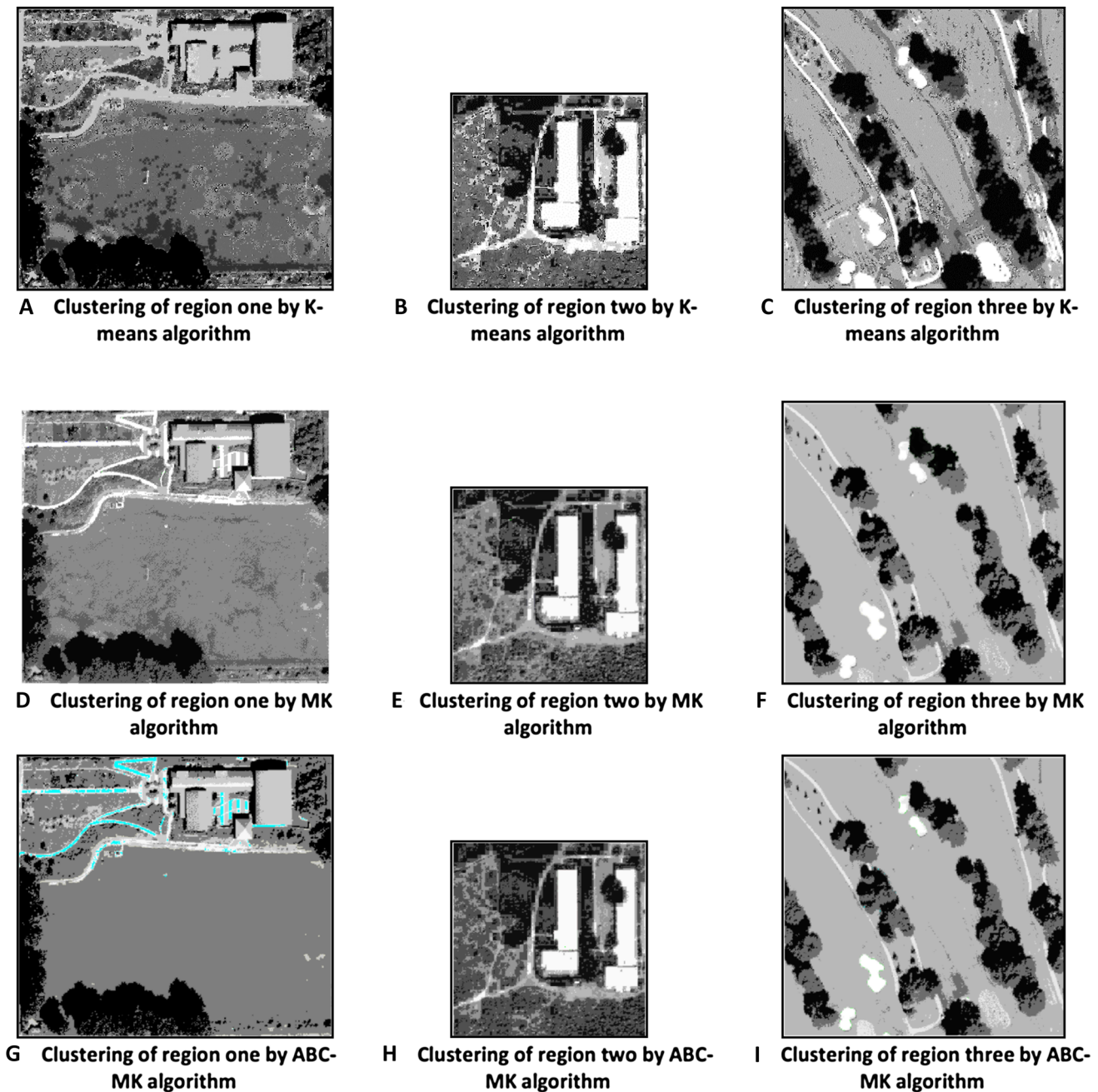
Table 1 reports the assessment results of algorithms for the regions understudy. The values obtained from K-Means and modified K-Means show that the hybrid cost function succeeded to optimize the results. The entropy value in Table 1 indicates enhanced quality-trustworthiness and reduced uncertainty of the associated results with modified K-Means algorithm, as compared to those of the K-Means algorithm. The associated uncertainty values with modified K-Means algorithm in regions 1, 2, and 3 decreased from 51.2322, 49.7665, and 49.7651 to 41.6859, 33.9081, and 41.7540, respectively.

Standard deviation served as a measure of compactness for the clustering assessment. The obtained standard deviations for the regions understudy show that, the hybrid cost function contributed to decreased standard deviation of the distances from each cluster member to the cluster center. This indicates the closeness of cluster members to one another. According to Table 1, corresponding standard deviations to K-Means algorithm were 50.8733, 47.0922, and 43.6543 for regions 1, 2, and 3, respectively. In modified K-Means algorithm, these values were observed to be significantly decreased to 23.1087, 34.3867, and 33.5498, respectively. To estimate separation level of the clusters with different algorithms, cluster distances were assessed. By increasing the inter-cluster space, the hybrid cost function attempts to assign similar pixels to the same cluster while maximizing inter-cluster differences. The cluster distances in K-Means algorithm were 432.5019, 239.6501, and 467.9342 for regions 1, 2, and 3,

<sup>3</sup> Single Linkage (SL).

<sup>4</sup> Complete Linkage (CL).

<sup>5</sup> Complete Linkage (CC).



**Fig. 5** Clustering of under study regions using different algorithms

respectively. In the modified K-Means algorithm, however, the distances increased to 492.5623, 245.5923, and 481.0512, respectively.

Moreover, Table 1 shows that, the combinations of ABC algorithm with other algorithms contributed into optimized results. This can be explained by the fact that, the ABC algorithm prevents the clustering algorithms from being trapped in local optima. Considering the values reported in Table 1, compared to modified K-Means, the ABC-MK algorithm decreased the corresponding entropy

to region 1 by 5.3057. It further decreased the compactness and increased the cluster distance by 2.1359 and 31.3458, respectively. Optimizations experienced in other regions are also reported in Table 1.

According to Table 1, the corresponding values of the hybrid cost function to regions 1, 2, and 3 underwent a significant decrease by 77.33, 45.0008, and 30.3791, respectively, in modified K-Means algorithm. The decrease was due to the use of ABC algorithm. Furthermore, Table 1 shows that, one can sort the algorithms in order of

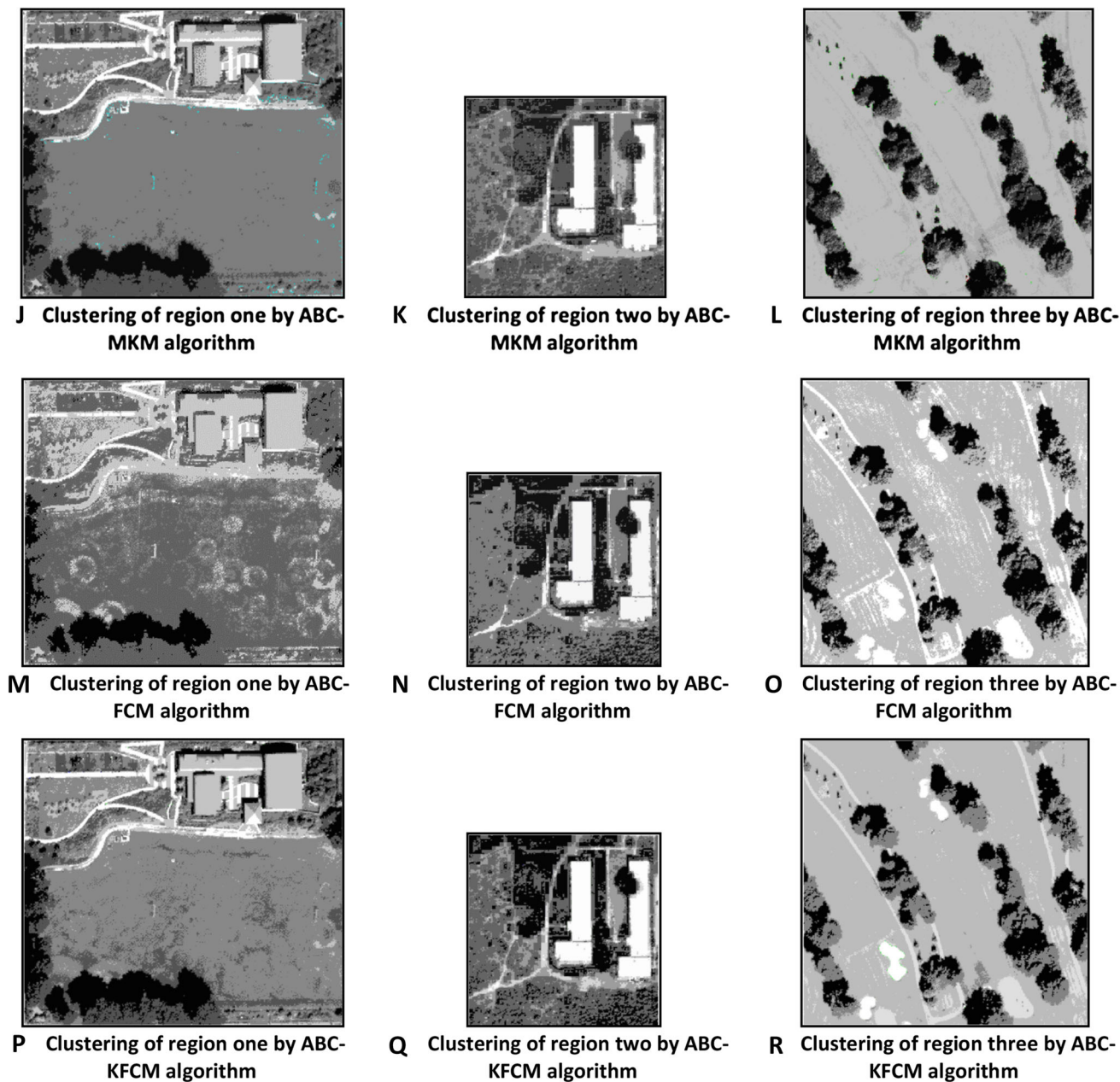


Fig. 5 continued

decreasing uncertainty, compactness, cost function, and increasing separation, as ABC-KFCM, ABC-MK, ABC-FCM, ABC-MKM, modified K-Means, and K-Means. The results indicated that the hybrid algorithm consisting of the proposed cost function, KFCM algorithm, and ABC algorithm was the most optimized method. Therefore, this hybrid algorithm can be recommended for the clustering of high-resolution satellite images.

The only issue remained is to assess the clustering method against the reality; the mission was accomplished using Kappa coefficient which assessed clustering results of satellite

images of high-spatial resolution against the reality. Figure 7 shows the values obtained by Kappa coefficient for the provided algorithms in each study area. Once hybrid cost function was implemented, values of Kappa coefficient increased from 0.636, 0.718 and 0.628 to 0.753, 0.814 and 0.715, respectively. Following the incorporation of ABC algorithm into modified K-Means algorithm, Kappa coefficient exhibited 0.118, 0.080, and 0.142 enhancements for regions 1, 2, and 3, respectively. Considering the obtained significant values (reported in Table 1), ABC-KFCM algorithm yielded Kappa coefficients at 0.942, 0.969, and 0.947 confidence level for regions 1, 2,



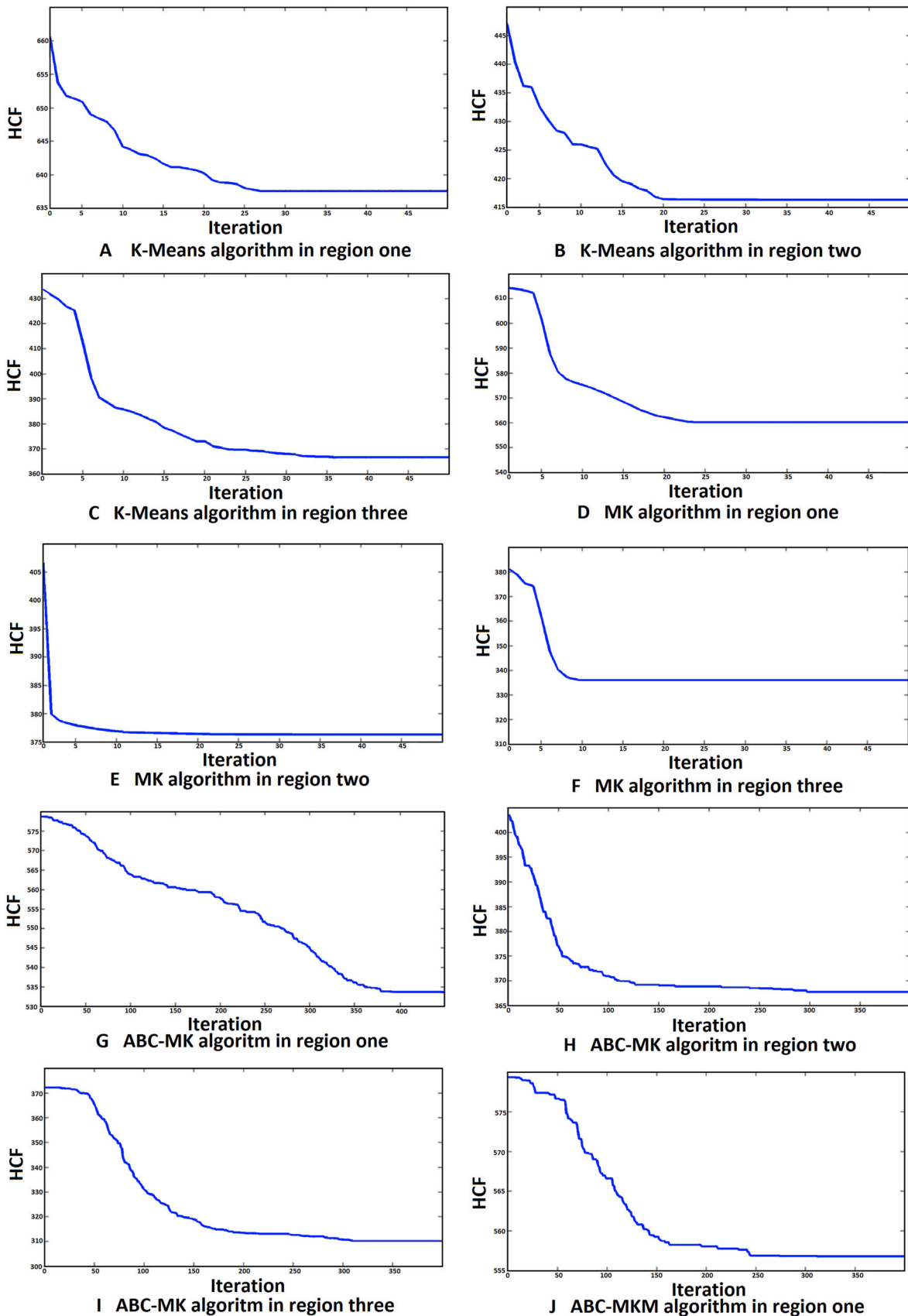


Fig. 6 Hybrid cost function value in different iterations and for utilized algorithms

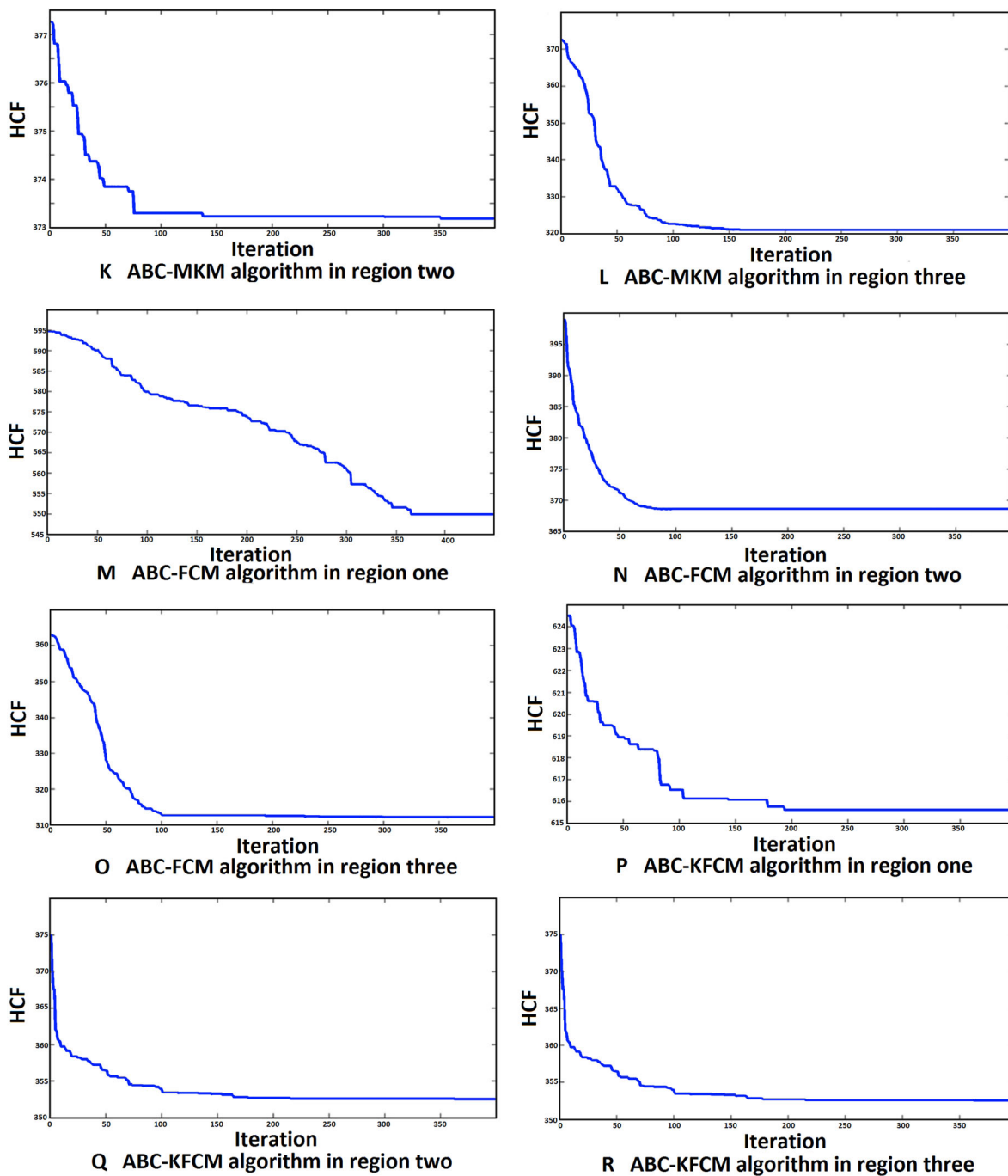


Fig. 6 continued

and 3, respectively. These values were 0.306, 0.251, and 0.319 higher than those yielded by K-Means algorithm.

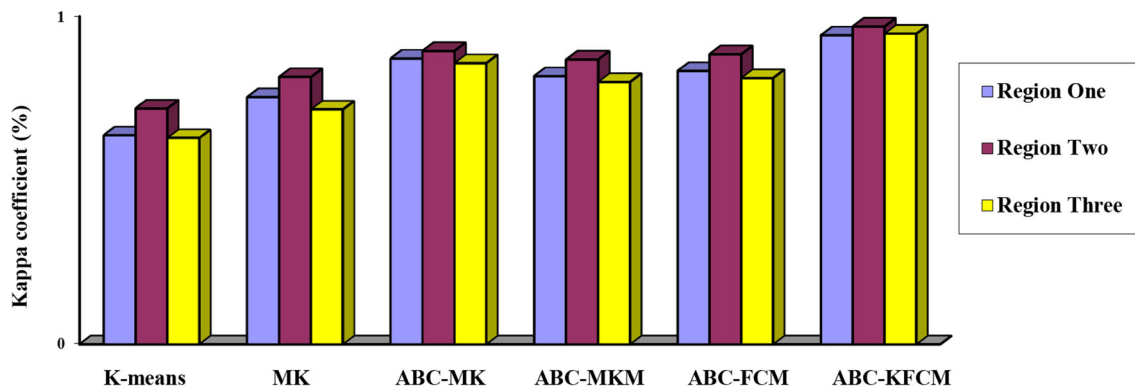
## Conclusion

This study showed the effect of hybrid cost function and intelligent hybrid methods on uncertainty and quality of cluster analysis for high-spatial resolution satellite images.

The optimization level of clustering quality and increased trustworthiness of each cluster value were initially assessed by introducing a hybrid cost function. Clustering quality was assessed by Kappa coefficient, compactness/separation levels, and cost function values. Moreover, uncertainties were assessed through entropy measurements. These assessments were conducted on the three study areas across a high-spatial resolution satellite image. The results showed that, in all three regions, the uncertainty level

**Table 1** Results of the proposed algorithms assessment

Under study regions	Assessment method	K-Means	MK	ABC-MK	ABC-MKM	ABC-FCM	ABC-KFCM
Region one	Entropy technique	51.2322	41.6859	36.3802	38.9149	37.2232	33.3393
	Compactness	50.8733	23.1087	20.9692	21.5658	21.2228	20.6443
	CC	432.5019	492.5623	523.9081	502.8710	513.3912	540.8172
	HCF	637.5600	560.2300	533.5143	557.1207	549.8820	516.3122
Region two	Entropy technique	49.7665	33.9081	28.9756	32.3348	30.2695	26.6447
	Compactness	47.0922	34.3867	28.8939	32.1312	30.1358	28.3625
	CC	239.6501	245.5923	251.2493	249.5121	250.9081	256.7150
	HCF	421.2958	376.2950	367.4229	373.1778	368.3120	352.6061
Region three	Entropy technique	49.7651	41.7540	37.3049	38.4636	38.4829	37.0906
	Compactness	43.6543	33.5498	15.7420	31.4790	22.7265	15.1538
	CC	467.9342	481.0512	507.6829	500.0058	501.9275	517.9715
	HCF	366.566	336.1869	310.1425	321.2165	312.1470	284.6500



**Fig. 7** Values of Kappa coefficient for the provided algorithms

decreased according to the entropy values. Furthermore, clustering compactness and cost function decreased, while the separation of clusters and Kappa coefficient increased; i.e., every indicator showed the enhancement in the clustering quality.

Uncertainty and clustering quality of high-spatial resolution satellite images were then assessed by combining ABC, hybrid cost function, and clustering methods (i.e. K-Means, K-Medoids, FCM, and KFCM). According to the quality indicators, the results showed that the combination of the ABC algorithm, the hybrid cost function, and KFCM leads to even further enhancements in clustering certainty and quality of high-spatial resolution satellite images. Therefore, the proposed hybrid algorithm was proved to serve as the most optimized method, compared to other methods considered, for the clustering of high-spatial resolution satellite images. It is recommend to address the problem with other optimization methods, so as to compare the methods and find an optimal approach.

**References**

Asagharpour, M. J. (2010). *Multi-criteria decision making*. Tehran: University of Tehran.

Berry, M. J. A., & Linoff, G. (1996). *Data mining techniques for marketing, sales and customer support*. Hoboken: Wiley.

Bezdek, J. C. (1980). A convergence theorem for the fuzzy ISODATA clustering algorithms. In *IEEE transactions on PAMI-2 pattern analysis and machine intelligence* (pp. 1–8).

Bidhendi, S. K., Shirazi, A. S., Fotoohi, N., & Ebadzadeh, M. M. (2007). Material classification of hyperspectral images using unsupervised fuzzy clustering methods. In *Third international IEEE conference on signal-image technologies and internet-based system, 2007, SITIS'07* (pp. 619–623). IEEE.

Bilgin, G., Erturk, S., & Yildirim, T. (2008). Unsupervised classification of hyperspectral-image data using fuzzy approaches that spatially exploit membership relations. *IEEE Geoscience and Remote Sensing Letters*, 5, 673–677.

Chen, X., Zhou, Y., & Luo, Q. (2014). A hybrid monkey search algorithm for clustering analysis. *The Scientific World Journal*, 2014, 938239.

Chunfei, Z., & Zhiyi, F. (2013). An improved k-means clustering algorithm. *Journal of Information and Computational Science*, 10, 193–199.

- Das, S., & Sil, S. (2010). Kernel-induced fuzzy clustering of image pixels with an improved differential evolution algorithm. *Information Sciences*, *180*, 1237–1256.
- Dunn, J. C. (1973). A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, *3*, 32–57.
- Esmim, A. A., & Matwin, S. (2012). Data clustering using hybrid particle swarm optimization. In H. Yin, J. F. Costa, & G. Barreto (Eds.), *Intelligent data engineering and automated learning—IDEAL 2012* (pp. 159–166). Berlin: Springer.
- Goel, S., Sharma, A., & Goel, A. (2011). Development of swarm based hybrid algorithm for identification of natural terrain features. In *2011 International conference on computational intelligence and communication networks (CICN)* (pp. 293–296).
- Halkidi, M., Batistakis, Y., & Vazirgiannis, M. (2001). On clustering validation techniques. *Journal of Intelligent Information Systems*, *17*, 107–145.
- Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques* (3rd ed.). Amsterdam: Elsevier.
- Jia, L., Li, M., Zhang, P., Wu, Y., & Zhu, H. (2016). SAR image change detection based on multiple kernel K-means clustering with local-neighborhood information. *IEEE Geoscience and Remote Sensing Letters*, *13*, 856–860.
- Kannan, S. R., Ramathilagam, S., & Pandiyarajan, R. (2011). Modified bias field fuzzy C-means for effective segmentation of brain MRI. In M. Gavrilova & C. J. K. Tan (Eds.), *Transactions on computational science VIII* (pp. 127–145). Berlin: Springer.
- Karaboga, D. (2005). *An idea based on honey bee swarm for numerical optimization*. Technical report-tr06, Erciyes university.
- Karaboga, D., & Ozturk, C. (2010). Fuzzy clustering with artificial bee colony algorithm. *Scientific Research and Essays*, *5*, 1889.
- Karaboga, D., & Ozturk, C. (2011). A novel clustering approach: Artificial Bee Colony (ABC) algorithm. *Applied Soft Computing*, *11*, 652–657.
- Ke, S., Jie, L., & Xueying, W. (2009). K mean cluster algorithm with refined initial center point. *Journal of Shenyang Normal University*, *27*, 448–451.
- Liao, T., & CompSc, B. (2003). Image segmentation-hybrid method combining clustering and region merging. BCompSc Thesis, Monash University.
- Mann, A. K., & Kaur, N. (2013). Review paper on clustering techniques. *Global Journal of Computer Science and Technology*, *13*(5), 42–47.
- Mather, P., & Tso, B. (2009). *Classification methods for remotely sensed data*. Boca Raton: Taylor & Francis Group.
- Niazmardi, S., Naeini, A. A., Homayouni, S., Safari, A., & Samadzadegan, F. (2012). Particle swarm optimization of kernel-based fuzzy c-means for hyperspectral data clustering. *Journal of Applied Remote Sensing*, *66*(1), 063601–15.
- Omran, M., Salman, A., & Engelbrecht, A. P. (2002). Image classification using particle swarm optimization. In *4th Asia-Pacific conference simulated evolution and learning, Singapore*.
- Ooi, W. S., & Lim, C. P. (2006). Fuzzy clustering of color and texture features for image segmentation: A study on satellite image retrieval. *Journal of intelligent and fuzzy systems*, *17*, 15.
- Ouadfel, S., Abdelmalik, T.-A., & Batouche, M. (2012). Spatial information based image clustering with a swarm approach. *IAES International Journal of Artificial Intelligence*, *1*, 149–160.
- Pour, H. E., & Homayouni, S. (2016). Clustering of hyperspectral image using fuzzy C-means based on spectral similarity measures. *Computations and Materials in Civil Engineering*, *1*, 49–56.
- Rosenberger, C., & Chehdi, K. (2000). Unsupervised clustering method with optimal estimation of the number of clusters: application to image segmentation. In *Pattern recognition, 2000. Proceedings. 15th international conference on 659 edn, Barcelona*.
- Rozanda, N. E., Ismail, M., & Permana, I. (2015). Segmentation google earth imagery using K-means clustering and normalized RGB color space. In *Computational intelligence in data mining. Proceedings of the International Conference on CIDM, 20-21 December 2014, (Vol. 1, pp. 375–386)*. India: Springer.
- Salman, A., Omran, M., & Engelbrecht, A. P. (2005). Particle swarm optimization method for image clustering. *International Journal of Pattern Recognition and Artificial Intelligence*, *19*, 297–321.
- Santos, A., & Pedrini, H. (2016). A combination of k-means clustering and entropy filtering for band selection and classification in hyperspectral images. *International Journal of Remote Sensing*, *37*, 3005–3020.
- Sathishkumar, K., Balamurugan, E., & Narendran, P. (2013). An efficient artificial bee colony and fuzzy C means based co-regulated biclustering from gene expression data. In R. Prasath & T. Kathirvalavakumar (Eds.), *Mining intelligence and knowledge exploration* (pp. 120–129). Berlin: Springer.
- Shi, G. (2014). *Data mining and knowledge discovery for geoscientists* (1st ed.). Amsterdam: Elsevier.
- Tsai, H.-S., Hung, W.-L., & Yang, M.-S. (2012). A robust kernel-based fuzzy C-means algorithm by incorporating suppressed and magnified membership for MRI image segmentation. In J. Lei, F. Wang, H. Deng, & D. Miao (Eds.), *Artificial intelligence and computational intelligence* (pp. 744–754). Berlin: Springer.
- Xiaojun, L., Junying, L., & Haitao, L. (2012). Improved fuzzy C-means clustering algorithm based on cluster density. *Journal of Computational Information Systems*, *8*, 727–737.
- Xu, X.-M., & Xiao, Y.-H. (2012). KBAC: K-means based adaptive clustering for massive dataset. *Journal of Chinese Computer Systems*, *33*, 2268–2272.
- Yang, Y., & Huang, S. (2007). Image segmentation by fuzzy C-means clustering algorithm with a novel penalty term. *Computers and Artificial Intelligence*, *26*, 17–31.
- Yang, M.-S., & Tsai, H.-S. (2008). A Gaussian kernel-based fuzzy c-means algorithm with a spatial bias correction. *Pattern Recognition Letters*, *29*, 1713–1725.
- Youssef, S. M. (2011). A new hybrid evolutionary-based data clustering using fuzzy particle swarm optimization. In *2011 23rd IEEE international conference on tools with artificial intelligence (ICTAI)* (pp. 717–724).
- Yu, H., & JinZhi, B. (2010). Optimized K-means clustering analysis based on genetic algorithm. *Computer System Application*, *19*, 52–55.
- Zhang, D.-Q., & Chen, S.-C. (2003a). Clustering incomplete data using kernel-based fuzzy c-means algorithm. *Neural Processing Letters*, *18*, 155–162.
- Zhang, D. Q., & Chen, S. C. (2003b). Kernel-based fuzzy and possibilistic c-means clustering. In *International conference on artificial neural networks* (pp. 122–125). Springer, Istanbul.
- Zhao, X., & Zhang, S. (2011). An improved KFCM algorithm based on artificial bee colony. In H. Deng, D. Miao, F. Wang, & J. Lei (Eds.), *Emerging research in artificial intelligence and computational intelligence* (pp. 190–198). Berlin: Springer.
- Zhaoxia, T. (2011). K-means clustering algorithm based on improved genetic algorithm. *Journal of Chengdu University (Natural Science Edition)*, *30*, 162–164.