Learning or assessment of classification algorithms relying on biased ground truth data: what interest?

Kacem Chehdi, Claude Cariou

Journal of
Applied Remote Sensing

RemoteSensing.SPIEDigitalLibrary.org

Learning or assessment of classification algorithms relying on biased ground truth data: what interest?

Kacem Chehdi
Claude Cariou

SPIE.

Kacem Chehdi, Claude Cariou, "Learning or assessment of classification algorithms relying on biased ground truth data: what interest?," *J. Appl. Remote Sens.* **13**(3), 034522 (2019), doi: 10.1117/1.JRS.13.034522.

Learning or assessment of classification algorithms relying on biased ground truth data: what interest?

Kacem Chehdi* **and Claude Cariou**
Université de Rennes/Enssat/TSI2M, CNRS, IETR—UMR 6164, Lannion, France

Abstract. The use of ground truth (GT) data in the learning and/or assessment of classification algorithms is essential. Using a biased or simplified GT attached to a remote sensing image to partition does not allow a rigorous explanation of the physical phenomena reflected by such images. Unfortunately, this scientific problem is not always treated carefully and is generally neglected in the relevant literature. Furthermore, the impacts of obtained classification results for decision-making are negative. This is inconsistent when considering investments in both the development of sophisticated sensors and the design of objective classification algorithms. Any GT must be validated according to a rigorous protocol before utilization, which is unfortunately not always the case. The evidence of this problem is provided, using two popular hyperspectral images (Indian Pine and Pavia University) that misleadingly are frequently used without care by the remote sensing community since the associated GTs are not accurate. The heterogeneity of the spectral signatures of some GT classes was proven using a semisupervised and an unsupervised classification method. Through this critical analysis, we propose a general framework for a minimum objective assessment and validation of the GT accuracy, before exploiting them in a classification method. [DOI: 10.1117/1.JRS.13.034522]

Keywords: ground truth; bias; assessment; classification; hyperspectral imaging; remote sensing.

Paper 190300 received Apr. 19, 2019; accepted for publication Aug. 12, 2019; published online Sep. 13, 2019.

1 Introduction

During the last decade, hyperspectral imagery has become an appropriate Earth observation means to help decision-making and is presently considered an excellent information source to ease analysis and interpretation of imaged objects in a variety of applications, beyond the field of remote sensing. With the development of hyperspectral imaging technology, hyperspectral imagery allows a better characterization of physical phenomena and more accurate discrimination of observed materials than traditional three-bands in the visible range (RGB) or even multispectral images (a few to tens of spectral bands).

Aerial hyperspectral imagery provides detailed and objective information on a scene by imaging narrow spectral bands over a continuous range of channels, producing the spectra of all pixels in the scene. Hyperspectral remote sensing is used in a wide array of applications due to its large spectral range (several hundreds of spectral bands covering the visible and infrared domains) and its fine spatial resolution (a few tens of centimeters). With such richness of information, the interest in hyperspectral image (HSI) data has increased during the recent years in many application fields. Among these fields, we can mention the qualitative and quantitative inventory of vegetation species and their spatial distribution,[1,2] the early detection of vegetation diseases[3,4] and invasive species,[5,6] the identification of marine algae,[7,8] the human and animal impacts on the environment,[9,10] etc.

Despite its importance and the wide range of current and potential applications it encompasses, hyperspectral imagery exploitation is still a big challenge due to difficulty in analyzing image datasets, which can be very large in both the spatial and spectral dimensions.[11,12]

*Address all correspondence to Kacem Chehdi, E-mail: kacem.chehdi@univ-rennes1.fr

To highlight and exploit this wealth of information given by HSIs, classification is a central stage in decision-making processes. It helps to summarize the image information content by assigning a unique label to similar pixels in the image, objectively based on its spectral signature. The classification methods can be categorized into three families, namely (1) supervised, (2) semisupervised, and (3) unsupervised.[13–15] Supervised methods require *a priori* knowledge of the ground truth (GT) label information in the learning and assessment stages.[16,17] In the case of semisupervised methods, the knowledge of the number of classes (often given by the GT) and/or some threshold values, or the number of iterations for iterative methods are required to perform the classification task.[18,19] Finally, unsupervised methods objectively aggregate the objects (pixels) in classes without any knowledge (neither the number of classes to discriminate nor learning samples). They estimate the number of classes and aggregate pixels in classes, by using one or several optimization criteria.[20]

Whatever the family of classification method considered, a reliable GT is always necessary because this knowledge is essential during the stages of evaluation and validation of classification results or algorithms; otherwise, classification methods will have no scientific credibility.

For instance, imagine we have an aerial HSI of cultivated areas for which the reference class information (GT) is wrongly summarized to single class content (e.g., wheat). It is very likely that this image exhibits spectral variations due to the existence of several distinct classes, though it is claimed as homogeneous and wrongly reduced to a single land cover area in the GT map. These spectral variations detected by the hyperspectral imager may come from regions in which the seeded plants did not grow uniformly for multiple reasons (plant disease, local moisture, and/or path through the plant crop, etc.) If one wants to assess an unsupervised (no prior knowledge) classification algorithm to this image, the chosen algorithm, without much a priori information, will probably be able to objectively discriminate these variations and to discover several classes that account for these variations and highlight some informational content not present in the original GT map. On the one hand, forcing pixels to belong to a wrong class during the learning stage (for supervised classification) or assuming a lower number of classes with respect to reality (for semisupervised classification) can have high negative impact: the measured classification accuracy does not significantly reflect the physical reality of the observed image since pixels with very different spectral signatures are merged into "virtual" classes. When considering the absolute reference of GT, a homogenous class formed of objects having the same or very close characteristics is necessary. A GT, therefore, must take into account the physical characteristics of objects present in the imaged scene. On the other hand, during the elaboration of GT or with the help of end-users, how the classes are forced to merge to form virtual classes must be considered; for example, in a crop field, how pixels belonging to bare soil should be grouped with those belonging to growing corn. The practical consequences of such knowledge-based (sometimes arbitrary) class merging are not very critical in the present context but might be disastrous in other application areas; for example, in the medical field, one can imagine what would be the consequences of confusing a tumor with a sane tissue after image partitioning.

Another important point is the evaluation of classification methods based on a false or simplified GT. With such a GT, unsupervised classification methods are doomed to failure and unjustifiably disqualified in contrast to supervised or semisupervised methods, though they are likely to provide classification maps closer to the physical reality.

To illustrate the problem addressed here, an analysis of the GT data associated with two well-known HSIs, namely "Indian Pine" (AVIRIS) and "Pavia University" (ROSIS), is conducted in this research. Both images have been extensively used in the remote sensing literature dealing with classification or clustering of HSI pixels. Indeed, so far, more than 200 scientific published papers mention these datasets in their abstract or keywords. By analyzing some specific classes defined in the GT map and field observations, when available, we demonstrate the fact that these reference maps are ill-conditioned and should be at least reconsidered before being used for classification purposes.

We must specify that the problem raised here does not aim to propose a method for selecting learning samples. It is rather an objective critical analysis that underlines the use of inconsistent GT data for the assessment of classification algorithms as well as the incoherent results given by certain algorithms, which closely follow the biased GTs. This scientifically worrying problem is

becoming significant and unfortunately creates a lot of confusion in the related scientific literature. It calls into question the credibility of the contribution of generation sensors and also of the accurate and objective analysis by sophisticated algorithms of the information that these sensors can acquire. This problem is not systematically avoided despite the existence of credible scientific reasons. The present paper underlines the fact that any GT should not systematically be considered as an absolute reference. Before any use, it must be validated according to a rigorous protocol, which is unfortunately not always the case. It is, therefore, important to remember that the fineness and richness of the data provided by the generation of imaging sensors, and the development of increasingly sophisticated algorithms must contribute to more objective decision-making. This paper gives a comprehensive analysis and further details of the work published by Chehdi and Cariou.[21] The steps of the proposed analysis can be used as a basic approach to validate a GT dataset.

The remaining of the paper is organized into two sections. Section 2 presents (i) a spectral signature analysis of two popular HSIs based on their associated GT maps, (ii) an assessment of the homogeneity of the GT classes by using semisupervised and unsupervised classification methods, (iii) a description of the impacts of a biased GT, and (iv) a general framework to assess and validate a given GT database. The last section provides a conclusion.

2 Spectral Signature Analysis of Biased Ground Truth of HSIs and Impacts

In the remote sensing field, the GT data associated with the acquired images are sometimes wrongly considered as "reference" data because they are incorrect or extremely simplified. This problem is particularly frequent in airborne and spaceborne remote imaging, where GT data are often utilized in an abusive and inappropriate manner. Before this finding is proven, it is very important to first recall definitions and the meaning of the GT authenticity.

2.1 *Ground Truth Definition*

According to the Oxford English dictionary,[22] there are three definitions of GT, depending on its usage:

i. Information that has been checked or facts that have been collected at source;
ii. Information obtained by direct measurement at ground level, rather than by interpretation of remotely obtained data (as aerial or satellite images, etc.), especially as used to verify or calibrate remotely obtained data;
iii. Information obtained by direct observation of a real system, as opposed to a model or simulation; a set of data that is considered to be accurate and reliable and is used to calibrate a model, algorithm, procedure, etc. In addition (specifically in image recognition technologies), information obtained by direct visual examination, especially as used to check or calibrate an automated recognition system.

These definitions converge and bring no confusion to the interpretation of the noun "GT." They are also in line with the definition given by Claval[23] in the sense "that it guarantees the authenticity of the collected observations."

2.2 *Ground Truth Authenticity*

Since the advent of technological remote sensing means, several authors have pointed out the risk of abandoning the precision and authenticity of the so-called "microlevel" knowledge (e.g., Rundstrom and Kenzer[24]), in favor of the "macrolevel" generalization. Nevertheless, the fieldwork, called "intimate sensing" by Porteous,[25] is still a necessary complement of knowledge, even at the macroscopic scale.

Whatever the application domain or the theme that a "GT" is associated with, this latter, therefore, must guarantee the authenticity and accuracy of observations and must be faultless since it is a reference, a model. In a decision-making framework based on image processing and

analysis, a GT map must be consistent with the corresponding image data since the latter are bound to the physical characteristics of objects or real materials that are present in the imaged scene. Moreover, each area declared as homogeneous must refer to the same content. Such GT area must, therefore, be coherent with the corresponding area in the HSI that objectively represents the real scene, indicating that the pixels of a homogeneous region must have similar spectral features; otherwise, the results of the objective analysis of images exploited in the decision-making process will never match those of the simplified GT. This means that any analysis method using untrue GT data will provide biased and nonrigorously exploitable results as well as, irrelevant conclusions.

To illustrate this, analysis results focusing on two significant examples, namely the cases of the Indian Pine and Pavia University datasets, are presented. These are the most widely used benchmark datasets (HSIs and associated superimposable GT maps) referred to in the remote sensing community for classification purposes. For each dataset, we first recall the characteristics of the image and the corresponding GT. Thereafter, the different analyses based on spectral signatures of the pixels are performed to put in evidence the inhomogeneity of GT classes. The anomalies of these two GTs are pointed out by calculating the spectral dispersions within the reference classes.

Furthermore, the need to subdivide the classes of the original GTs for a better coherence with the HSIs based on the spectral features is emphasized by using semisupervised and unsupervised methods. Finally, the approximations made in constructing the GT maps associated with HSIs and their negative impacts in the analysis and interpretation of their informational content are also discussed.

2.3 *Analysis of Two Biased Ground Truth*

2.3.1 *Indian Pine GT classes*

The AVIRIS Indian Pine HSI[26] has a spatial size of 145×145 pixels, where each pixel is characterized by a set of 220 spectral values (features). The spectral range is from 400 to 2499 nm. The ground spatial resolution is approximately 20 m/pixel. The corresponding GT map is made of 16 classes.

Figure 1 displays the HSI visualized under two different wavelength triplets to highlight the variations in the regions corresponding to each original GT class, as well as, the image of class labels of the associated GT. Figure 2 presents the nature of each supposed homogeneous class and the corresponding number of pixels.

Spectral signature analysis of hyperspectral images. In an HSI, a pixel is characterized by its spectral signature, a set of features corresponding to radiance or reflectance in contiguous spectral bands.

Let $\boldsymbol{X} = \{x_1, x_2, \ldots, x_N\}$ be the set of elements (pixels) to be partitioned. Each pixel x_i is characterized by the feature vector $\mathbf{A}_i = \mathbf{A}(x_i) = (a_{i1}, a_{i2}, \ldots, a_{iB})^T$, where B is the number of features (spectral bands).

Consider a partition of $\mathbf{X}$ into K indexed subsets or classes $\{C_n\}_{1 \le n \le K}$, and $l_i \in \{1, \ldots, K\}$ the label associated to pixel x_i so that the n'th class $C_n = \{x_i : l_i = n\}_{1 \le i \le N}$ and $|C_n| = M_n$. The average spectral signature (barycenter) of class C_n is given by

$$\boldsymbol{g}_n = \frac{1}{M_n} \sum_{x_i \in C_n} \boldsymbol{A}(x_i). \tag{1}$$

The metric used here to calculate the dispersion of a class C_n is the L_1-norm distance. This metric computes the global error without compensation (sum of absolute errors) between the spectral signature of an object and a reference or between two spectral signatures. This norm has been proven to be relevant for high dimensional datasets.[27]

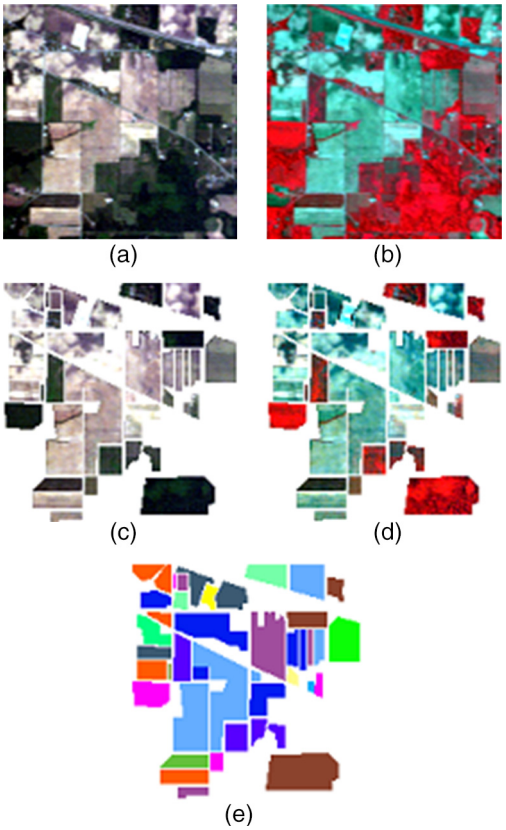


Fig. 1 Original "Indian Pine" image. (a) and (b) visualization based on two compositions of three different spectral bands (26, 16, 6) / (37, 21, 5), respectively; (c) and (d) the selected regions of the images (a) and (b), respectively, corresponding to the GT class labels given in (e).

Class label	Class name	#GT pixels		Class label	Class name	#GT pixels
C_1	Alfalfa	54		C_9	Oats	20
C_2	Corn no-till	1434		C_{10}	Soybeans no-till	968
C_3	Corn min-till	834		C_{11}	Soybeans min-till	2468
C_4	Corn	234		C_{12}	Soybeans clean-till	614
C_5	Grass/Pasture	497		C_{13}	Wheat	212
C_6	Grass/trees	747		C_{14}	Woods	1294
C_7	Grass/pasture-mowed	26		C_{15}	Bldg-Grass-Tree-Drives	380
C_8	Hay-windrowed	489		C_{16}	Stone-steel towers	95
Total GT pixels: 10 336						

Fig. 2 Data from the Indian Pine GT.

The total dispersion of class C_n is defined by

$$D_n = \sum_{x_i \in C_n} d(x_i, \boldsymbol{g}_n), \tag{2}$$

where $d(x_i, \boldsymbol{g}_n)$ is the L_1-norm distance between a pixel x_i and the barycenter $\boldsymbol{g}_n$ of class C_n is given as

$$d(x_i, \boldsymbol{g}_n) = \sum_{k=1}^{B} |a_{ik} - g_{nk}|. \tag{3}$$

In order to account for the population size within a class, we also calculate the average total dispersion of class C_n:

$$\overline{D}_n = \frac{D_n}{M_n}. \tag{4}$$

For the GT data under study, $n = 1,2,\ldots,16$; that is, $K = 16$.

Table 1 shows the results of the total dispersion and the averaged total dispersion as well as the dispersion rank of each GT class using the L_1-norm distance for the Indian Pine dataset. The four GT classes that exhibit the highest total dispersion are (in decreasing order) C_{11}, C_2, C_{12}, and C_{14}. This ranking is different when considering the average dispersions. Apart from C_{12}, these classes contain the highest number of pixels. Due to space limitation, we subsequently have limited the spectral signature analysis to C_{11} (soybeans min-till) and C_2 (corn no-till) GT classes. Figure 3 shows the selected regions of the original image corresponding to these GT classes, and Fig. 4 shows the spectral signatures of the pixels, the average spectral signature, and their standard deviation within the C_{11} and C_2 GT classes. The wavelengths of the first band and the last band are 400 and 2499 nm, respectively.

The high variations of the spectral signatures inside each GT class confirm the dissimilarity of the pixels that form these two classes. This conclusion is consistent with the disparity of these classes observed with just three bands of the original HSI, as shown in Figs. 1 and 3 and hence, no further criterion is required for the confirmation of this fact. The most homogeneous class for

Table 1 Total dispersion and average dispersion of "Indian Pine" spectral signatures per GT class using the L_1-norm distance.

Supposed GT classes	Dispersion in each class and dispersion rank		Average dispersion in each class and dispersion rank	
C_1	600 623	14	11 123	14
C_2	48 443 601	2	33 782	4
C_3	22 594 034	6	27 091	7
C_4	11 205 200	9	47 885	1
C_5	19 825 704	7	39 891	3
C_6	14 369 730	8	19 237	11
C_7	197 015	16	7 577	16
C_8	7 035 441	11	14 387	12
C_9	225 766	15	11 288	13
C_{10}	23 134 763	5	23 900	8
C_{11}	58 301 631	1	23 623	9
C_{12}	26 973 156	3	43 930	2
C_{13}	1 773 877	13	8 367	15
C_{14}	26 682 300	4	20 620	10
C_{15}	11 169 485	10	29 393	6
C_{16}	2 927 468	12	30 815	5

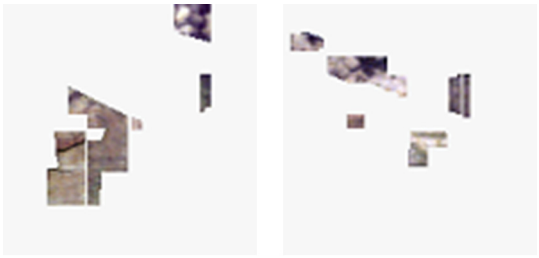


Fig. 3 Indian Pine original images visualized with three spectral bands (26, 16, 6) corresponding to the label of C_{11}, claimed as Soybeans min-till, and C_2, claimed as Corn no-till.

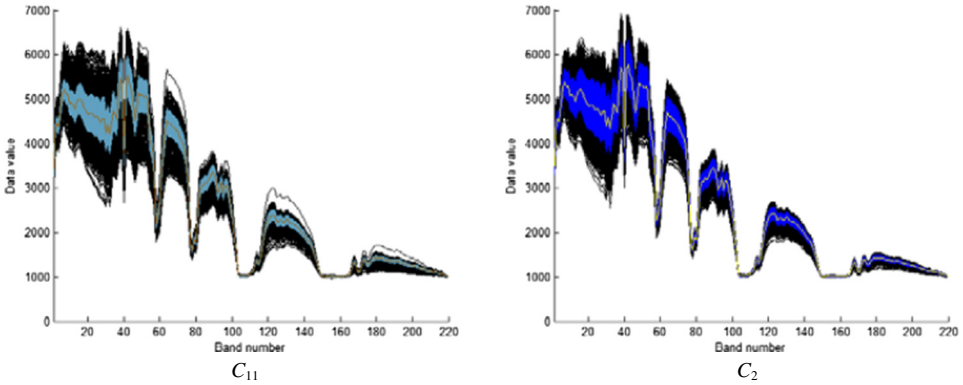


Fig. 4 Indian Pine dataset: Spectral signatures (black), average spectral signature (central curve), and ±standard deviation interval (blue) of C_{11} and C_2 GT classes.

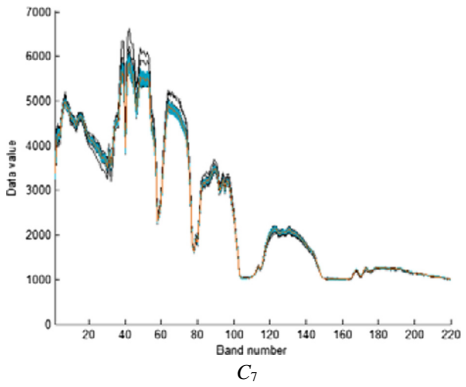


Fig. 5 Indian Pine dataset: Spectral signatures (black), average spectral signature (central curve), and ±standard deviation interval (blue) of the assumed homogeneous class C_7.

this GT is C_7, even if a few pixels are distant from the class barycenter. This fact is confirmed by observing the weaker variations of the standard deviation around the average spectral signature (Fig. 5).

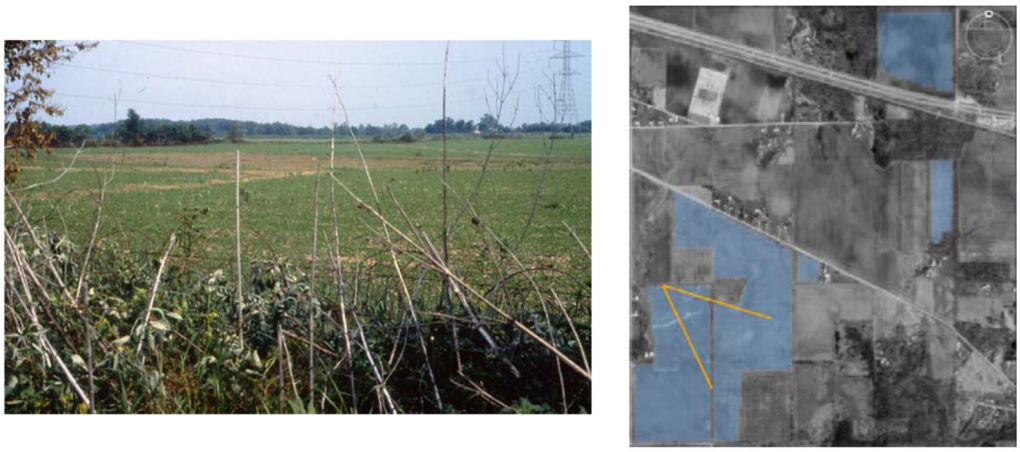
Discussion. The examples of C_{11} and C_2 classes mentioned earlier indicate that some regions of the HSI declared in the GT as relating to two classes of vegetation species do not exhibit coherent and similar spectral signatures in the acquired image. The corresponding variations can even be visually detected from visible bands on Fig. 3. One might ask whether such variations really exist from the field viewpoint and are not part of some artifacts, for example,

caused by the sensor itself. In fact, some answers to the issue of heterogeneity of most original GT classes reside in the supplemental material provided with the HSI, that is, the observation notes and field pictures associated with the field work of Baumgardner et al.,[26] which is barely referred to in the HSI classification literature. This ~70-page document including handwritten notes taken approximately at the time of the Indian Pine flight survey, as well as the pictures taken by the field specialists, contain rich information that has only partially been reported in the GT map. For instance, let us consider the field numbered as 3 to 10 in the observation notes document. This field corresponds to the bottommost leftmost field among those of C_{11} class (cf. Fig. 3). The vegetative canopy reported for this field in the observation notes is soybeans, drilled in 8-in. rows, and a plant height of 4 to 5 in., with very few weed infestations. In the same report, the soil characteristics also mention a minimum tillage system, not freshly tilled with corn residues on the surface. These observations, which are only partly reported in C_{11} GT class (soybeans min-till), seem to indicate that the same, uniform soil and vegetation conditions are available over the whole field. However, this is not the case, as can be seen from the picture of this field taken during the field observations,[26] shown in Fig. 6(a). This picture clearly exhibits local variations along lineaments traversing the north part of the field (particularly, the WSW-ENE lineament) taken from the north end of the field and in the direction of south-east. The first line of trees and bushes at the background correspond to the east end of the field. Figure 6(b) shows the C_{11} class regions overlaid on a Google Earth archive image acquired 3 months before the hyperspectral acquisition. The two orange lines in Fig. 6(b) delineate approximately the field-of-view of the picture in Fig. 6(a). It is observed that the local variations of gray levels in this image are in accordance with those observed in the HSI.

As earlier mentioned, this crop field is claimed by the GT map as uniformly grown with soybeans on a minimum tillage soil. However, the central part of the picture in Fig. 6(a) showing brown areas (probably bare soil) partly contradicts the original GT class map. Besides, this area is very likely to correspond to the lineaments detected in both the HSI [cf. C_{11} of Fig. 3 and in Fig. 6(b)].

The variations of the spectral signatures in classes C_{11} and C_2 (cf. Fig. 4) probably have two origins, with one deriving from the influence of the soil composition and moisture[28] and the other from the inclusion in these classes of objects of different natures. For example, the variations observable in field plot 3-43 (upmost of class C_{11}), which are very important have unfortunately not been reported in the GT map at all, though they were reported in the field observation document.[26] Moreover, a close examination of this document reveals that the lack of information reporting is not specific to this field plot.

From this example, it is clear that the users and developers of classification algorithms must give attention to the GT maps provided with the HSI for their absolute truthfulness.



(a) (b)

Fig. 6 (a) Picture "field_3-10a.tif" available in Ref. 26, displaying field 3 to 10 southeastward from its north end and (b) C_{11} class regions overlaid on a Google Earth archive image (March 23, 1992).

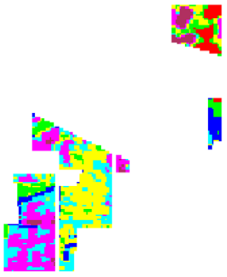
The above analysis based on both spectral considerations and visual perception clearly shows that the actual number of classes can only be greater than the original 16 GT classes. Therefore, subsequently, we propose to reconsider the number of classes by taking into account the spectral homogeneity. For this, two classification methods are used, the first one is semisupervised (K-means algorithm[14]) and the second one is unsupervised [affinity propagation (AP) algorithm[15]]. Note that the objective here is not to promote any particular classification method but to prove the coherence and effectiveness of splitting individual GT classes into subclasses representing similar objects and so, to preserve as much as possible, the physical content of the observed image. K-means is chosen for its simplicity and the possibility to vary the number of classes. AP is one of the most recent unsupervised classification methods.

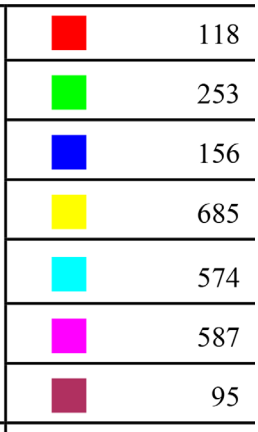
Subdivision of original GT classes by using the K-means algorithm. As shown in Fig. 4, the high variations of spectral signatures in each class C_{11} and C_2 are clearly marked, especially in the visible range (band 1: 400 nm to band 37: 715 nm). Hence, the presence of some subclasses in most supposed GT classes is evident. Each GT class can, therefore, be subdivided into subclasses. Using the popular K-means algorithm (semisupervised method),[14,29] where the number of classes and sometimes the number of iterations and/or the percentage of label changes are required (*a priori* knowledge), we demonstrate that the subdivision into subclasses is unavoidable since it leads to an objective aggregation of pixels representing identical materials. Here, we do not search the exact number of subclasses; instead, for each class, the number of subclasses has been estimated by seeking the coherence and the homogeneity of spectral signatures. The visual consistency of the subclasses formed is also taken into consideration. In this example, the number of classes retained is one that highlights both the easily visible and spatially coherent structures in the images of Fig. 3 and the low variability of the spectral signatures within each created subclass.

After partitioning the pixels of each GT class, the different subclasses issued from its subdivision are analyzed by calculating within subclasses dispersion and intersubclass distances. Due to space limitation, only the results of GT classes C_{11} and C_2 subdivisions are given in this study.

The partitioning results of these classes by applying the K-means algorithm using the ENVI software (100 iterations and 5% change threshold) are presented in Figs. 7 and 8, respectively.

At this level of analysis, it is not necessary to use a sophisticated classification method to prove our findings. For example, it is obvious that the spectral signature of the assumed bare soil area observed in Fig. 6(a) is different from the one in the nearby vegetated areas in the same field.





(a) (b) (c)

Fig. 7 Partitioning result of GT class C_{11} into seven subclasses by the K-means algorithm. (a) Original image of GT class C_{11} visualized with three bands (26, 16, 6), (b) image of subclasses, and (c) labels of subclasses in (b) and their number of pixels.

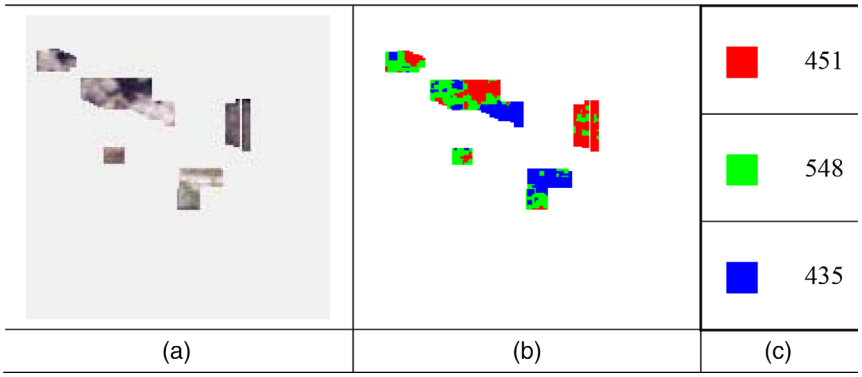


Fig. 8 Partitioning result of GT class C_2 into three subclasses by the K-means algorithm. (a) Original image of GT class C_2 visualized with three bands (26, 16, 6), (b) image of subclasses, and (c) labels of subclasses in (b) and their number of pixels.

This implies that these two areas belong to different classes. This is clearly confirmed by the subdivision result obtained, as shown in Fig. 7(b). Any opposite result would be unacceptable. We recall that the important point here is to emphasize the obvious inconsistencies between the original GT and the results of classification methods, with the help of an objective analysis of the spectral signatures of each GT class. At this level, it is important to underline, even if it is obvious, that if some GT classes are incorrectly labeled, supervised, and semisupervised classification algorithms fed with such erroneous GT are prone to reproduce the same errors.

For class C_{11}, the number of subclasses retained is 7. Figure 9 details the L_1-norm distances between class C_{11} barycenter and its subclasses barycenters as well as between pairwise subclasses. In this example, the total dispersion of class C_{11} is very high (58 301 631) as compared to the average dispersions of its subclasses (34 607). These subclasses highlight the undeniable existence of heterogeneous objects in the original C_{11} class. This is also confirmed by the nonnegligible gap between the average spectral signatures of each subclass issued from the subdivision compared with the original class before subdivision (see Figs. 7 and 10).

In the case of GT class C_2, the number of retained subclasses is 3. Figure 11 shows the average spectral signature of each subclass. The total L_1-norm distances between $SC_{2,1}$ and $SC_{2,3}$ subclasses barycenters is the highest, as seen in Fig. 12. Again, the value of the total dispersion in class C_2 is higher (48 443 601) as compared to the average dispersions of its subclasses (30 998).

To quantify the homogeneity of the subclasses obtained after partitioning classes C_{11} and C_2, the average dispersion within each subclass is also reported, respectively, in Figs. 9 and 12. These average dispersions are in all cases lower than the corresponding intersubclass distances, which shows the well-foundedness of partitioning each of these classes.

	$SC_{11,1}$	$SC_{11,2}$	$SC_{11,3}$	$SC_{11,4}$	$SC_{11,5}$	$SC_{11,6}$	$SC_{11,7}$
C_{11}	78 569	34 468	29 840	6 817	9 302	25 209	58 045
$SC_{11,1}$		44 117	61 713	72 115	87 850	103 770	136 590
$SC_{11,2}$			31 767	28 012	43 748	59 677	92 507
$SC_{11,3}$				29 962	30 780	45 880	78 710
$SC_{11,4}$					15 743	31 670	64 507
$SC_{11,5}$						15 933	48 774
$SC_{11,6}$							32 848
Average within subclass dispersions	14596	13168	16344	8443	8625	7529	13454

Fig. 9 L_1-norm distances between GT class C_{11} barycenter and subclasses barycenters, pairwise distances between subclasses barycenters, and average within subclass dispersions.

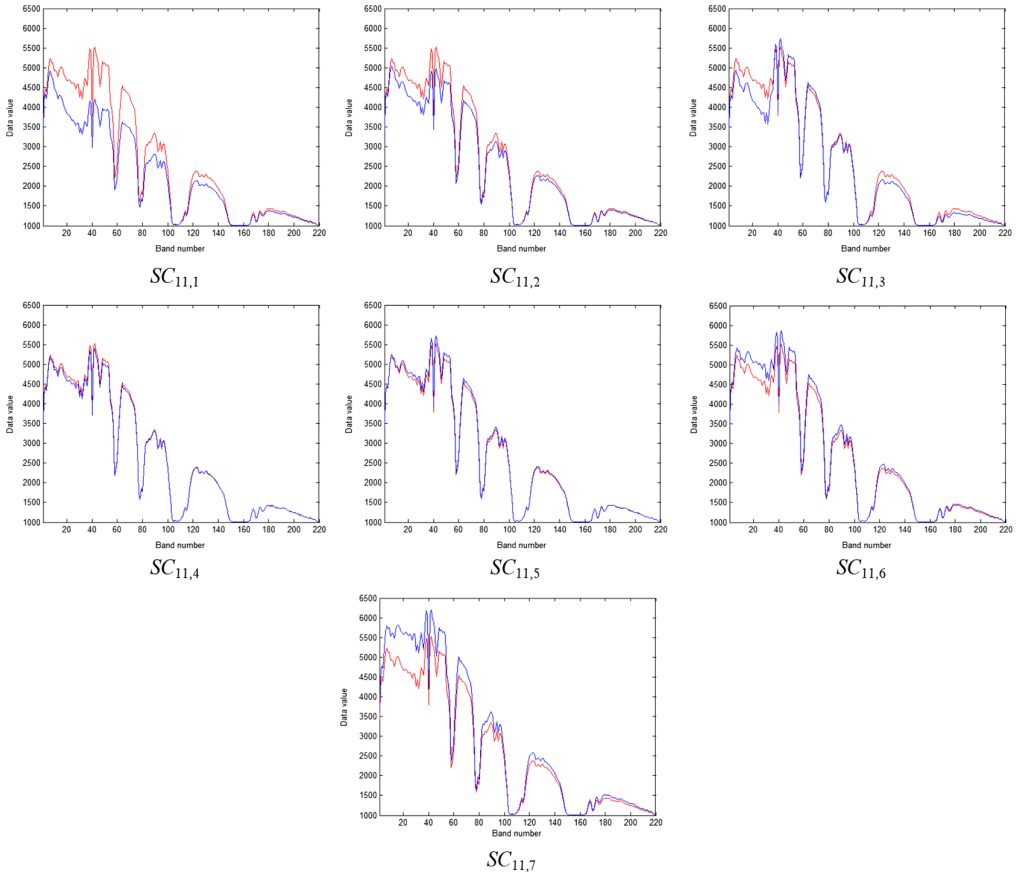


Fig. 10 Average spectral signature of the GT class C_{11} (red) and average spectral signatures of each subclass ($SC_{11,i}$) issued from the subdivision (blue).

If the correct classification rate from the above subdivision results is calculated, taking into account as reference the biased GT classes, with evidence, it will be low and unsatisfactory. On the contrary, if one considers the coherence of the spectral signatures of each formed subclass with respect to data that would result from a correct GT, this rate should inevitably be good and unbiased. The number of likely subclasses after subdivision of each original GT class, independently of the others, is given in Table 2. The partition of each original GT into subclasses reflecting the actual presence of physically distinct objects is consistent with the precise content of the HSI. This result provides more precise information to the end-user as regards the observed reality of the environment and allows a better analysis and interpretation of data.

Subdivision of original GT classes using an unsupervised algorithm. To assess the homogeneity of GT classes, we also applied an unsupervised method, which automatically estimates the number of classes partitions pixels without any prior knowledge. This method is named AP.[15] The principle of this method and its main stages are summarized in Sec. 4.

The classification of GT pixels for C_{11} and C_2 by the AP method is presented, respectively, in Figs. 13 and 14. For these results, each original GT class is partitioned independently of the others. However, some subclasses can be common to several GT classes. To avoid this, the AP method was applied to all the pixels of the hyperspectral original image corresponding only to the set of GT classes. By setting the value of the preference parameter p to the minimum of the pairwise similarity matrix [$p = \min(S)$], and the damping parameter λ to 0.9, the estimated number of classes is 17, hence greater than the one provided by the GT. Figure 15 shows the partitioning result. We note, on the one hand, that the 17 classes obtained do not correspond exactly to the 16 original classes of the GT and, on the other hand, that each GT class is subdivided into

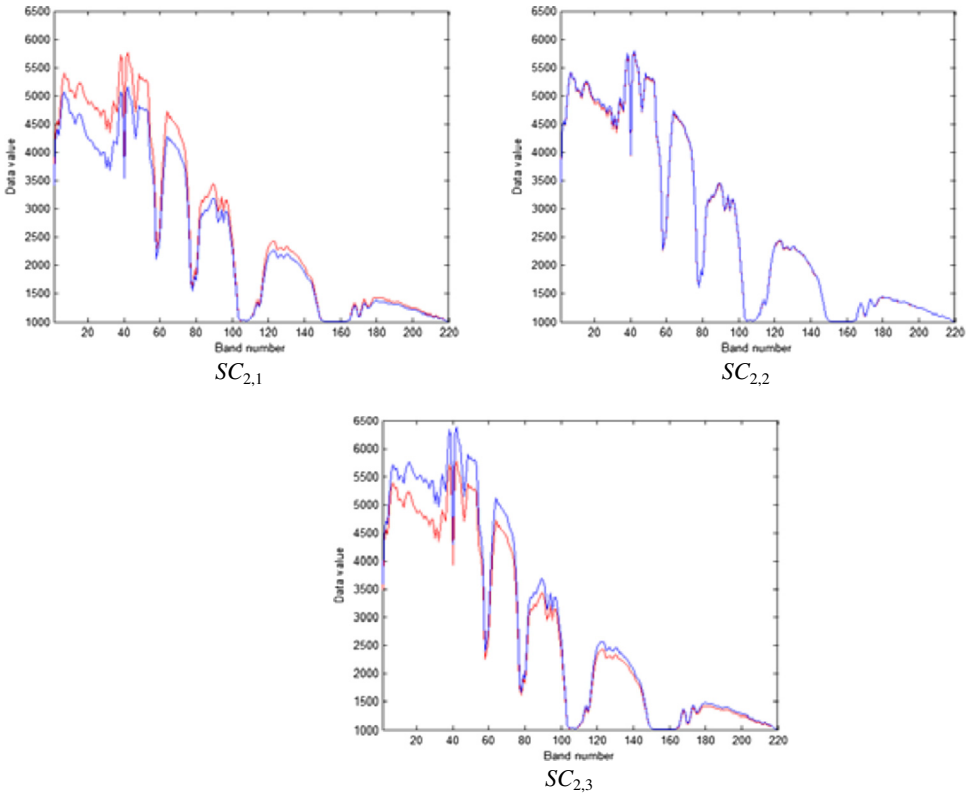


Fig. 11 Average spectral signature of the GT class C_2 (red) and average spectral signatures of each subclass ($SC_{2,i}$) issued from the subdivision (blue).

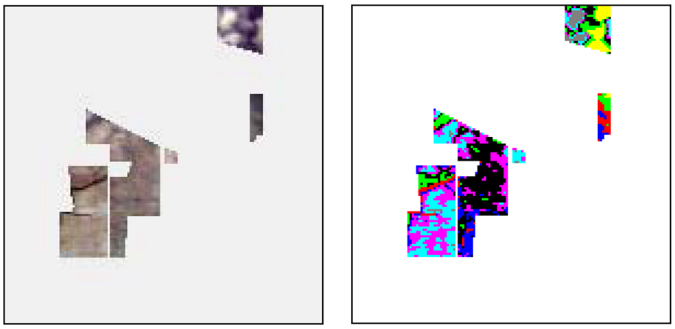
	$SC_{2,1}$	$SC_{2,2}$	$SC_{2,3}$
C_2	46 021	3 101	43 872
$SC_{2,1}$		49 076	89 891
$SC_{2,2}$			40 824
Average within subclass dispersion	20291	14604	14793

Fig. 12 L_1-norm distances between GT class C_2 barycenter and subclasses barycenters, pairwise distances between subclasses barycenters, and average within subclass dispersions.

Table 2 Number of likely subclasses after subdivision of supposed GT classes (Indian Pine) by the K-means algorithm.

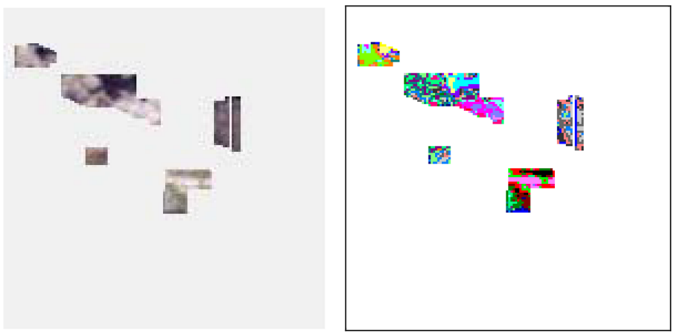
Supposed GT classes	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	C_{11}	C_{12}	C_{13}	C_{14}	C_{15}	C_{16}
Number of likely subclasses	2	3	4	3	4	4	2	3	2	4	7	3	2	3	5	2

several subclasses (cf. Table 3), which are composed of pixels belonging to several original GT classes, as detailed in Table 4. It is therefore difficult to challenge the results after observing the spectral signatures of the pixels, as illustrated in Fig. 16. This proves that all the pixels of each original GT class do not physically represent the same objects. For example, bare soil can no longer be confused with vegetated areas because they do not have the same spectral signatures. For these objective reasons, the dissimilar spectral signatures of pixels belonging to an original



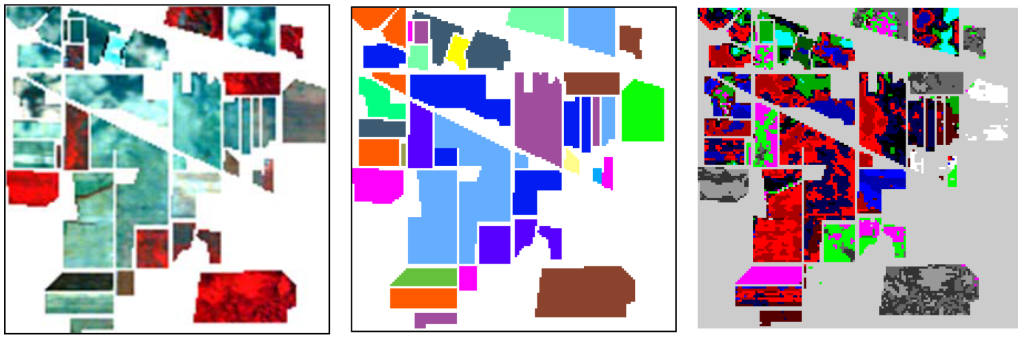
(a) (b)

Fig. 13 Partitioning result of C_{11} GT class into subclasses by AP algorithm. (a) Original image of GT class visualized with three bands (26, 16, 6) and (b) image of subclasses [$p = \min(S)$: eight subclasses estimated].



(a) (b)

Fig. 14 Partitioning result of C_2 GT class into subclasses by AP algorithm. (a) Original image of GT class visualized with three bands (26, 16, 6) and (b) image of subclasses [$p = \min(S)$: six subclasses estimated].



(a) (b) (c)

Fig. 15 Partitioning result of GT classes into 17 classes by AP algorithm. (a) The selected regions of the images [visualized with three bands (26, 16, 6)] corresponding to labels of GT classes (b), and (c) image of classes of (a) [$p = \min(S)$, 17 classes estimated].

Table 3 Number of estimated subclasses by AP after global subdivision of all GT classes (Indian Pine).

Supposed GT classes	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	C_{11}	C_{12}	C_{13}	C_{14}	C_{15}	C_{16}
Estimated number of subclasses	3	12	10	10	12	8	3	4	2	11	12	12	2	5	12	5

Table 4 Confusion matrix between GT classes (Indian Pine) and AP classes (number of common pixels).

		AP Classes																	
		C_{1AP}	C_{2AP}	C_{3AP}	C_{4AP}	C_{5AP}	C_{6AP}	C_{7AP}	C_{8AP}	C_{9AP}	C_{10AP}	C_{11AP}	C_{12AP}	C_{13AP}	C_{14AP}	C_{15AP}	C_{16AP}	C_{17AP}	Number of pixels per GT class
GT classes	C_1	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	18	31	54
	C_2	9	1	54	297	190	0	124	2	239	276	148	90	0	0	0	4	0	1434
	C_3	15	2	75	81	215	0	59	0	205	64	42	76	0	0	0	0	0	834
	C_4	0	16	53	0	6	0	42	1	1	58	50	2	0	0	0	5	0	234
	C_5	0	4	12	1	3	0	6	32	0	0	2	0	126	14	188	95	14	497
	C_6	0	264	2	0	0	0	10	386	0	0	0	0	35	31	1	18	0	747
	C_7	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	20	3	26
	C_8	0	1	0	0	0	0	0	2	0	0	0	0	0	0	0	293	193	489
	C_9	0	8	0	0	0	0	0	12	0	0	0	0	0	0	0	0	0	20
	C_{10}	5	1	158	125	191	0	75	1	203	0	2	199	0	0	0	8	0	968
	C_{11}	81	8	210	454	804	0	84	8	556	49	6	188	0	0	0	20	0	2468
	C_{12}	65	3	33	133	44	1	110	0	41	65	31	86	0	0	0	2	0	614
	C_{13}	0	211	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	212
	C_{14}	0	7	0	0	0	0	0	14	0	0	0	0	521	474	278	0	0	1294
	C_{15}	8	97	8	0	1	0	2	103	1	0	0	1	38	108	2	11	0	380
	C_{16}	0	0	0	3	1	77	0	0	3	11	0	0	0	0	0	0	0	95
Number of pixels per AP class		183	623	614	1094	1455	78	512	561	1249	523	281	642	720	627	469	494	241	
Number of original GT classes forming each AP class		6	13	12	7	9	2	9	10	8	6	7	7	4	4	4	11	4	

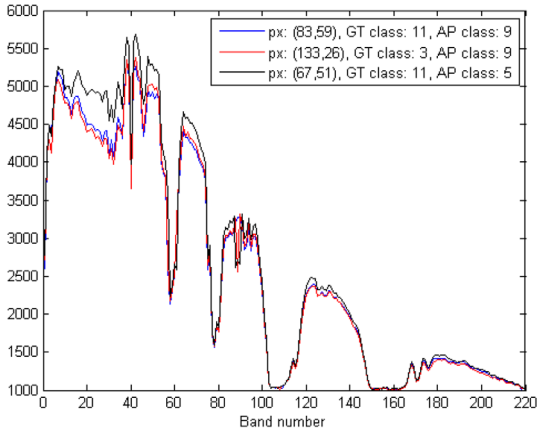


Fig. 16 Dissimilar spectral signatures of two pixels of the original GT class C_{11} (blue and black) separated by AP into two classes (5 and 9) and similar spectral signatures of pixels belonging to two originals GTs classes C_{11} and C_3 merged by AP in the same subclass 9.

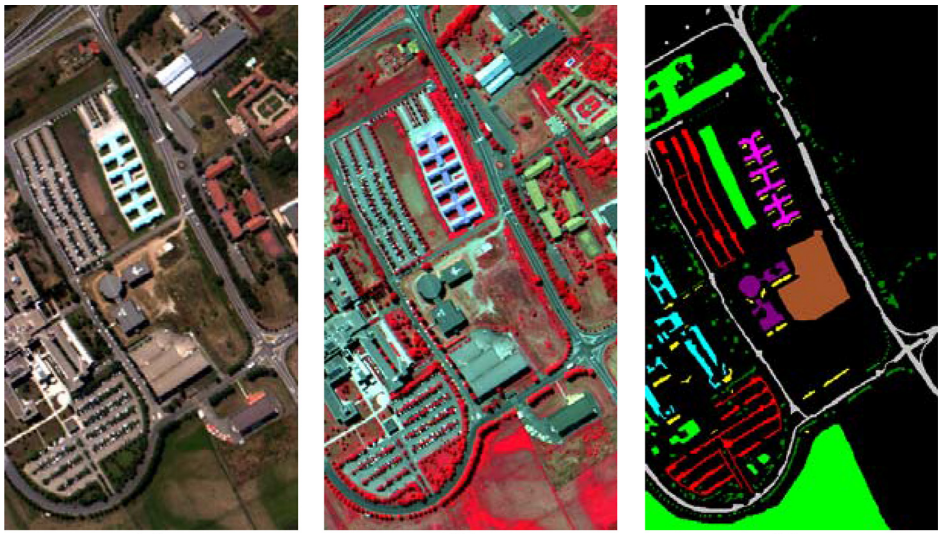
GT class are divided into subclasses, whereas similar spectral signatures of pixels belonging to different original GTs are merged in identical subclasses.

In conclusion, the analysis of the Indian Pine dataset shows many inconsistencies of the associated GT map, with identical class labels given to very distant spectral signatures. This can greatly reduce the quality of the classification results (for supervised classification using learning samples) and their assessment (for semisupervised and unsupervised methods).

2.3.2 *Pavia University GT classes*

The ROSIS Pavia University HSI (610×340 pixels) is characterized by 103 spectral bands located in the visible and near-infrared range ([430 to 860] nm).[30] The spatial resolution is 1.3 m/pixel.

Figure 17 displays this image as well as the nine labels (classes) GT map. Figure 18 presents the nature and the number of pixels of the GT classes.



(a) (b) (c)

Fig. 17 Original "Pavia University" image: (a) and (b) visualization based on two compositions of three different spectral bands (46, 27, 10) and (94, 62, 37) respectively; (c) image of the GT classes labels.

Class label	Class name	#GT pixels
C_1	Asphalt	6 852
C_2	Meadows	18 686
C_3	Gravel	2 207
C_4	Trees	3 436
C_5	(Painted) metal sheets	1 378
C_6	Bare soil	5 104
C_7	Bitumen	1 356
C_8	Self-blocking bricks	3 878
C_9	Shadow	1 026
Total GT pixels		43923

Fig. 18 Data from the Pavia University GT.

Table 5 Total dispersion and average dispersion of "Pavia University" spectral signatures per GT class using the L_1-norm distance.

Classes	Total dispersion and ranking		Average dispersion and ranking	
C_1	145 390 780	3	21 219	5
C_2	508 339 270	1	27 204	4
C_3	41 665 360	7	18 879	6
C_4	94 962 990	5	27 638	3
C_5	98 346 393	4	71 369	1
C_6	194 096 180	2	38 028	2
C_7	11 634 247	8	8 580	9
C_8	55 043 145	6	14 194	7
C_9	9 356 833	9	9 120	8

C_1 C_2 C_6

Fig. 19 Original images corresponding to each label for GT classes C_1, C_2, and C_6 visualized using three spectral bands (46, 27, 10).

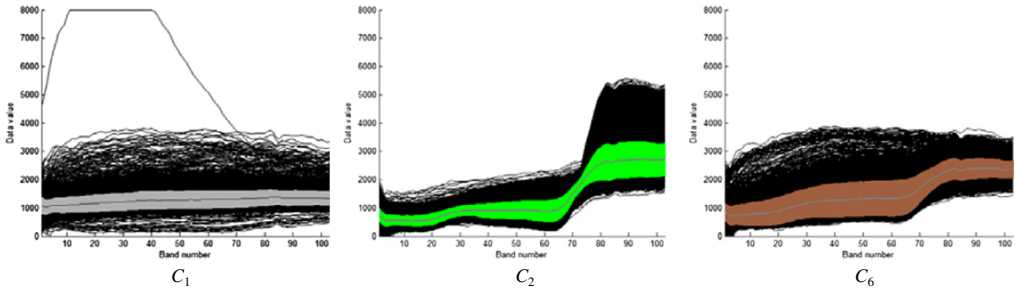


Fig. 20 Spectral signatures (black), average spectral signature (central curve), and ±standard deviation interval (color) for classes C_1, C_2, and C_6.

Spectral signature analysis. Similar to the Indian Pine case, Table 5 shows the values of total dispersions and average dispersions around the barycenters of each original GT class using L_1-norm distance. The classes presenting the highest total dispersion values in decreasing order are C_2, C_6, and C_1, whereas the ranking is different if one considers the average dispersions. These three classes are the ones that contain the highest numbers of pixel samples. Figure 19

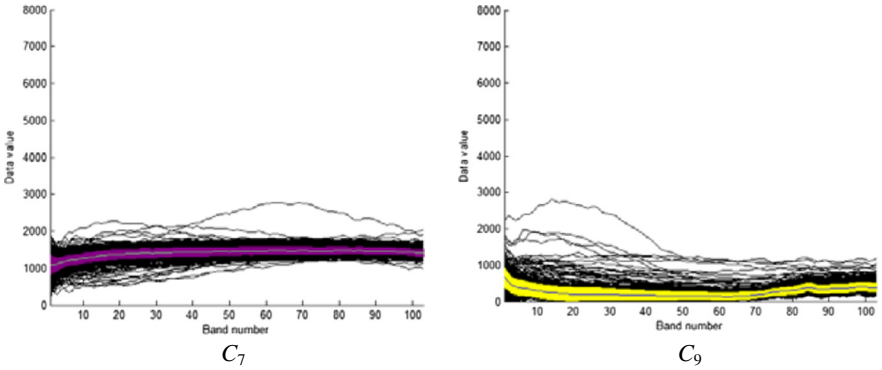


Fig. 21 Spectral signatures (black), average spectral signature (central curve), and ±standard deviation interval (color) for classes C_7 and C_9.

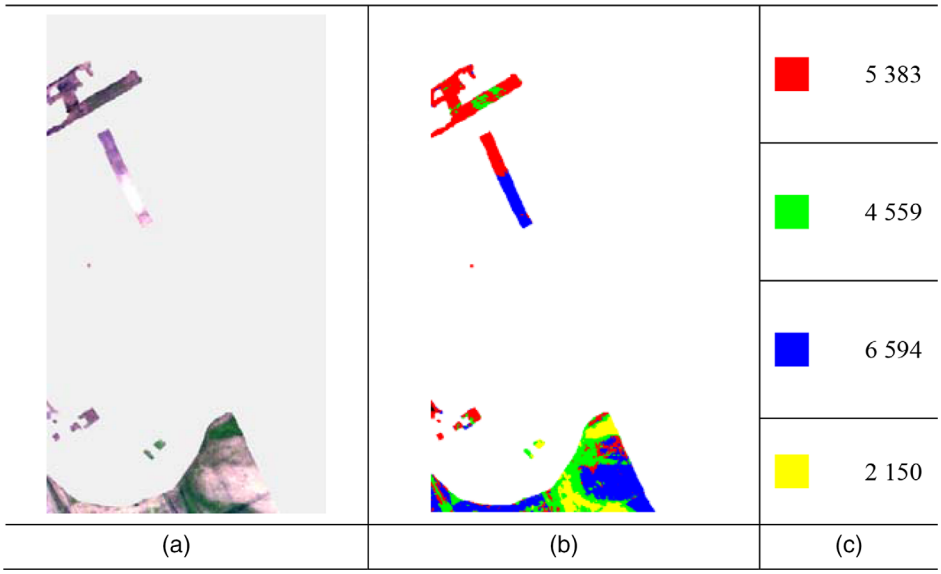


Fig. 22 Partitioning result of class C_2 into four subclasses by the K-means algorithm. (a) Original image of class C_2 visualized with three bands (46, 27, 10), (b) image of subclasses, and (c) number of pixels per subclass.

displays the areas of the original image corresponding to these classes, and Fig. 20 shows the spectral signatures, the average, and the standard deviation of these heterogeneous classes. The wavelengths of the first band and the last band are 430 and 860 nm, respectively. The high variations of these spectral signatures confirm the disparity within these classes. Figure 21 showing the variations of the spectral signatures of GT classes C_7 and C_9 also confirm the existence of dissimilar pixels, despite these two classes exhibit the lowest dispersions.

Subdivision of original GT classes using the K-means algorithm. The analysis of spectral signatures for GT classes C_1, C_2, and C_6 of Fig. 20 along with the observed image of Fig. 19 confirms the obvious presence of several subclasses in each of them. Similarly, as earlier,

	$SC_{2,1}$	$SC_{2,2}$	$SC_{2,3}$	$SC_{2,4}$
C2	20 585	14 594	22 585	61 565
$SC_{2,1}$		26 113	36 794	72 366
$SC_{2,2}$			36 987	47 099
$SC_{2,3}$				83 325
Average within subclass dispersion	13083	13240	13377	15707

Fig. 23 L_1-norm distances between GT class C_2 barycenter and subclasses barycenters, pairwise distances between subclasses barycenters and average within subclass dispersions.

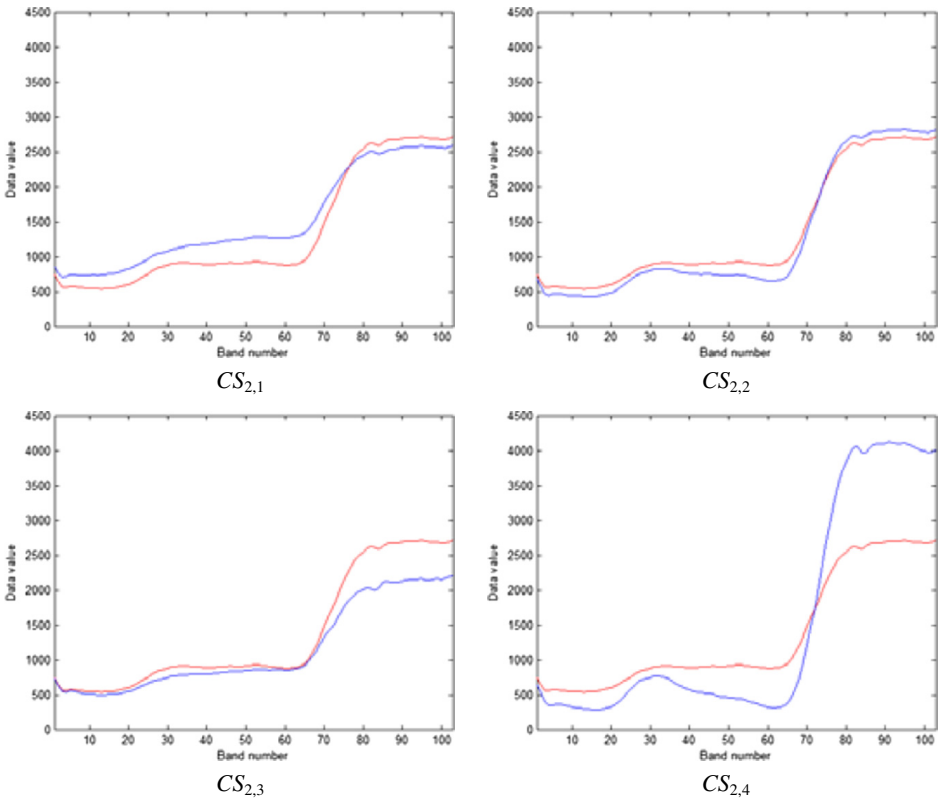


Fig. 24 Average spectral signature of the GT class C_2 (red) and average spectral signatures of each subclass ($SC_{2,i}$) issued from the subdivision (blue).

Table 6 Number of classes after subdivision of each GT class by the *K*-means algorithm (Pavia University).

Classes	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9
Number of likely subclasses	8	4	4	6	4	5	3	4	2

the subdivision of class C_2 was performed to illustrate this fact. The partitioning results by applying the *K*-means algorithm (100 iterations and 5% change threshold) are provided in Fig. 22. As mentioned above, the number of subclasses was estimated by seeking the coherence and the homogeneity of their spectral signatures. In this case, the number of subclasses retained is 4 for C_2. Figure 23 gives the total L_1-norm distances between class C_2 barycenter and the barycenters of its subclasses $C_{2,i}$ ($\{SC_{2,i}\}$, $1 \leq i \leq 4$) issued from its subdivision as well as between subclasses barycenters. The average dispersion of each subclass is also reported in this table. Again, these average dispersions are in all cases lower than the corresponding inter-subclass distances. Figure 24 shows the gap between the average signature of GT class C_2 and those of the subclasses $C_{2,i}$ issued from its subdivision, highlighting the important amount of dispersion between subclasses. The inconsistency of the average spectral signatures of the different subclasses formed after subdivision proves the invalidity of the C_2 GT class.

Table 6 shows the number of likely subclasses after subdivision of each original GT class. Each partition was obtained by maximizing the coherence of spectral signatures within subclasses. This provided a total number of 40 subclasses.

Subdivision of original C_2 GT class using the AP algorithm. By applying the AP algorithm in the same conditions as earlier mentioned to the C_2 GT class, the number of found subclasses is 65, indicating its strong heterogeneity (see Fig. 25). Recall that with the *K*-means algorithm, the number of classes was forced to 4. One can assume that these correspond to main classes and that the 65 subclasses of AP correspond to their subdivisions (cf. Fig. 25). As such, it is difficult to validate this assumption objectively from the original GT; this requires more precise additional information. Applying the AP algorithm to all GT pixels gives a partition with 71 subclasses (cf. Fig. 26).

This result perfectly illustrates the problem of evaluating an unsupervised partitioning method, as discussed earlier. Here, the number of classes is much higher than the number

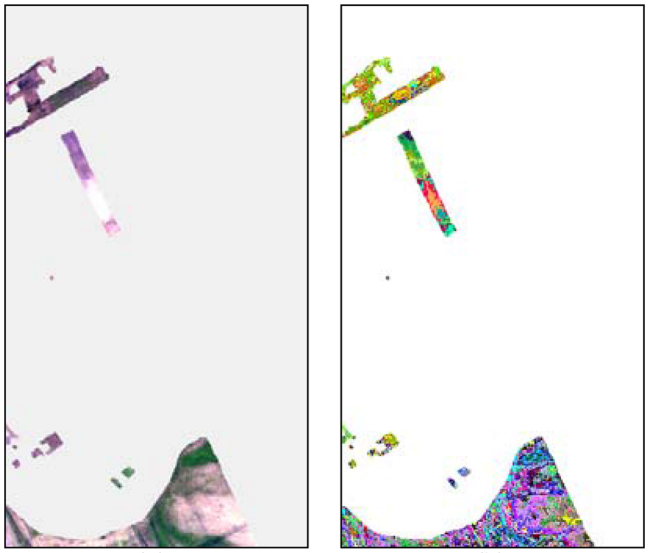
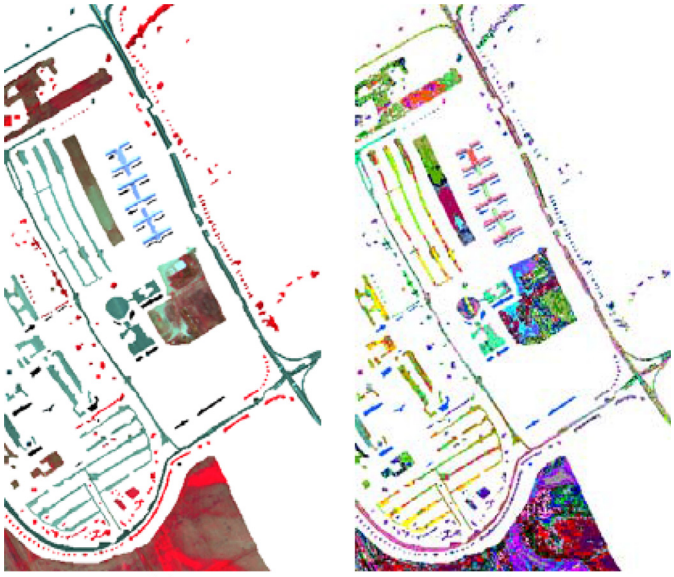


Fig. 25 Partitioning of class C_2 into subclasses. (a) Original image of class C_2 visualized with three bands (46, 27, 10) and (b) image of 65 subclasses obtained by AP [$p = \min(S)$].



(a) (b)

Fig. 26 Partitioning of the nine original GT classes into subclasses. (a) Original image of GT classes visualized with three bands (94, 62, 37) and (b) image of 71 subclasses obtained by AP [$p = \min(S)$].

of GT classes. The evaluation of the method as regards the GT data will certainly be negative, though the pixels of each formed subclass have objectively similar spectral signatures. It is difficult under such conditions to disqualify a method using nonscientifically based criteria.

In conclusion, these results confirm the heterogeneity of some GT classes for this dataset and the existence of actual subclasses that are not accounted for in the GT data.

2.3.3 *Impacts of biased ground truth in classification*

The various examples of results obtained with the spectral signature analysis of the Indian Pine and Pavia University HSIs show globally that the GT classes are not perfectly homogeneous (high variation of spectral signatures) and do not correctly represent the diversity and variability of the land cover.

Unfortunately, these two GTs have been widely used to assess the performance of classification methods.[31–36] As an indication, as of the end of 2018, we identified almost 300 scientific published papers using one or the other of these two GTs and more than 40 using both.

In the case of supervised classification methods, which exploit one or both of these GTs, we can cite the works based on the support vector machine (SVM) algorithm and its variants after optimization,[37–39] on the adaptive artificial immune network,[40] and on the adaptive simultaneous orthogonal matching pursuit.[41] Other approaches belonging to this category make use of classifiers like SVM and multilayer perceptron.[42–49] Whatever the method or approach within this category, the experimental results and the comparisons made by using these two GTs are hardly exploitable since specific regions considered as homogeneous in the associated GT are actually not. In addition to this difficulty, another problem is the choice of the learning samples and their number within each class, because the classification results are highly dependent on these. For supervised methods, the learning phase must not be conducted without scientific rigor under the pretext that they are able to reproduce what they learn. In the field of decision-making by vision, sensors provide physical measurements in order to characterize each imaged object. At a finer analysis level, the physical interpretation of these measurements must be respected. Therefore, any aggregation of heterogeneous measures by the end-user must be specified and argued.

In the category of semisupervised methods,[50–53] the classes are formed based on some objective optimization criterion free of learning samples. However, we can notice the requirement for the number of classes (which can be inferred from the GT data) and some other parameters

(threshold values and number of iterations, etc.) can influence the results when not properly and correctly chosen. For example, if the number of classes is less than the true one, the algorithms will force the objects of the unlisted classes to belong to those mentioned, if there is no defined rejection criterion. Regardless of the algorithm used, when the number of classes is underestimated by the end-user, the objects that normally do not really belong to any of the classes will be forced to belong to one of them.

In the category of unsupervised classification, the GT data are exploited exclusively during the assessment of the methods. The impact of using a biased GT is that the evaluation results are systematically average or even low[54] and therefore are doomed to be promptly disqualified. Consequently, for the assessment results to be reliable and usable, the GT data must be accurate. It is therefore essential that, in a GT, all the known subclasses of the main class must be mentioned for perfect adequacy with the precise information provided by the hyperspectral imagery. This point is very important because it is unjustified to downgrade an unsupervised classification algorithm that objectively provides accurate results and detects the real presence of objects in images.

The last key point is related to the comparison of performances between classification methods, for which GT data play an important role in the selection of the most relevant methods. Indeed, in the case where the reference data are biased, it is difficult (if not almost impossible) to objectively conclude that a method that reproduces similar errors as those contained in the GT is better than another. Actually, the question is: should one preferably consider methods giving results closer to the GT, even if this one exhibits anomalies (dissimilarity of spectral signatures within the same GT class); or methods that objectively subdivide the classes of a GT into subclasses showing similar spectral signatures (low within-class variance) and physically represent the same objects or structures? It is obvious that the second solution is scientifically more relevant and consistent as regards the means used, the reality of the phenomena observed, and the expectations of end-users.

In conclusion, the use of a biased GT in a classification process can only yield confusion when considered as an absolute reference. Beyond remote sensing, all the application fields requiring decision-making are concerned by this subject, the degree of confusion, and the negative impact being more or less serious depending on the application domain.

To avoid any confusion, and for a better exploitation of available remote sensing data, a GT must be accurate: during the field campaign, it must not be limited to the plot area and the type of land cover, but it must also include the areas, where the soil remains bare, its moisture, the presence of a river, of a path, etc.

For the collection of a reliable GT, the first condition is rigor. Field surveys must be taken under the same conditions following the same protocol. They must be a strict reflection of the physical reality of the terrain studied, without simplification or extrapolation. Nonaccessible areas with a high percentage of observable variations should be indicated and supplemented by other sources of information. Links between thematic classes should be clearly mentioned if they exist, and any aggregation of subclasses into classes should be specified and justified.

Moreover, any GT class must be analyzed and validated before using it. The analysis of the spectral features of images that accompany a GT, the use of some classification methods (semi-supervised and unsupervised) and the spectral signature analysis of the classes formed must be coherent with those of the GT classes. This cross-analysis is a minimal scientific process for the validation of a GT. Any simplified or erroneous GT should not be used. Otherwise, the results obtained would only be partial and hypothetical, and any decision based on it would have at least two principal negative impacts:

- The first one is scientific since any rigorous classification method susceptible to provide results close to reality would be mandatorily discarded if compared to other methods for which results are close to the erroneous GT. Mixing up distinct classes such as water and vegetation, even at a very coarse level of analysis is really of no interest; this does neither contribute to the progress in scientific research nor to the increase of the precision sought in the interpretation of the image informational content.
- The second impact is economical since involving very sophisticated means such as hyperspectral imaging is a very costly operation, which can lead to a low-value outcome when coupled to an inaccurate GT dataset.

3 Conclusion

The evaluation and validation steps are essential in the design process of classification algorithms. Hence, they must be conducted with high rigor. To carry out these steps, the GT data are indispensable and their precision strongly conditions the quality of the algorithms' outputs from which conclusions can be drawn. In this paper, to illustrate this problem, two examples of HSIs, namely Indian Pine and Pavia University, whose GT data are not precise, were studied. The analysis of the spectral signatures of the GT classes revealed this fundamental problem, where several classes exhibit a high heterogeneity. The existence within a GT class of pixels having very different spectral signatures indicates the presence of objects, which are physical of different nature. The main direct consequences of such a situation are:

- The difficulty in evaluating a classification method, although supervised, semisupervised, or unsupervised and mostly when evaluation is conducted toward a comparative study. Indeed, the results are hardly exploitable and prone to subjectivity when an erroneous GT is taken as an absolute reference. The problem is even more complex in the case, where supervised or semisupervised methods are used, for which learning samples and/or the number of classes are imposed. In these cases, using biased data as references in classification algorithms cannot be scientifically credible, especially when used in methods requiring a learning stage. Such results are unsuitable for serious exploitation by the end-user;
- The incoherence as regards the original motivation for introducing hyperspectral imagery. Hyperspectral remote sensing data were introduced because they give more accurate and more relevant information both spectrally and spatially compared to satellite multicomponent imagery or others. Exploiting HSIs by relying on inaccurate GT data goes totally against this objective, not mentioning the cost of image acquisition as well as the cost to elaborate the GT.

A GT must be constructed hierarchically according to several information levels depending on the physical detailed meaning of the objects to be classified. If users wish to aggregate classes representing physically different objects, they are free to do so but within another procedure outside the evaluation of a classification algorithm. For example, if vegetated and nonvegetated areas are mixed up, in a single class, the question of using precise and expensive sensors arises. We recall that the choice to aggregate classes with physically different objects is not appropriate because the primary purpose of classification is to discriminate objects and obtain the maximum relevant information about their distribution in coherent classes.

In conclusion, the use of biased GTs is of no interest and their exploitation for the development of methods and decision-making tools can bring nothing else than confusion. To fully exploit relevant information from HSIs acquired by sophisticated sensors, the accompanying GT data must, therefore, rely on objective criteria that depend only on the reality of the observed data, whatever the intended application domain. Such an approach leads to maximum exploitation of the data wealth, up to the technical and financial resources invested in the equipment and acquisition campaigns. To allow this, the GT data must make mention of any detail collected at source and carefully check and validate the information before utilization. Besides, classification methods introducing a minimum of empirical knowledge can also be incorporated into the GT data verification and validation processes. The exploitation of accurate information provided by hyperspectral imagery requires a lot of scientific rigor, and this must also be the case for the associated GT data. The more accurate the GTs, the more significant the results of any classification algorithm.

4 Appendix A: The Main Stages of the Affinity Propagation Method

Classification by AP first requires the calculation of a similarity matrix S. Each element $s(x_i, x_k)$ of this matrix indicates the similarity between pixels or objects (data points) x_i and x_k. Any type of similarities can be used. Here, the negative squared L_2-norm distance was used, that is, $s(x_i, x_k) = -\|A_i - A_k\|_2^2$.

Note that the diagonal elements $s(x_k, x_k)$ of matrix S are not computed in the same way as elements $s(x_i, x_k)$ for $x_i \neq x_k$. More precisely, $s(x_k, x_k) = p$ for all x_k, with p being the preference parameter. In our case, it is initialized to the minimum value of the elements of S.

The AP algorithm calculates degrees of availability and responsibility to the other pixels in an iterative way for each pixel. Initially, all pixels are considered as potential exemplars, though for each one, a preference parameter p value is allocated so that it can be chosen as an exemplar. Two procedures of message transmission (responsibility and availability) are used to exchange messages between pixel x_i and a candidate exemplar x_k. The responsibility $r(x_i, x_k)$ [Eqs. (5) and (6)] is the message sent from pixel x_i to candidate exemplar x_k, indicating how well-suited pixel x_k would be as the exemplar for pixel x_i. Alternatively, the availability $a(x_i, x_k)$ [Eqs. (7) and (8)] is the message sent from candidate exemplar x_k to pixel x_i, indicating how likely pixel x_i would choose candidate x_k as its exemplar. This procedure identifies for each pixel the exemplar that maximizes the sum of responsibility and availability denoted by $E^*(x_i)$ [Eq. (9)]. For pixel x_i, the pixel x_k that maximizes $\{r(x_i, x_k) + a(x_i, x_k)\}$ either identifies x_i as an exemplar if $x_k = x_i$ or identifies x_k as its exemplar.

The updated messages [Eqs. (10) and (11)] are damped by a constant factor, $\lambda \in]0,1[$, to avoid numerical oscillations that may arise under some circumstances.

Each iteration of the AP algorithm consists of (i) updating all responsibilities given the availabilities, (ii) updating all availabilities given the responsibilities, and (iii) combining availabilities and responsibilities to monitor the exemplar decisions and terminate the classification process.

The main steps of the algorithm are given below.

Step 1: Initialization

For N pixels to be classified, R, A, and S are the responsibility, availability, and similarity matrices of size $N \times N$, respectively. $r(x_i, x_k)$, $a(x_i, x_k)$, and $s(x_i, x_k)$ are, respectively, their elements for pixels x_i and x_k. $a(x_i, x_k) = 0$, for all x_i, x_k.

Step 2: Responsibility updates:

$$r(x_i, x_k) = s(x_i, x_k) - \max_{x_j, x_j \neq x_k}\{a(x_i, x_j) + s(x_i, x_j)\} \quad \text{for } x_i \neq x_k, \tag{5}$$

$$r(x_k, x_k) = p - \max_{x_j, x_j \neq x_k}\{r(x_k, x_j) + a(x_k, x_j)\}. \tag{6}$$

Step 3: Availability updates:

$$a(x_i, x_k) = \min\left[0, r(x_k, x_k) + \sum_{x_j, x_j \neq \{x_i, x_k\}} \max\{0, r(x_j, x_k)\}\right] \quad \text{for } x_i \neq x_k, \tag{7}$$

$$a(x_k, x_k) = \sum_{(x_j, x_j \neq x_k)} \max\{0, r(x_j, x_k)\}. \tag{8}$$

Step 4: Making assignments:

$$E^*(x_i) = \arg\max_{x_k}\{(r(x_i, x_k) + a(x_i, x_k)\}, \tag{9}$$

where $E^*(x_i)$ identifies the pixel x_k as exemplar of x_i.

Updated messages are sent iteratively after regularization at the current iteration l of responsibilities $\widehat{R}_l$ and availabilities $\widehat{A}_l$ as follows:

$$\widehat{R}_l = \lambda \widehat{R}_{l-1} + (1 - \lambda) R_l, \tag{10}$$

$$\widehat{A}_l = \lambda \widehat{A}_{l-1} + (1 - \lambda) A_l. \tag{11}$$

References

1. B. D. Bue et al., "Leveraging in-scene spectra for vegetation species discrimination with MESMA-MDA," *ISPRS J. Photogramm. Remote Sens.* **108**, 33–48 (2015).
2. K. L. Dudley et al., "A multi-temporal spectral library approach for mapping vegetation species across spatial and temporal phenological gradients," *Remote Sens. Environ.* **167**, 121–134 (2015).
3. W. Kong et al., "Application of hyperspectral imaging to detect sclerotinia sclerotiorum on oilseed rape stems," *Sensors* **18**(1), 123 (2018).
4. P. Schmitter et al., "Unsupervised domain adaptation for early detection of drought stress in hyperspectral images," *ISPRS J. Photogramm. Remote Sens.* **131**, 65–76 (2017).
5. K. Peerbhay et al., "Detecting bugweed (Solanum mauritianum) abundance in plantation forestry using multisource remote sensing," *ISPRS J. Photogramm. Remote Sens.* **121**, 167–176 (2016).
6. S. Stagakis, T. Vanikiotis, and O. Sykioti, "Estimating forest species abundance through linear unmixing of CHRIS/PROBA imagery," *ISPRS J. Photogramm. Remote Sens.* **119**, 79–89 (2016).
7. A. A. Mogstad and G. Johnsen, "Spectral characteristics of coralline algae: a multi-instrumental approach, with emphasis on underwater hyperspectral imaging," *Appl. Opt.* **56**(36), 9957–9975 (2017).
8. M. Mehrubeoglu, M. Y. Teng, and P. V. Zimba, "Resolving mixed algal species in hyperspectral images," *Sensors* **14**(1), 1–21 (2014).
9. J. Lopatin et al., "Mapping plant species in mixed grassland communities using close range imaging spectroscopy," *Remote Sens. Environ.* **201**, 12–23 (2017).
10. S. J. Walsh et al., "QuickBird and Hyperion data analysis of an invasive plant species in the Galapagos Islands of Ecuador: implications for control and land use management," *Remote Sens. Environ.* **112**(5), 1927–1941 (2008).
11. J. Senthilnath et al., "Crop stage classification of hyperspectral data using unsupervised techniques," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **6**(2), 861–866 (2013).
12. P. Pahlavani and B. Bigdeli, "A mutual information-Dempster-Shafer based decision ensemble system for land cover classification of hyperspectral data," *Front. Earth Sci.* **11**(4), 774–783 (2017).
13. A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Stat. Soc.* **39**(1), 1–38 (1977).
14. J. B. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. Fifth Berkeley Symp. Math. Stat. and Probab.*, L. M. Le Cam and J. Neyman, Eds., University of California Press, Vol. 1, pp. 281–297 (1967).
15. B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science* **315**(5814), 972–976 (2007).
16. P. Ghamisi et al., "Automatic framework for spectral-spatial classification based on supervised feature extraction and morphological attribute profiles," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **7**(6), 2147–2160 (2014).
17. H. Qu et al., "Dimensionality-varied deep convolutional neural network for spectral-spatial classification of hyperspectral data," *J. Appl. Remote Sens.* **12**(1), 016007 (2018).
18. P. Ghamisi et al., "A novel evolutionary swarm fuzzy clustering approach for hyperspectral imagery," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **8**(6), 2447–2456 (2015).
19. H. Li et al., "Performance evaluation of cluster validity indices (CVIs) on multi/hyperspectral remote sensing datasets," *Remote Sens.* **8**(4), 295 (2016).
20. K. Chehdi, M. Soltani, and C. Cariou, "Pixel classification of large size hyperspectral images by affinity propagation," *J. Appl. Remote Sens.* **8**(1), 083567 (2014).
21. K. Chehdi and C. Cariou, "The true false ground truths: what interest?" *Proc. SPIE* **10004**, 100040M (2016).
22. M. Waite, *Oxford English Dictionary*, Oxford University Press, Oxford (2012).
23. P. Claval, "Le rôle du terrain en géographie," *Confins* (17), 23 (2013).
24. R. A. Rundstrom and M. S. Kenzer, "The decline of fieldwork in human geography," *Prof. Geogr.* **41**(3), 294–303 (1989).

25. J. D. Porteous, "Intimate sensing," *Area* **18**(3), 250–251 (1986).
26. M. Baumgardner, L. Biehl, and D. Landgrebe, "220 band AVIRIS hyperspectral image data set: June 12, 1992 Indian Pine Test Site 3," Purdue University Research Repository, https://purr.purdue.edu/publications/1947/1 (2015).
27. C. C. Aggarwal, A. Hinneburg, and D. A. Keim, "On the surprising behavior of distance metrics in high dimensional space," *Lect. Notes Comput. Sci.* **1973**, 420–434 (2001).
28. C.-W. Ahn, M. F. Baumgardner, and L. L. Biehl, "Delineation of soil variability using geostatistics and fuzzy clustering analyses of hyperspectral data," *Soil Sci. Soc. Am. J.* **63**(1), 142 (1999).
29. J. Xu et al., "A novel hyperspectral image clustering method with context-aware unsupervised discriminative extreme learning machine," *IEEE Access* **6**, 16176–16188 (2018).
30. Grupo Inteligencia Computacional (UPV/EHU), "Hyperspectral remote sensing scenes," http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes (2014).
31. A. B. Santos, A. D. A. Araujo, and D. Menotti, "Combining multiple classification methods for hyperspectral data interpretation," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **6**(3), 1450–1459 (2013).
32. Y. Y. Tang, Y. Lu, and H. Yuan, "Hyperspectral image classification based on three-dimensional scattering wavelet transform," *IEEE Trans. Geosci. Remote Sens.* **53**(5), 2467–2480 (2015).
33. X. Zhu, N. Li, and Y. Pan, "Optimization performance comparison of three different group intelligence algorithms on a SVM for hyperspectral imagery classification," *Remote Sens.* **11**(6), 734 (2019).
34. H. Huang, Z. Li, and Y. Pan, "Multi-feature manifold discriminant analysis for hyperspectral image classification," *Remote Sens.* **11**(6), 651 (2019).
35. S. S. Sawant and P. Manoharan, "New framework for hyperspectral band selection using modified wind-driven optimization algorithm," *Int. J. Remote Sens.* **40**(20), 7852–7873 (2019).
36. T. Zhan et al., "Hyperspectral classification using an adaptive spectral-spatial kernel-based low-rank approximation," *Remote Sens. Lett.* **10**(8), 766–775 (2019).
37. A. Plaza et al., "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.* **113**(Suppl. 1), S110–S122 (2009).
38. Y. Tarabalka et al., "SVM- and MRF-based method for accurate classification of hyperspectral images," *IEEE Geosci. Remote Sens. Lett.* **7**(4), 736–740 (2010).
39. F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.* **42**(8), 1778–1790 (2004).
40. Y. Zhong and L. Zhang, "An adaptive artificial immune network for supervised classification of multi-/hyperspectral remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.* **50**(3), 894–909 (2012).
41. J. Zou et al., "Classification of hyperspectral urban data using adaptive simultaneous orthogonal matching pursuit," *J. Appl. Remote Sens.* **8**(1), 085099 (2014).
42. A. B. Santos et al., "Feature selection for classification of remote sensed hyperspectral images: a filter approach using genetic algorithm and cluster validity," in *Proc. Int. Conf. Image Process. Comput. Vision, Pattern Recognit.*, Vol. 2 (2012).
43. X. Huang and L. Zhang, "A comparative study of spatial approaches for urban mapping using hyperspectral ROSIS images over Pavia City, northern Italy," *Int. J. Remote Sens.* **30**(12), 3205–3221 (2009).
44. B. Bigdeli, F. Samadzadegan, and P. Reinartz, "A multiple SVM system for classification of hyperspectral remote sensing data," *J. Indian Soc. Remote Sens.* **41**(4), 763–776 (2013).
45. J. Li et al., "Generalized composite kernel framework for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.* **51**(9), 4816–4829 (2013).
46. K. Kavitha, S. Arivazhagan, and B. Suriya, "Classification of Pavia University hyperspectral image using Gabor and SVM classifier," *Int. J. New Trends Electron. Commun.* **2**(3), 9–14 (2014).
47. S. Jia, X. Zhang, and Q. Li, "Spectral-spatial hyperspectral image classification using $\ell 1/2$ regularized low-rank representation and sparse representation-based graph cuts," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **8**(6), 2473–2484 (2015).

48. L. C. B. Dos Santos et al., "Unsupervised hyperspectral band selection based on spectral rhythm analysis," in *Brazilian Symp. Comput. Graph. Image Process.*, pp. 157–164 (2014).
49. K. Makantasis et al., "Deep supervised learning for hyperspectral data classification through convolutional neural networks," in *IEEE Int. Geosci. Remote Sens. Symp.*, pp. 4959–4962 (2015).
50. Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.* **28**(1), 84–95 (1980).
51. Z. Wang, N. M. Nasrabadi, and T. S. Huang, "Spatial-spectral classification of hyperspectral images using discriminative dictionary designed by learning vector quantization," *IEEE Trans. Geosci. Remote Sens.* **52**(8), 4808–4822 (2014).
52. J. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York (1981).
53. A. Ma, Y. Zhong, and L. Zhang, "Spectral-spatial clustering with a local weight parameter determination method for remote sensing imagery," *Remote Sens.* **8**(2), 124 (2016).
54. C. McCann et al., "Novel histogram based unsupervised classification technique to determine natural classes from biophysically relevant fit parameters to hyperspectral data," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **10**(9), 4138–4148 (2017).

Kacem Chehdi received his PhD and HDR "Habilitation à Diriger des Recherches" degrees in signal processing from the University of Rennes 1, France, in 1986 and 1992, respectively. He is currently with the University of Rennes 1. He is a full professor of signal and image processing since 1993. He leads a research team for multicomponent and multimodal image processing. His research interests include blind restoration and filtering, unsupervised classification, and adaptive decision-making systems.

Claude Cariou received his PhD in electronics from the University of Brest, France, in 1991. Since 1992, he has been an assistant professor at the University of Rennes 1, Ecole Nationale Supérieure des Sciences Appliquées et de Technologie (ENSSAT), Lannion, France. His research interests include image analysis, pattern recognition, unsupervised classification, texture modeling and segmentation, image registration, and feature selection, especially with application to multi- and hyperspectral remote sensing images.