

Continuous Iterative Guided Spectral Class Rejection Classification Algorithm: Part 2

Rhonda D. Phillips *Student Member, IEEE*, Layne T. Watson *Fellow, IEEE*,
Randolph H. Wynne *Member, IEEE*, and Naren Ramakrishnan *Member, IEEE*

Abstract—This paper describes in detail the continuous iterative guided spectral class rejection (CIGSCR) classification method based on the iterative guided spectral class rejection (IGSCR) classification method for remotely sensed data. Both CIGSCR and IGSCR use semisupervised clustering to locate clusters that are associated with classes in a classification scheme. In CIGSCR and IGSCR, training data are used to evaluate the strength of the association between a particular cluster and a class, and a statistical hypothesis test is used to determine which clusters should be associated with a class and used for classification and which clusters should be rejected and possibly refined. Experimental results indicate that the soft classification output by CIGSCR is reasonably accurate (when compared to IGSCR), and the fundamental algorithmic changes in CIGSCR (from IGSCR) result in CIGSCR being less sensitive to input parameters that influence iterations. Furthermore, evidence is presented that the semisupervised clustering in CIGSCR produces more accurate classifications than classification based on clustering without supervision.

I. INTRODUCTION

The conversion of the iterative guided spectral class rejection (IGSCR) classification method ([1], [2], [3]) from a hard classification to a soft classification method called continuous iterative guided spectral class rejection (CIGSCR) based on soft clustering will require the soft cluster evaluation and refinement methods developed in Part 1. IGSCR evaluates hard clusters using a statistical hypothesis test based on a binomial random variable. This discrete random variable can be used to model hard cluster and class memberships. CIGSCR uses the hypothesis test developed in Part 1 that is based on both discrete random variables modeling class memberships of training data and continuous random variables modeling soft cluster memberships. The iterative cluster refinement in IGSCR assumes samples are attributed to only one cluster (hard clustering), but the soft cluster refinement proposed in Part 1 uses soft memberships to seed new clusters, taking advantage of soft clustering and providing an alternative mechanism for cluster refinement.

This paper describes how the soft cluster evaluation and refinement are incorporated into the IGSCR framework to form CIGSCR, and provides experimental results demonstrating that CIGSCR can produce superior classifications to IGSCR, especially in circumstances that are ideally suited for soft clustering and classification. Section II provides a detailed description of the CIGSCR algorithm including pseudocode, and Section III discusses how alternative clustering distance (dissimilarity) functions may be used within CIGSCR. Section IV presents experimental results and detailed discussion, and Section V concludes the paper.

II. CIGSCR ALGORITHM

CIGSCR, like IGSCR, begins by clustering an image, but unlike IGSCR, CIGSCR uses soft clusters where each sample has partial membership in each cluster. Each soft cluster is then evaluated using the association significance test based on one of two standard normal random variables defined in Part 1,

$$\hat{z} = \frac{\sqrt{n_c}(\bar{w}_{c,j} - \bar{w}_j)}{S_{\bar{w}_j}}, \quad (1)$$

R.D. Phillip, L.T. Watson, and N. Ramakrishnan are with the Departments of Computer Science and Mathematics, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061.

R.H. Wynne is with the Department of Forestry, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061.

where n_c is the number of samples labeled with the c th class ($1 \leq c \leq C$), $\bar{w}_{c,j}$ is the average weight of samples labeled with the c th class for the j th cluster ($1 \leq j \leq K$), \bar{w}_j is the sample mean of all weights in the j th cluster, and $S_{\bar{w}_j}$ is the sample standard deviation of the weights for the j th cluster, and

$$\hat{z} = \frac{y_{c,j} - n_c \bar{w}_j}{\sqrt{p_c \sum_{d=1}^C n_d (S_{\bar{w}_{d,j}}^2 + (1 - p_c) \bar{w}_{d,j}^2)}}, \quad (2)$$

where $y_{c,j}$ is the sum of weights of samples labeled with the c th class for the j th cluster, p_c is the estimated probability of the c th class, and $S_{\bar{w}_{d,j}}$ is the sample standard deviation of weights of samples labeled with the d th class for the j th cluster. Clusters that fail the test are refined in subsequent iterations. Recall that unless termination criteria are met, a new cluster is introduced using information in an existing cluster, effectively splitting that cluster into two clusters. The full CIGSCR algorithm is precisely defined in the pseudocode below using the soft k -means clustering algorithm and the maximum likelihood decision rule as discussed in Part 1.

Algorithm CIGSCR

Input: X (3-dimensional image)

ϕ^{-1} (set of (row, col) indices for each class)

K_{init} (number of initial clusters)

K_{max} (maximum number of clusters)

C (number of classes)

ϵ (convergence threshold)

α (Type-I error for one-sided hypothesis test)

β (distance function penalty)

Output: DR (decision rule classification)

IS (iterative stacked classification)

begin

Initialize cluster means U along the mean plus or minus the standard deviation of the image X ;

$K := K_{init}$;

for $iteration := K_{init}$ **step 1 until** K_{max} **do**

begin

$w := 0$; $convergence := 1$;

while $convergence > \epsilon$ **do**

begin

$num := 0$; $denom := 0$;

for $i := 1$ **step 1 until** $rows$ **do**

for $j := 1$ **step 1 until** $cols$ **do**

for $k := 1$ **step 1 until** K **do**

begin

$$(*) \quad \hat{w}_{i,j,k} := \frac{1/\|X^{(ij)} - U^{(k)}\|_2^2}{\sum_{l=1}^K 1/\|X^{(ij)} - U^{(l)}\|_2^2};$$

(update sums for mean calcs.)

$$num^{(k)} := num^{(k)} + \hat{w}_{i,j,k}^2 X^{(ij)};$$

$$denom_k := denom_k + \hat{w}_{i,j,k}^2;$$

end

(update cluster means)

for $k := 1$ **step 1 until** K **do**

$$U^{(k)} := \frac{num^{(k)}}{denom_k};$$

$$convergence := \max_{i,j,k} |w_{i,j,k} - \hat{w}_{i,j,k}|;$$

```

     $w := \hat{w};$ 
end
for  $k := 1$  step 1 until  $K$  do
  begin
    Determine majority class  $c$  of cluster  $k$ ;
     $c_k := c$ ;
    (**)  $Z_k := \frac{\sqrt{n_c}(\bar{w}_{c,k} - \bar{w}_k)}{s\bar{w}_k};$ 
  end
  if any class is not associated with a cluster then
    begin
       $c :=$  first unassociated class
       $k := \operatorname{argmax}_k \frac{\bar{w}_{c,k}}{w_{c,k}}$ 
       $K := K + 1$ 
      
$$U(K) = \frac{\sum_{ij \in \phi^{-1}(c)} w_{ij,k} X^{(ij)}}{\sum_{ij \in \phi^{-1}(c)} w_{ij,k}};$$

    end
    elseif (any( $Z_k < Z(\alpha)$ ,  $k = 1, \dots, K$ )) then
      begin
         $k := \operatorname{argmin}_k Z_k$ ;
         $K := K + 1$ ;
        
$$U(K) = \frac{\sum_{ij \in \phi^{-1}(c_k)} w_{ij,k} X^{(ij)}}{\sum_{ij \in \phi^{-1}(c_k)} w_{ij,k}};$$

      end
    else
      exit for loop;
    end
  end
for  $k := 1$  step 1 until  $K$  do
  begin
    (initialize for covariance calcs.)
     $\Sigma_k := 0$ ;
     $denom_k := 0$ ;
  end
  (IS classification)
  for  $i := 1$  step 1 until  $rows$  do
    for  $j := 1$  step 1 until  $cols$  do
      begin
         $csum := 0$ ;
        for  $k := 1$  step 1 until  $K$  do
          if ( $Z_k > Z(\alpha)$ ) then
             $csum_{c_k} := csum_{c_k} + w_{ij,k}$ ;
          for  $c := 1$  step 1 until  $C$  do
             $IS_{ij,c} := \frac{csum_c}{\sum_{k=1} csum_k}$ ;
          end
        end
        (calculate covariance matrices)
      end
    end
  end

```

```

begin
   $\Sigma_k := \Sigma_k + w_{ij,k}$ 
   $\cdot (X^{(ij)} - U^{(k)})(X^{(ij)} - U^{(k)})^T$ ;
   $denom_k := denom_k + w_{ij,k}$ ;
end
end
for  $k := 1$  step 1 until  $K$  do
   $\Sigma_k := 1/denom_k \cdot \Sigma_k$ ;
  (DR classification)
for  $i := 1$  step 1 until  $rows$  do
  for  $j := 1$  step 1 until  $cols$  do
  begin
     $csum := 0$ ;
    for  $k := 1$  step 1 until  $K$  do
      if  $(Z_k > Z(\alpha))$  then
        begin
           $p := \frac{2e^{-\frac{1}{2}(X^{(ij)} - U^{(k)})^T \Sigma_k^{-1} (X^{(ij)} - U^{(k)})}}{\pi^{B/2} |\Sigma_k|^{\frac{1}{2}}}$ ;
           $csum_{c_k} := csum_{c_k} + p$ ;
        else
           $csum_{c_k} := 0$ ;
        end
      end
    for  $c := 1$  step 1 until  $C$  do
       $DR_{ij,c} := \frac{csum_c}{\sum_{k=1}^C csum_k}$ ;
    end
  end
end

```

(2) could be used in place of the less sophisticated Z_k calculation in line (**).

III. DISTANCE FUNCTIONS

For positive real numbers ρ_{ij} , $i = 1, \dots, n$; $j = 1, \dots, K + 1$, and weights w_{ij} computed in a particular way (such as in soft k -means), the addition of a cluster will result in a smaller value of the objective function $J(\rho) = \sum_{i=1}^n \sum_j w_{ij}^2 \rho_{ij}$ (refer to Part 1). Although the soft clustering iteration for the objective function $J(\rho) = \sum_{i=1}^n \sum_j w_{ij}^2 \rho_{ij}$ is only guaranteed to converge when ρ_{ij} is Euclidean distance squared (between the i th sample $X^{(i)}$ and the j th cluster prototype $U^{(j)}$), [4] suggests that other functions may be used. The Euclidean distance squared is a special case of a radial function: $f : \mathbb{R}^B \rightarrow \mathbb{R}$ is *radial* if $f(x) = f(y)$ for $\|x\|_2 = \|y\|_2$. Thus $\rho_{ij} = f(x^{(i)} - U^{(j)}) = \|x^{(i)} - U^{(j)}\|_2^2$ is radial. Some alternative radial functions include

$$f(x) = \exp(\|x\|_2^q)$$

and

$$f(x) = \|x\|_2^q$$

where $q \geq 1$ and $\rho_{ij} = f(x^{(i)} - U^{(j)})$. The advantage of using a radial function is that distances can be magnified so the difference between large and small cluster weights will be more extreme, approaching hard clustering.

None of the aforementioned metrics or radial functions influence the assignment of cluster weights based on the pre-labeled points. Semisupervised clustering uses prior information to influence a clustering method. Although the association significance test and iteration are indirectly doing this, a modified objective function could directly use prior information to influence clusters. Consider the modified objective function component

$$J_i = \sum_{j=1}^K w_{ij}^2 \rho_{ij} (1 + \beta L_{ij}), \quad i = 1, \dots, n,$$

where the term βL_{ij} is the penalty associated with assigning a labeled pixel to a cluster with a different associated label [5]. $\phi(i) = c$ is the class label of the i th labeled pixel, and let $\phi(i) = \Omega \notin \{c_1, \dots, c_C\}$ if the i th pixel is unlabeled,

$$C(j) = \begin{cases} c, & \text{if the } j\text{th cluster is associated with the} \\ & c\text{th class,} \\ \Omega, & \text{otherwise,} \end{cases}$$

$$L_{ij} = \begin{cases} 1, & \text{if } \phi(i) \neq \Omega, \phi(i) \neq C(j), C(j) \neq \Omega, \\ 0, & \text{otherwise.} \end{cases}$$

The distance function $f(x^{(i)} - U^{(j)}) = d_{ij} = \rho_{ij}(1 + \beta L_{ij})$ can be substituted for ρ_{ij} in the CIGSCR algorithm to magnify the weights of pixels labeled with the c th class to clusters associated with the c th class. Note that in place of the distance function used in line (*) of the above algorithm, one of the radial functions of Euclidean distance or a distance function with a penalty (β) could be used.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

The first dataset used to obtain experimental results for IGSCR and CIGSCR is a mosaicked Landsat Enhanced Thematic Mapper Plus (ETM+) satellite image taken from Landsat Worldwide Reference System (WRS) path 17, row 34, located in Virginia, USA, shown in Fig. 1. This image, hereafter referred to as VA1734, was acquired on November 2, 2003 and consists largely of forested, mountainous regions, and a few developed regions that are predominantly light blue and light pink in Fig. 1. Fig. 1 contains a three color representation of VA1734 where the red color band in Fig. 1 corresponds to the near infrared wavelength in VA1734, the green color band in Fig. 1 corresponds to the red wavelength in VA1734, and the blue color band in Fig. 1 corresponds to the green wavelength in VA1734. Fig. 2 contains a zoomed area of interest.

The training data for this image was created by the interpretation of point locations from a systematic, hexagonal grid over Virginia Base Mapping Program (VBMP) true color digital orthophotographs. A two class classification was performed (forest/nonforest), and classification parameters and results are given in Table 1 (DR classification) and Table 2 (IS/IS+ classification). Classification images for this dataset are given in Figs. 3 through 8.

Validation data in the form of point locations at the center of USDA Forest Service Forest Inventory and Analysis (FIA) ground plots were used to assess the accuracy of this classification. Since these validation data are typically used to evaluate crisp classifications, only homogeneous FIA plots were used (either 100 percent forest or nonforest), and these plots were obtained between 1997 and 2001. Accuracy was assessed based on an error matrix where classification results for specific points (not included in the training data set) are compared against known class values. The accuracies reported in Tables 1–4 were obtained by first converting all soft classifications to hard classifications for the purpose of comparing hard classification values to hard ground truth values. The classification results reported in Tables 1–4 used 10, 15, 20, and 25 initial clusters for IGSCR and CIGSCR. Experimental runs of IGSCR used homogeneity thresholds (test probabilities of observing the majority class in a particular cluster) of .5 and .9, with $\alpha = .01$ for all IGSCR classifications. A threshold of .9 would indicate a homogeneous cluster, but a threshold of .5 is perhaps more analogous to the new association significance test used in CIGSCR. Experimental runs of CIGSCR used traditional Euclidean distance squared in addition to two proposed radial functions $f(X^{(i,j)} - U^{(k)}) = \|X^{(i,j)} - U^{(k)}\|_2^4$ and $f(X^{(i,j)} - U^{(k)}) = \exp(-\|X^{(i,j)} - U^{(k)}\|_2)$. For all reported CIGSCR runs, $\alpha = .0001$ (values of \hat{z} tend to be high for the association significance test). All reported CIGSCR classifications used hypothesis test (2). Only three out of 24 total CIGSCR classifications reported in this paper were different using (1) and (2), and the difference in resulting classification accuracies was not significant and did not show that one test consistently resulted in higher classification accuracies than the other test. Values of \hat{z} are slightly smaller using (2) than (1), resulting in more potential for cluster refinement. Additionally, the distance function with penalty was used in classification, although results are not reported in Tables 1 and 2 because incorporating the penalty into the distance function did not increase classification accuracies in any experimental runs. Large values of β produced less accurate classification results. Finally, classification was performed using just clustering without the semisupervised framework to evaluate the effect of the combination of the association significance test and iteration in CIGSCR on classification accuracies.

The second dataset used to obtain experimental results for IGSCR and CIGSCR is a hyperspectral image of the Appomattox Buckingham State Forest in Virginia, USA. The AVIRIS 224-band, low-altitude flight lines were acquired in the winter of 1999 and ranged from approximately 400-2500nm (10nm spectral resolution) with 3.4m spatial resolution [6]. The AVIRIS data were geometrically and radiometrically corrected (to level 1B at-sensor radiance, units of microwatts per square centimeter per nanometer per steradian) by the Jet Propulsion Laboratory (JPL; Pasadena, California, USA). The three flight lines used for this study were registered (8–12 control points per flight line) to an existing 0.5m orthophoto of the area. Resampling resulted in root mean square errors (RMSE) ranging between 0.23 and 0.24 pixels [6].

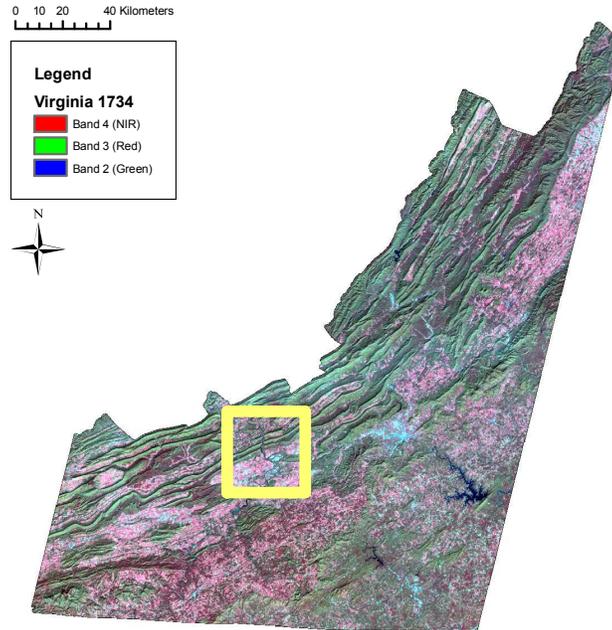


Fig. 1. Landsat ETM+ path 17/row 34 over Virginia, USA with area of interest highlighted.

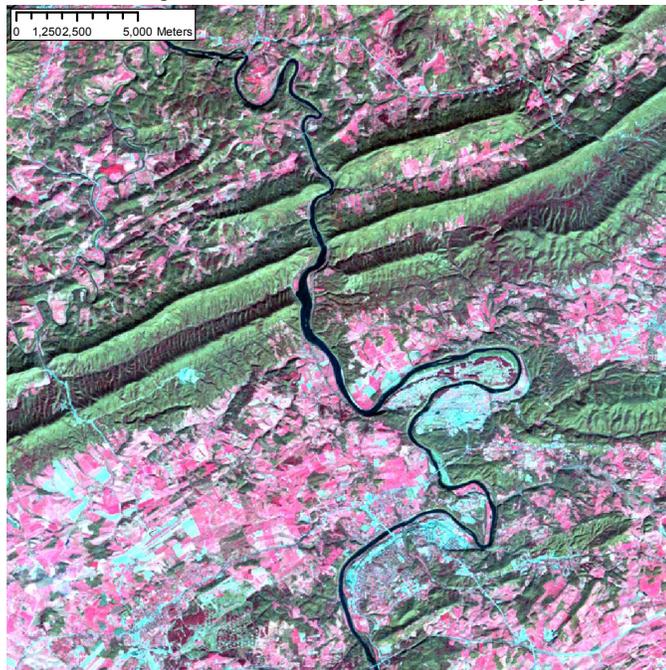


Fig. 2. Landsat ETM+ path 17/row 34 over Virginia, USA area of interest.

Training data were acquired by collecting 142 field locations [6]surrounded by homogeneous areas of single pine species (64 loblolly (*Pinus taeda*), 30 shortleaf (*Pinus echinata*), and 48 Virginia pine (*Pinus virginiana*)) with differentially corrected global positioning system (GPS) coordinates. These locations were used in a region growing algorithm to obtain a sufficient number of points for training and validation, and nonpine training data were acquired using knowledge of the area and maps of known stands in the region. The image (shown in Fig. 9 and hereafter referred to as ABSF) contains various tree stands that include the three species of pines listed above, hardwoods, and mixed (evergreens and hardwoods).

400 points were randomly selected to serve as validation data for these four classes (loblolly, shortleaf, and Virginia pines, and nonpine). Classification results for these data are reported in Tables 3 and 4, Fig. 10 contains the IGSCR IS classification

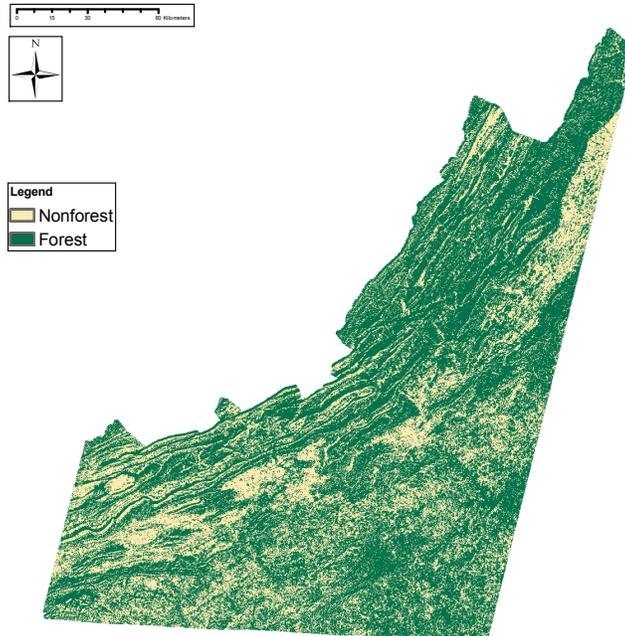


Fig. 3. IGSCR DR classification using 10 initial clusters and a homogeneity threshold of 90%.

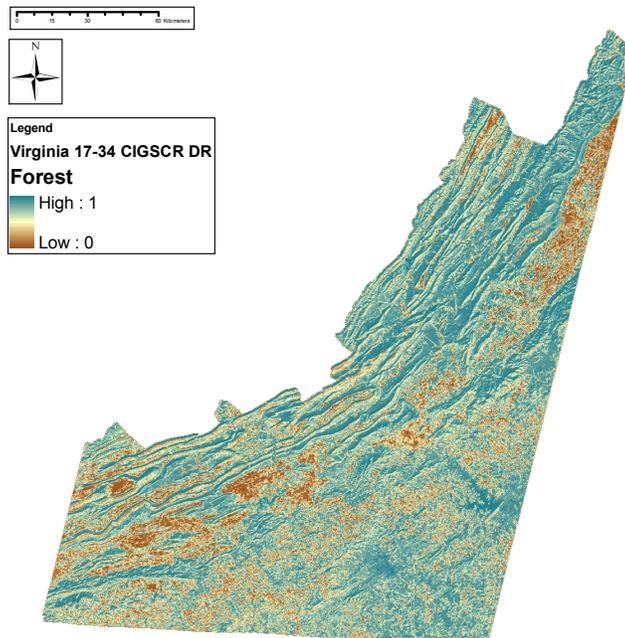


Fig. 4. CIGSCR DR classification using 10 initial clusters and Euclidean distance squared.

image using 25 initial clusters and a homogeneity threshold of .5, and Figs. 11a–d contain the CIGSCR IS classification images using 10 initial clusters and Euclidean distance to the fourth power. Classifications were run using the same parameters as classifications reported in Tables 1 and 2. An asterisk (*) indicates that the classification failed because at least one class had no associated clusters. Tables 5 and 6 report the number of pure clusters (IGSCR), and the number of clusters produced and number of associated clusters (CIGSCR).

A. Discussion.

The soft clustering and soft classification in CIGSCR can result in qualitatively different classifications than IGSCR. Even

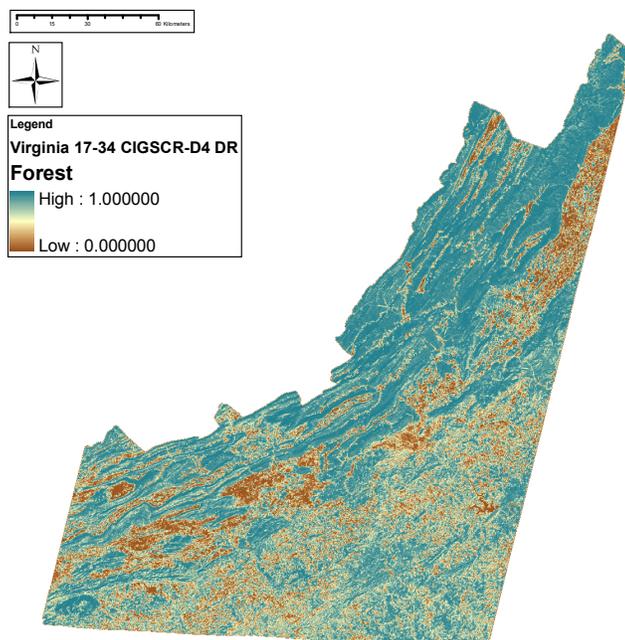


Fig. 5. CIGSCR DR classification using 10 initial clusters and Euclidean distance to the fourth power.

when the final classifications are similar, CIGSCR provides more information through soft classification. The soft classifications in Figs. 4 and 5 compared to the hard classification in Fig. 3 show that even when the hard clustering/classification in IGSCR and the soft clustering/classification in CIGSCR identify the same general regions as likely to be forest or likely to be nonforest, the soft classifications in Figs. 4 and 5 provide extra information relating to how strongly a particular sample is forest or nonforest. The dark green and dark brown colors indicate a high probability of forest and nonforest, respectively. Lighter shades of both colors indicate lower probabilities of membership in respective classes, and the beige regions indicate that the probabilities of that region being forest or nonforest are almost equal. The classifications in Figs. 10 and 11a–d show that in addition to providing more information, CIGSCR can produce qualitatively different classifications than IGSCR. The classifications present in Figs. 10 and 11a–d are the IS classifications that result from clustering, showing that soft clustering in CIGSCR produces different clustering and classification than the hard clustering in IGSCR. The regions identified by CIGSCR as being likely to contain individual pine species are different from the regions identified by IGSCR, although both algorithms identified similar nonpine regions.

Based on accuracies reported in Tables 1 and 2, CIGSCR is less sensitive to the number of initial clusters than IGSCR, especially when the alternative radial functions are used. As shown in Tables 1 and 2, IGSCR can be sensitive to the number of initial clusters and the homogeneity threshold. The set of clusters ultimately used for classification in IGSCR is directly affected by the number of initial clusters and the homogeneity test, and furthermore, when all clusters fail the homogeneity test, the iteration terminates and no more clusters are found. The number of clusters used for classification can vary widely depending on the number of iterations completed as each iteration potentially produces several pure clusters. The low accuracies reported for the IGSCR IS+ classifications in Table 2 occur when a small number of iterations occurs, which can be greatly influenced by the number of initial clusters and the homogeneity test. The classification accuracies reported for CIGSCR in Tables 1 and 2 are more consistent as CIGSCR does not have the same sensitivity issues. First, the association significance test no longer requires a user input threshold like the homogeneity test. The homogeneity test evaluates the observed values against a user supplied probability of observing a specific class (within a cluster), but the association significance test determines if the average cluster memberships per class are statistically significantly different (requiring no user specified probability). Secondly, the iteration in CIGSCR is fundamentally different from the iteration in IGSCR. While each iteration in IGSCR locates multiple clusters, each iteration in CIGSCR adds one additional cluster, and terminating this iteration potentially excludes many fewer clusters from the final classification than terminating the iteration in IGSCR (especially when few iterations occur). As classification methods are already sensitive to training data and clustering methods are sensitive to initial prototype locations, classifications being sensitive to fewer parameters is a desirable property.

The CIGSCR classifications shown in Figs. 4, 5, 7, and 8 experimentally validate the discussion in Section 7 that radial functions magnify the difference between the largest and smallest cluster weights and will more closely approximate hard

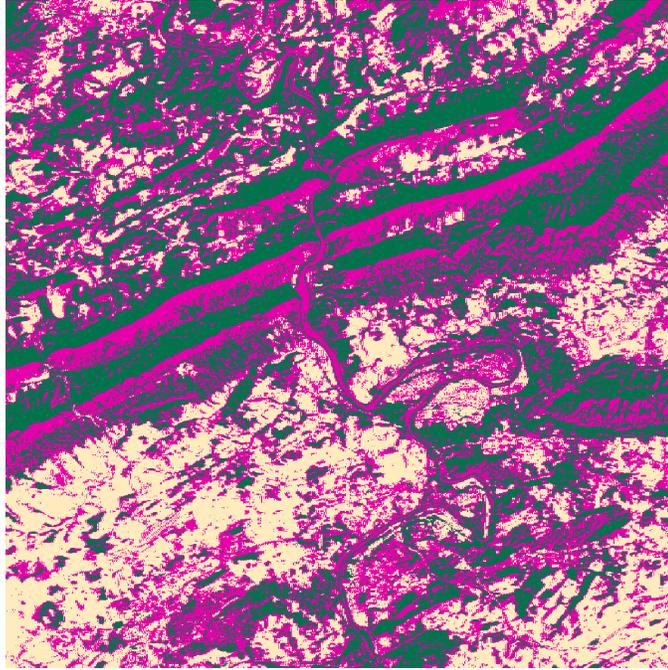


Fig. 6. IGSCR IS classification using 10 initial clusters and a homogeneity threshold of 90%.

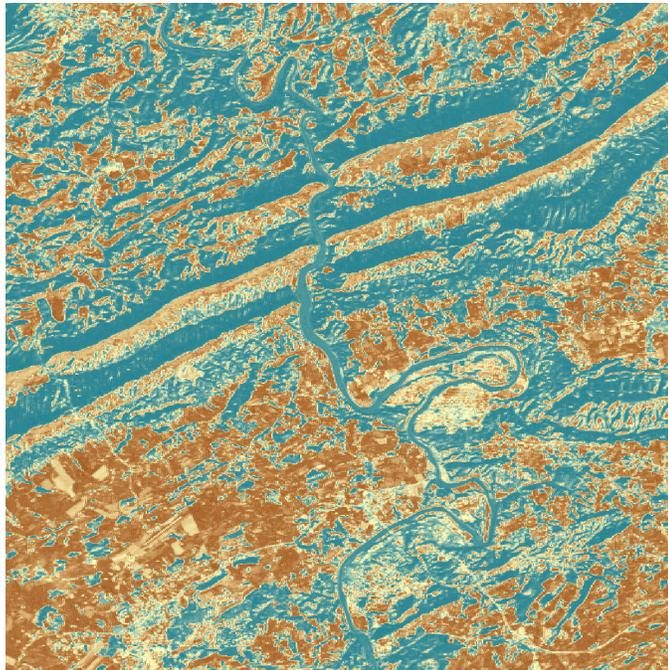


Fig. 7. CIGSCR IS classification using 10 initial clusters and Euclidean distance squared.

clustering. The classifications based on clustering with Euclidean distance to the fourth power have significantly fewer samples with almost equal probabilities of being in either class (corresponding to the beige color in the classification images). The classifications based on clustering with an exponential function of Euclidean distance (not pictured) are even closer to hard classification. Some beige areas remain in Figs. 5 and 8, indicating that although classifications based on these functions become more like hard classifications, in practice these classifications retain desirable properties of soft classification. Based on accuracies reported in Tables 1 and 2, these CIGSCR classifications with alternative radial functions are often the most accurate classifications for a given number of initial clusters. CIGSCR with alternative radial functions is accurate, can approximate

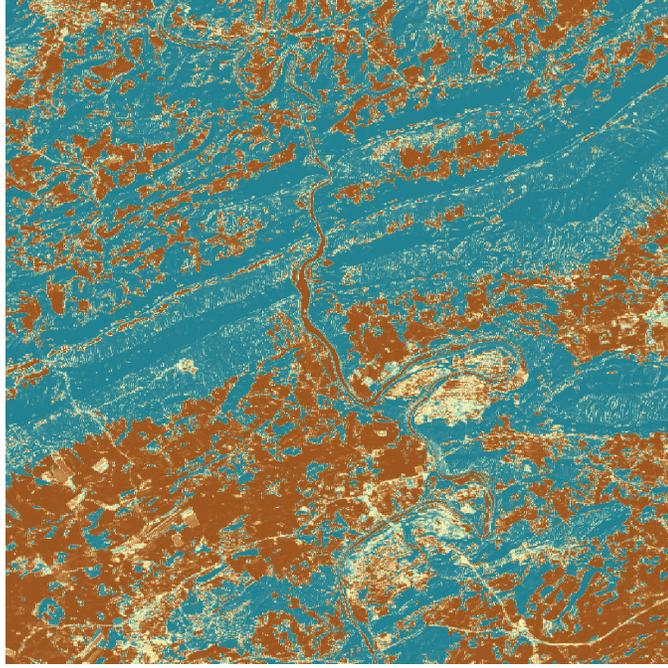


Fig. 8. CIGSCR IS classification using 10 initial clusters and Euclidean distance to the fourth power.

TABLE 1
IGSCR AND CIGSCR DECISION RULE (DR) CLASSIFICATION ACCURACIES FOR VA1734.

no. init.	IGSCR ($\alpha = .01$)		CIGSCR ($\alpha = .0001$)			clustering (no iteration)
	clusters	$p = .5$	$p = .9$	$\rho = \ x - U\ _2^2$	$\rho = \ x - U\ _2^4$	
10	85.81	75.49	88.74	87.07	87.70	72.26
15	88.22	74.56	80.50	88.53	86.97	73.72
20	84.78	89.57	79.87	89.68	88.74	76.54
25	87.49	84.25	81.44	89.47	88.74	77.58

TABLE 2
IGSCR ITERATIVE STACKED PLUS (IS+) AND CIGSCR ITERATIVE STACKED (IS)
CLASSIFICATION ACCURACIES FOR VA1734.

no. init.	IGSCR ($\alpha = .01$)		CIGSCR ($\alpha = .0001$)			clustering (no iteration)
	clusters	$p = .5$	$p = .9$	$\rho = \ x - U\ _2^2$	$\rho = \ x - U\ _2^4$	
10	68.30	75.39	83.63	84.67	85.09	72.26
15	86.34	74.56	76.96	86.03	85.19	72.99
20	84.46	88.95	75.60	85.40	86.86	76.85
25	66.63	83.94	78.52	88.32	87.28	76.75

hard classification when hard classification is desired, still provides more information than strict hard classification, and is less sensitive to input parameters than IGSCR.

All classification methods can be expected to perform poorly when training data are insufficient (samples within the dataset are not represented in the training set). This is especially true in IGSCR where spectrally pure hard clusters containing multiple training samples must be located in order for samples to be labeled with that particular class. In the VA1734 dataset, an example of a spectral class with insufficient training data is water, and although water is technically nonforest, water is often classified as forest because water and forest are spectrally similar in certain wavelength regions. This is the case in Fig. 7 where the New River running vertically through the zoomed area of interest has been identified as forest. In Fig. 6, this region is “unclassified” meaning that these pixels are not part of a pure cluster as expected (few or no water training samples are identified for this image/training dataset). Another misclassification occurs as a result of shadows in the forested mountains running diagonally in