

# Semisupervised Hyperspectral Image Classification Using Soft Sparse Multinomial Logistic Regression

Jun Li, José M. Bioucas-Dias, *Member, IEEE*, and Antonio Plaza, *Senior Member, IEEE*

**Abstract**—In this letter, we propose a new semisupervised learning (SSL) algorithm for remotely sensed hyperspectral image classification. Our main contribution is the development of a new soft sparse multinomial logistic regression model which exploits both *hard* and *soft* labels. In our terminology, these labels respectively correspond to labeled and unlabeled training samples. The proposed algorithm represents an innovative contribution with regard to conventional SSL algorithms that only assign *hard* labels to unlabeled samples. The effectiveness of our proposed method is evaluated via experiments with real hyperspectral images, in which comparisons with conventional semisupervised self-learning algorithms with *hard* labels are carried out. In such comparisons, our method exhibits state-of-the-art performance.

**Index Terms**—Hyperspectral image classification, semisupervised learning (SSL), soft labels, sparse multinomial logistic regression (SMLR), unlabeled training samples.

## I. INTRODUCTION

RE MOTELY sensed hyperspectral image classification is an active area of research [1]. It takes advantage of the detailed information contained in each pixel (vector) of a hyperspectral image to generate thematic maps from detailed spectral signatures. A relevant challenge for supervised classification techniques (which assume prior knowledge in the form of class labels for different spectral signatures) is the limited availability of labeled training sets, since their collection generally involves expensive ground campaigns [2]. While the collection of labeled samples is generally difficult, expensive, and time consuming, unlabeled samples can be generated in a much easier way. This observation has fostered the idea of adopting semisupervised learning (SSL) techniques in hyperspectral image classification. The main assumption of such techniques is that the new (unlabeled) training samples [3] can be obtained from a (limited) set of available labeled samples without significant effort/cost [4].

Manuscript received December 28, 2011; revised April 19, 2012; accepted May 29, 2012. This work was supported in part by the European Community's Marie Curie Research Training Networks Programme under Contract MRTNCT-2006-035927 (Hyperspectral Imaging Network) and by the Portuguese Science and Technology Foundation, under project PEst-OE/EEI/LA0008/2011.

J. Li and A. Plaza are with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escola Politécnica de Cáceres, University of Extremadura, 10071 Cáceres, Spain (e-mail: junli@unex.es; aplaza@unex.es).

J. M. Bioucas-Dias is with the Instituto de Telecomunicações, Instituto Superior Técnico, Universidade Técnica de Lisboa, 1049-001 Lisbon, Portugal (e-mail: bioucas@lx.it.pt).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/LGRS.2012.2205216

The area of SSL has experienced a significant evolution in terms of the adopted models, which comprise complex generative models [5], self-learning models [6], multiview learning models [7], transductive support vector machines (SVMs) [8], and graph-based methods [9]. A survey of SSL algorithms is available in [10]. Most SSL algorithms use some type of regularization which encourages the fact that “similar” features are associated to the same class. The effect of such regularization is to push the boundaries between classes toward regions with low data density, where the usual strategy adopted first associates the vertices of a graph to the complete set of samples and then builds the regularizer depending on the variables defined at the vertices. This trend has been successfully adopted in several remote sensing image classification studies [11]–[13]. In general terms, these algorithms add *hard* labels to the set of unlabeled samples and then use jointly the labeled and unlabeled sets to improve the obtained classification results [11].

Although the aforementioned methods generally exhibit good performance, difficulties may arise from the viewpoint of the complexity of the model and its high computational cost. Furthermore, when a *hard* label is inappropriately estimated (which may often happen when limited training samples are available), the learning process is mostly driven by unlabeled samples. Most importantly, it is well known that hyperspectral images are dominated by mixed pixels. In this case, assigning a single (*hard*) label to a pixel as a whole may be a potential source of errors if several spectral constituents participate in the spectral signature associated to the pixel, as it is common in practice [14]. This calls for new developments in the area of SSL, able to exploit unlabeled information (e.g., by means of *soft* labels) in a more effective way.

The use of *soft* classification labels has been very rarely studied in the context of hyperspectral imaging. In [15], *soft* labels are combined with harmonic energy minimization into an external classifier based on the graph strategy. In [16], posterior marginals are considered as *soft* labels leading to good classification performance. In [2], a novel fuzzy-input fuzzy-output SVM classifier ( $F^2$ SVM) was designed to address subpixel classification problems. The proposed  $F^2$ SVM algorithm uses an input fuzzy membership function to model the subpixel abundances of unknown patterns in the learning process.

In this letter, we develop a new *soft* sparse multinomial logistic regression (SMLR) ( $S^2$ MLR) model which belongs to the family of self-learning algorithms. Compared with the original model in which it is inspired [17], the  $S^2$ MLR makes use of both *hard* and *soft* labels. As opposed to other self-learning methods which expand the training set by using unlabeled samples with *hard* labels only, the proposed SSL algorithm exploits the concept of *soft* labels to generate unlabeled training samples using a recently proposed subspace-based MLR algorithm

(MLR<sub>sub</sub>).<sup>1</sup> This algorithm is specifically designed to address the presence of highly mixed pixels in real hyperspectral images. In addition, the MLR<sub>sub</sub> provides high confidence for the posterior probability estimates for such pixels, which is also a highly desirable feature for our proposed model. The remainder of this letter is organized as follows. Section II introduces the S<sup>2</sup>MLR model in mathematical terms and describes the proposed SSL strategy. Section III evaluates the effectiveness of the proposed method via experiments with real hyperspectral data sets. Quantitative comparisons with classic self-learning algorithms are included for illustrative purposes. Section IV concludes with some remarks and hints at plausible future research lines.

## II. S<sup>2</sup>MLR MODEL

Let  $\mathcal{K} \equiv \{1, \dots, c\}$  denote a set of  $c$  class labels; let  $\mathcal{S} \equiv \{1, \dots, n\}$  be a set of integers indexing the  $n$  pixels of a hyperspectral image; let  $\mathbf{x} \equiv (\mathbf{x}_1, \dots, \mathbf{x}_n)$  be a hyperspectral image of  $d$ -dimensional feature vectors; let  $\mathbf{y} \equiv (\mathbf{y}_1, \dots, \mathbf{y}_n)$  be the output, where  $\mathbf{y}_i = [y_i^{(1)}, \dots, y_i^{(c)}]^T$  denotes a “1-of- $c$ ” encoding of the  $c$  classes,  $\sum_{k=1}^c y_i^{(k)} = 1$ ,  $y_i^{(k)} \in \{0, 1\}$  for *hard* labels, and  $y_i^{(k)} \in [0, 1]$  for *soft* labels; let the sets  $\{(\mathbf{y}_i, \mathbf{x}_i) : i = l_1, \dots, l_n\}$  and  $\{(\mathbf{y}_i, \mathbf{x}_i) : i = u_1, \dots, u_n\}$  respectively denote the labeled and unlabeled training sets, where  $l_n$  and  $u_n$  are the numbers of labeled and unlabeled training samples, respectively. In the following, we describe the classic SMLR model and our proposed variation (S<sup>2</sup>MLR).

### A. SMLR

The MLR [17] models the posterior class probabilities as follows:

$$P\left(y_i^{(k)} = 1 | \mathbf{x}_i, \boldsymbol{\omega}\right) \equiv \frac{\exp\left(\boldsymbol{\omega}^{(k)T} \mathbf{h}(\mathbf{x}_i)\right)}{\sum_{k=1}^K \exp\left(\boldsymbol{\omega}^{(k)T} \mathbf{h}(\mathbf{x}_i)\right)} \quad (1)$$

where  $\boldsymbol{\omega} \equiv [\boldsymbol{\omega}^{(1)T}, \dots, \boldsymbol{\omega}^{(K-1)T}]^T$  denotes the regressors and  $\mathbf{h}(\mathbf{x}_i) \equiv [h_1(\mathbf{x}_i), \dots, h_l(\mathbf{x}_i)]^T$  is a vector of  $l$  fixed functions of the input, often termed *features*. In this letter, we use the Gaussian radial basis function kernel:  $K(\mathbf{x}_i, \mathbf{x}_j) \equiv \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\sigma^2)$ , which is widely used in hyperspectral image classification [19]. Following the SMLR algorithm introduced in [20], we model  $\boldsymbol{\omega}$  as a random vector with Laplacian density  $p(\boldsymbol{\omega}) \propto \exp(-\lambda \|\boldsymbol{\omega}\|_1)$ , where  $\lambda$  is the regularization parameter controlling the degree of sparsity of  $\boldsymbol{\omega}$ . Under the present setting, learning the class densities amounts to estimating the logistic regressors  $\boldsymbol{\omega}$  [13], [20], [21]. By adopting the maximum *a posteriori* (MAP) estimation criterion, we have

$$\hat{\boldsymbol{\omega}}_{\text{MAP}} = \arg \max_{\boldsymbol{\omega}} \ell_l(\boldsymbol{\omega}) + \log p(\boldsymbol{\omega}) \quad (2)$$

where  $\ell_l(\boldsymbol{\omega})$  is the log-likelihood function on the labeled information and

$$\ell_l(\boldsymbol{\omega}) = \sum_{i=l_1}^{l_n} \left( \sum_{k=1}^c y_i^{(k)} \boldsymbol{\omega}^{(k)T} \mathbf{h}(\mathbf{x}_i) - \log \sum_{k=1}^c \exp\left(\boldsymbol{\omega}^{(k)T} \mathbf{h}(\mathbf{x}_i)\right) \right). \quad (3)$$

Notice that  $y_i^{(k)}$  denotes *hard* labels, i.e.,  $y_i^{(k)} \in \{0, 1\}$ .

<sup>1</sup>More details about the MLR<sub>sub</sub> algorithm can be found in [18]. The source code for this algorithm is available online: [http://www.lx.it.pt/~jun/mlrsub\\_demo.zip](http://www.lx.it.pt/~jun/mlrsub_demo.zip).

### B. S<sup>2</sup>MLR

Since our approach is semisupervised, the classifier is learnt from both the labeled and the unlabeled data. Based on the supervised optimization problem in (2), we heuristically obtain the semisupervised MAP estimation of  $\boldsymbol{\omega}$  by introducing the unlabeled information

$$\hat{\boldsymbol{\omega}}_{\text{MAP}} = \arg \max_{\boldsymbol{\omega}} \ell_l(\boldsymbol{\omega}) + \ell_u(\boldsymbol{\omega}) + \log p(\boldsymbol{\omega}) \quad (4)$$

where  $\ell_u(\boldsymbol{\omega})$  has the same structure of  $\ell_l(\boldsymbol{\omega})$  given by (3) and

$$\ell_u(\boldsymbol{\omega}) = \sum_{i=u_1}^{u_n} \left( \sum_{k=1}^c \hat{y}_i^{(k)} \boldsymbol{\omega}^{(k)T} \mathbf{h}(\mathbf{x}_i) - \log \sum_{k=1}^c \exp\left(\boldsymbol{\omega}^{(k)T} \mathbf{h}(\mathbf{x}_i)\right) \right) \quad (5)$$

where, now,  $\hat{y}_i^{(k)}$  denotes the *soft* labels in contrast with the *hard* labels. In this letter, the *soft* labels  $\hat{y}_i^{(k)}$  are heuristically replaced by the probabilities given by the MLR<sub>sub</sub> algorithm [18], i.e.,  $\hat{y}_i^{(k)} \equiv E[y_i^{(k)} | \mathbf{x}_i, \hat{\boldsymbol{\theta}}] = p(y_i^{(k)} = 1 | \mathbf{x}_i, \hat{\boldsymbol{\theta}})$ , where  $\hat{\boldsymbol{\theta}}$  is learnt from the MLR<sub>sub</sub> algorithm. As shown in [18], the MLR<sub>sub</sub> is well suited to deal highly mixed analysis scenarios, i.e., dominated by pixel vectors with materials appearing in more than one class and very difficult to separate. The MLR<sub>sub</sub> provides reliable probabilities that are ideal to be used as *soft* labels in our problem. However, it should be noticed that any other estimates can be adopted, as long as the *soft* labels are reliable. For instance, in [16], the posterior marginals are considered, whereas in [22], the fractional abundances obtained from linear spectral unmixing are used, providing good performance.

As an SSL technique, the proposed approach generally deals with a very low number of labeled samples. This leads to difficult classification problems regardless of whether a supervised or an SSL technique is adopted. In the former case, poor generalization capability is expected. In the latter case, the tradeoff between a large number of unlabeled samples versus a small number of labeled samples could bias the learning process. This is also an important concern since, in a self-learning algorithm, the *hard/soft* labels are inferred from the labeled training set. Therefore, in this letter, we use an iterative scheme to increment the number of unlabeled samples. Let  $u_t$  be the number of new unlabeled samples included at each iteration. In general, we take  $u_t < l_n$ . This setting provides a reasonable balance between the labeled and unlabeled information. At the same time, it has the advantage of iteratively refining the learning process in case of poor generalization.

At this point, it should be noted that our adopted procedure is iterative, i.e., we select a number of unlabeled samples which are less than the total number of labeled samples at each iteration, but after several iterations, we have more unlabeled samples than labeled samples. In other words, in our approach, we start with a very small number of labeled samples and then grow the number of unlabeled samples until we end up with a much larger number of unlabeled samples as compared to the number of initial labeled samples, i.e.,  $l_n \ll u_n$ . However, under the present setup, a relevant question is “which  $u_t$  samples should be chosen.” In this letter, we generate unlabeled training samples based on a first-order neighborhood system, where the

$u_t$  unlabeled samples are selected according to the following criterion. Let us consider the first-order neighborhood system of  $D_{l+u}$ , which denotes the set made up of both labeled and unlabeled training samples. If both the MLR and the MLR<sub>sub</sub> classifiers provide the same estimate  $\hat{y}_i$  for a given pixel  $\mathbf{x}_i$ , then  $\{\hat{y}_i, \mathbf{x}_i\}$  is selected as a new unlabeled sample. Here, the MLR<sub>sub</sub> classifier combines MLR with a subspace projection method that better characterizes noise and mixed pixels [18].

We emphasize that the considered optimization problems (2) and (4) can be solved by SMLR [20] and by fast SMLR (FSMLR) [23]. Although the original SMLR algorithm gives very good results, it is however limited to data sets with products  $(l + u) \times c$  not larger than, for example, 1000, whereas the FSMLR is unbearable when the number of training samples increases. Therefore, most hyperspectral data sets are beyond the reach of these algorithms, particularly in SSL which tries to use a large amount of unlabeled training samples. This difficulty has been recently addressed by the introduction of the logistic regression via variable splitting and augmented Lagrangian (LORSAL) algorithm [24], which is able to deal with training sets with a few thousand training samples, regardless of the number of classes. LORSAL plays a central role, for example, in [13]. In this letter, following our previous work, we resort to the LORSAL algorithm to estimate the regressors for the proposed SSL algorithm. The computational complexity of the proposed S<sup>2</sup>MLR using LORSAL algorithm is  $O(c(l_n + u_n)^2)$ . It is also important to note that the computational complexity only depends on the total number of labeled and unlabeled trainings; there is no difference between using soft and using hard labels for training purposes.

Algorithm 1 shows a pseudocode for the proposed S<sup>2</sup>MLR algorithm. In lines 2–3 of Algorithm 1, *soft* labels are estimated from the MLR<sub>sub</sub> algorithm. Here,  $\tau$  controls the loss of spectral information after projecting the data into a subspace. Notice that  $\hat{\theta}$  is also iteratively updated by taking advantage of unlabeled training samples. Line 4 of Algorithm 1 estimates the regressor using the LORSAL algorithm. In line 5, function  $\psi(\cdot)$  generates the unlabeled training set  $D_{u_t}$  based on the aforementioned criterion. In our experiments, we set  $\tau = 0.95$  and  $\lambda = 0.001$ . These are suboptimal settings; however, we have empirically found that these parameters provide very good performance [18].

---

#### Algorithm 1 S<sup>2</sup>MLR

**Require:**  $\mathcal{D}_{l+u}, \mathbf{x}, \lambda, \tau$

1: **repeat**

2:  $\hat{\theta} := \text{MLR}_{\text{sub}}(\mathcal{D}_{l+u}, \tau)$

3:  $\hat{y} := E[\mathbf{y}|\mathbf{x}, \hat{\theta}]$

4:  $\hat{\omega} := \text{LORSAL}(\mathcal{D}_{l+u}, \hat{y}, \lambda)$

5:  $D_{u_t} := \psi(\mathcal{D}_{l+u}, \mathbf{x}, \hat{\theta}, \hat{\omega})$

6:  $\mathcal{D}_{l+u} := \mathcal{D}_{l+u} + D_{u_t}$

7: **until** some stopping criterion is met

---

### III. EXPERIMENTAL RESULTS

In order to evaluate the proposed SSL algorithm, we conduct experiments with three widely used hyperspectral images col-

lected by the National Aeronautics and Space Administration's Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) [25] and by the Reflective Optics System Imaging Spectrometer (ROSIS) operated by the German Aerospace Agency (DLR). In all cases, we illustrate the performance achieved by the proposed algorithm with a conventional self-learning algorithm [26] in which *hard* labels are given by the obtained classification results based on the aforementioned sampling criterion. This is reasonable for generating *hard* labels, since, in hyperspectral images, it is very likely that two neighboring pixels belong to the same class. In order to have a fair comparison, the same unlabeled feature vectors are used for the proposed S<sup>2</sup>MLR algorithm along with *soft* labels estimated by the MLR<sub>sub</sub>. In our experiments, the labeled training samples are randomly selected from the available ground-truth data, whereas the remaining samples are used for validation. In order to increase the statistical significance of the results, each value of overall accuracy (OA) and kappa statistic ( $\kappa$ ) reported in this letter is obtained as the average of ten Monte Carlo (MC) runs.

#### A. Experiment 1: AVIRIS Kennedy Space Center

In our first experiment, we use an AVIRIS hyperspectral data set collected over the Kennedy Space Center,<sup>2</sup> Florida, in March 1996. The portion of this scene used in our experiments has dimensions of  $292 \times 383$  pixels. After removing water absorption and noisy bands, 176 bands were used for the analysis. The spatial resolution is 20 m by pixel. Twelve ground-truth classes were available, where the number of pixels in the smallest class is 105 and the number of pixels in the largest class is 761. In order to show the good capacity of the proposed SSL algorithm in dealing with ill-posed problems, a very limited number of 36 labeled samples (three per class) are used for training purposes.

Table I shows the classification results obtained for each of the ten conducted MC runs. The table reveals that the classification results are improved significantly by using unlabeled samples with regard to the supervised case, in which only labeled samples are used. Even in those cases with poor generalization (e.g., see MC4), the proposed SSL algorithm still significantly improves the obtained classification accuracies. It is also noticeable that the average OA and  $\kappa$  obtained by using *soft* labels are slightly better than the scores obtained using *hard* labels. Although the improvements in classification accuracy are not very significant, a detailed look into each MC run reveals that the *soft* classifier always outperforms the *hard* classifier (in our experiments, only MC2 was an exception to this statement). This is because the use of *soft* labels allows the proposed S<sup>2</sup>MLR model to tackle mixed pixels in conjunction with labeled samples, thus better optimizing the estimation of class boundaries and associating together all of the samples sharing common properties in the same region while preserving the relevance of labeled samples in the process. This allows for a better balancing in the joint exploitation of labeled and unlabeled samples during the classification process.

#### B. Experiment 2: AVIRIS Indian Pines

The well-known AVIRIS Indian Pines scene was used in our second experiment. The data were collected over northwestern

<sup>2</sup>Available online: <http://www.csr.utexas.edu/hyperspectral/data/KSC/>.

TABLE I  
OA (IN PERCENT) AND  $\kappa$  (IN PERCENT) RESULTS FOR TEN MC RUNS AFTER PROCESSING THE AVIRIS KENNEDY SPACE CENTER HYPERSPECTRAL IMAGE USING THE PROPOSED SSL WITH 36 LABELED (THREE PER CLASS) AND 1226 UNLABELED TRAINING SAMPLES. FOR ILLUSTRATIVE PURPOSES, THE RESULTS OBTAINED FOR THE SUPERVISED CASE (USING ONLY LABELED SAMPLES) ARE ALSO DISPLAYED

		Monte Carlo (MC) runs										Average	
		MC1	MC2	MC3	MC4	MC5	MC6	MC7	MC8	MC9	MC10		
Supervised	OA	70.22	67.75	66.99	59.44	69.03	65.46	64.78	68.32	60.44	70.50	66.29	
	$\kappa$	66.41	64.14	62.96	55.06	65.05	61.56	60.36	64.52	56.23	66.69	62.30	
SSL	<i>soft</i>	OA	73.43	75.55	78.08	74.76	79.67	77.03	75.04	71.47	75.67	74.73	75.54
		$\kappa$	69.95	72.64	75.22	71.58	77.03	74.31	71.59	68.21	72.76	71.50	72.48
	<i>hard</i>	OA	72.55	75.64	77.37	74.25	78.70	76.29	73.96	71.13	74.76	73.82	74.85
		$\kappa$	68.98	72.69	74.43	71.02	75.95	73.50	70.38	67.82	71.78	70.49	71.70

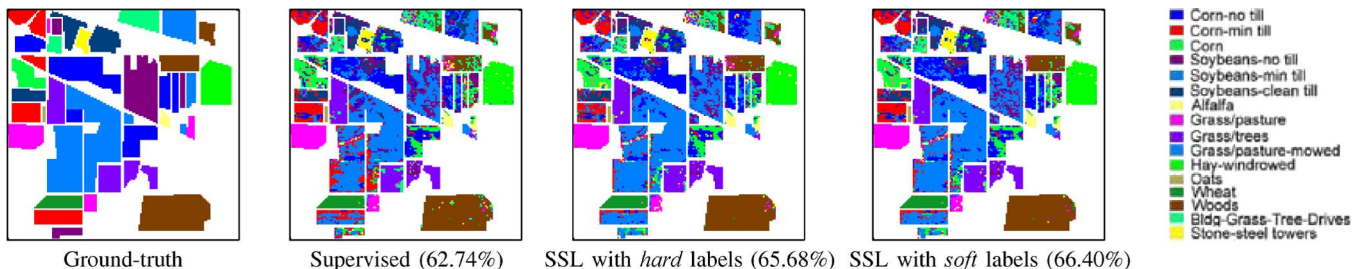


Fig. 1. Classification maps obtained for the AVIRIS Indian Pines image using 160 labeled samples (ten per ground-truth class) and 849 unlabeled samples, along with the obtained OAs.

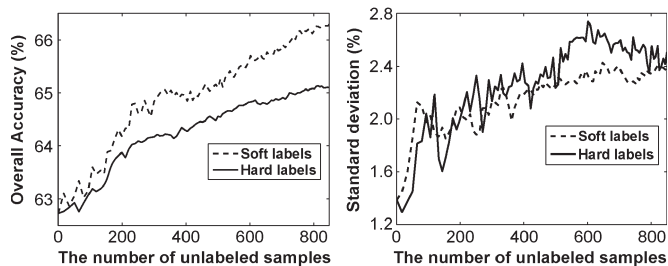


Fig. 2. OA (in percent) classification results and standard deviations (after ten MC runs) as functions of the number of unlabeled samples obtained for the AVIRIS Indian Pines image, using the proposed SSL algorithm with 160 labeled samples (ten per ground-truth class).

Indiana in June of 1992 [1] and contain  $145 \times 145$  pixels and 220 spectral bands. A total of 20 bands were removed prior to experiments due to noise and water absorption in those channels. The ground-truth data [shown in Fig. 1(a)] contain 16 mutually exclusive classes [with the class legends in Fig. 1(b)] and a total of 10 366 labeled pixels.<sup>3</sup> This image is a classic benchmark to validate the accuracy of hyperspectral image analysis algorithms and constitutes a challenging problem due to the significant presence of mixed pixels in all available classes and also because of the unbalanced number of available labeled pixels per class.

Fig. 2 reports the classification accuracies and standard deviations (after ten MC runs) obtained by the proposed SSL approach as functions of the number of unlabeled samples. It can be observed that very good results are obtained by using both *soft* and *hard* labels, where the classification results always increase as the number of unlabeled samples is increased (recall that these samples can be obtained at no cost in our proposed

framework). For instance, with 160 labeled samples (and no unlabeled samples), the supervised algorithm obtained an OA of 62.72% and a  $\kappa$  of 58.23%. However, by including 849 unlabeled samples (which come at very low cost), the proposed SSL algorithm increased the OA to 66.29% (with  $\kappa$  of 62.19%, using soft labels) and to 65.10% (with  $\kappa$  of 60.89%, using hard labels). Furthermore, the improvements obtained by using *soft* labels with regard to using *hard* labels can also be appreciated in Fig. 2, in which a significant improvement resulting from using *soft* labels can be already appreciated around 200 unlabeled samples. Such improvement becomes more relevant with an increasing number of unlabeled samples. Fig. 2 also shows that soft labels exhibit less standard deviation and produce more robust results as compared to hard labels. For illustrative purposes, Fig. 1(c)–(e) shows the classification maps obtained (in one of the MC runs) by the supervised framework, the SSL framework with hard labels, and the SSL framework with soft labels.

### C. Experiment 3: ROSIS University of Pavia

In our third experiment, we considered a hyperspectral image acquired in 2001 by the ROSIS instrument over the city of Pavia, Italy. The image scene, with a size of  $610 \times 340$  pixels, is centered at the University of Pavia. After removing 12 bands due to noise and water absorption, it comprises 103 spectral channels. Nine ground-truth classes, with a total of 3921 training samples and 42 776 test samples, were considered in experiments. Table II shows the classification results obtained by the proposed SSL algorithm in two different scenarios:  $l_n = 1800$  and  $u_n = 400$  and  $l_n = 2700$  and  $u_n = 600$ . In both cases, the proposed SSL algorithm with *soft* labels provides better results in terms of OA and  $\kappa$  when compared to the other tested methods.

<sup>3</sup>Available online: <http://dynamo.ecn.purdue.edu/biehl/MultiSpec>.

TABLE II  
OA (IN PERCENT) AND  $\kappa$  (IN PERCENT) CLASSIFICATION RESULTS  
OBTAINED BY THE PROPOSED SSL ALGORITHM  
FOR THE ROSIS DATA SET

$l_n = 1800, u_n = 400$			
	Supervised	SSL ( <i>soft</i> labels)	SSL ( <i>hard</i> labels)
OA	76.68	78.64	77.34
$\kappa$	71.20	73.28	71.93
$l_n = 2700, u_n = 600$			
	Supervised	SSL ( <i>soft</i> labels)	SSL ( <i>hard</i> labels)
OA	78.22	81.18	78.45
$\kappa$	72.97	76.22	73.19

#### IV. CONCLUSION AND FUTURE LINES

In this letter, we have developed a new S<sup>2</sup>MLR model which uses both *hard* and *soft* labels as opposed to other SSL algorithms which generally assign hard labels only when deriving unlabeled training samples. Our proposed strategy allows us to better model the phenomenon of mixed pixels present in hyperspectral images by the inclusion of soft labels. For this purpose, we use the posterior probabilities obtained by a recently proposed subspace-based multinomial logistic regression algorithm (MLR<sub>sub</sub>) as soft labels, mainly because these probabilities exhibit high confidence in the estimates provided for mixed pixels. The obtained classification accuracies for the proposed method have been evaluated via experiments with three different hyperspectral scenes, achieving state-of-the-art performance with very limited training samples. In the future, additional strategies for the generation of soft labels (e.g., obtaining the label estimates from the whole image) will be used to fully substantiate our findings. We will also target additional mechanisms for exploiting the unlabeled information, e.g., by means of active learning. There has been some work in the past concerning possibilistic *k*-nearest neighbors for land mine detection with classifiers [27] and matching pursuits [28] that has similarities to the proposed approach and may prove to be a robust approach that could also be placed into a Bayesian framework in future developments.

#### REFERENCES

- [1] D. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*. Hoboken, NJ: Wiley, 2003.
- [2] F. Bovolo, L. Bruzzone, and L. Carline, "A novel technique for subpixel image classification based on support vector machine," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2983–2999, Nov. 2010.
- [3] B. Shahshahani and D. A. Landgrebe, "The effect of unlabeled samples in reducing the small sample size problem and mitigating the Hughes phenomenon," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 5, pp. 1087–1095, Sep. 1994.
- [4] L. Bruzzone, M. Chi, and M. Marconcini, "A novel transductive SVM for the semisupervised classification of remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 11, pp. 3363–3373, Nov. 2006.
- [5] A. Blum and T. Mitchell, *Combining Labeled and Unlabeled Data With Co-Training*. San Mateo, CA: Morgan Kaufmann, 1998, pp. 92–100.
- [6] D. Yarowsky, "Unsupervised word sense disambiguation rivaling supervised methods," in *Proc. 33rd Annu. Meeting Assoc. Comput. Linguistics*, 1995, pp. 189–196.
- [7] U. Brefeld, T. Gärtner, T. Scheffer, and S. Wrobel, "Efficient co-regularised least squares regression," in *Proc. 23rd ICML Conf.*, 2006, pp. 137–144.
- [8] V. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.
- [9] A. Blum and S. Chawla, "Learning from labeled and unlabeled data using graph mincuts," in *Proc. 18th ICML Conf.*, 2001, pp. 19–26.
- [10] X. Zhu, "Semi-supervised learning literature survey," *Comput. Sci., Univ. Wisconsin-Madison, Madison, WI, Tech. Rep. 1530*, 2005.
- [11] G. Camps-Valls, T. Bandos, and D. Zhou, "Semi-supervised graph-based hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3044–3054, Oct. 2007.
- [12] A. Plaza, J. Benediktsson, J. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, M. Marconcini, J. Tilton, and G. Trianni, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, no. 1, pp. 110–122, Sep. 2009.
- [13] J. Li, J. Bioucas-Dias, and A. Plaza, "Semi-supervised hyperspectral image segmentation using multinomial logistic regression with active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4085–4098, Nov. 2010.
- [14] N. Keshava and J. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 44–57, Jan. 2002.
- [15] X. Zhu, Z. Ghahramani, and J. Lafferty, "Semi-supervised learning using Gaussian fields and harmonic functions," in *Proc. 20th ICML Conf.*, 2003, pp. 912–919.
- [16] J. Li, J. Bioucas-Dias, and A. Plaza, "Exploiting spatial information in semi-supervised hyperspectral image segmentation," in *Proc. IEEE WHISPERS*, 2010, pp. 1–4.
- [17] D. Böhning, "Multinomial logistic regression algorithm," *Ann. Inst. Stat. Math.*, vol. 44, no. 1, pp. 197–200, Mar. 1992.
- [18] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Spectral-spatial hyperspectral image segmentation using subspace multinomial logistic regression and Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 809–823, Mar. 2012.
- [19] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 1351–1362, Jun. 2005.
- [20] B. Krishnapuram, L. Carin, M. Figueiredo, and A. Hartemink, "Sparse multinomial logistic regression: Fast algorithms and generalization bounds," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 6, pp. 957–968, Jun. 2005.
- [21] J. Li, J. Bioucas-Dias, and A. Plaza, "Hyperspectral image segmentation using a new Bayesian approach with active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3947–3960, Oct. 2010.
- [22] A. Villa, J. Li, A. Plaza, and J. Bioucas-Dias, "A new semi-supervised algorithm for hyperspectral image classification based on spectral unmixing concepts," in *Proc. IEEE WHISPERS*, Jun. 2011, pp. 1–4.
- [23] J. Borges, J. Bioucas-Dias, and A. Marçal, "Bayesian hyperspectral image segmentation with discriminative class learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2151–2164, Jun. 2011.
- [24] J. Bioucas-Dias and M. Figueiredo, "Logistic Regression via Variable Splitting and Augmented Lagrangian Tools," Instituto Superior Técnico, Lisbon, Portugal, Tech. Rep., 2009.
- [25] R. O. Green, M. L. Eastwood, C. M. Sarture, T. G. Chrien, M. Aronsson, B. J. Chippendale, J. A. Faust, B. E. Pavri, C. J. Chovit, M. Solis, and M. R. Olah, "Imaging spectroscopy and the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)," *Remote Sens. Environ.*, vol. 65, no. 3, pp. 227–248, Jul. 1998.
- [26] J. Li, J. Bioucas-Dias, and A. Plaza, "Semi-supervised hyperspectral image classification based on a Markov random field and sparse multinomial logistic regression," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2009, vol. 3, pp. 817–820.
- [27] H. Frigui and P. D. Gader, "Detection and discrimination of land mines in ground-penetrating radar based on edge histogram descriptors and a possibilistic *k*-nearest neighbor classifier," *IEEE Trans. Fuzzy Syst.*, vol. 17, no. 1, pp. 185–199, Feb. 2009.
- [28] R. Mazhar, P. Gader, and J. Wilson, "Matching-pursuits dissimilarity measure for shape-based comparison and classification of high-dimensional data," *IEEE Trans. Fuzzy Syst.*, vol. 17, no. 5, pp. 1175–1188, Oct. 2009.