A Subspace-Based Multinomial Logistic Regression for Hyperspectral Image Classification

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

***Abstract*—In this letter, we propose a multinomial-logistic-regression method for pixelwise hyperspectral classification. The feature vectors are formed by the energy of the spectral vectors projected on class-indexed subspaces. In this way, we model not only the linear mixing process that is often present in the hyperspectral measurement process but also the nonlinearities that are separable in the feature space defined by the aforementioned feature vectors. Our experimental results have been conducted using both simulated and real hyperspectral data sets, which are collected using NASA's Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) and the Reflective Optics System Imaging Spectrographic (ROSIS) system. These results indicate that the proposed method provides competitive results in comparison with other state-of-the-art approaches.**

***Index Terms*—Hyperspectral imaging, pixelwise classification, subspace multinomial logistic regression (MLR).**

I. Introduction

HYPERSPECTRAL sensors provide images in hundreds of continuous (narrow) spectral bands that can be used to discriminate different objects on the earth surface [1]. Recently, multinomial logistic regression (MLR) has shown good performance in hyperspectral image classification. MLR is a discriminative approach that directly models posterior class distributions [2]–[5]. Recent examples on the use of MLR in hyperspectral classification problems can be found in [6]–[9]. In this type of classifiers, we highlight the MLRsub method [5] that was specifically designed with the linear spectral mixing process in mind. In the MLRsub method, the classification of a pixel (with its associated spectral vector in a given class) corresponds to the largest projection of that vector onto the class-indexed subspaces. In this letter, and in order to model possible nonlinear mixing effects, we allow the MLR regression vectors to define arbitrary linear combinations of the projections of the subspaces learned from the training set. In comparison with the work in [5], which originally proposed MLRsub, the proposed subspace-based MLR ($\text{MLRsub}_{\text{mod}}$) introduces two main contributions.

- First, the newly developed method uses the projection of the original spectral vectors onto class-dependent subspaces in order to enhance class separability. At this point, we can mention two main reasons that support the use of these projections. One reason is that the hyperspectral measurements are often linear mixtures of the endmembers' signatures, and then, each class corresponds to a given subset of the endmembers, thus defining a subspace. The other reason is that we claim that a relevant number of nonlinear phenomena present in the hyperspectral image yields spectral vectors that are linearly separable in the feature space defined by the class-indexed subspaces.
- Second, the proposed method consists of including the class prior probabilities in the proposed model. This is expected to introduce advantages in scenarios in which the number of training samples per class depends on the area covered by that particular class in the scene.

The remainder of this letter is structured as follows. Section II describes the newly proposed subspace-based MLR method. Section III presents experimental results using both simulated and real hyperspectral scenes. The numerical results illustrate that the performance of the MLRsub classification algorithm can be significantly improved by using the proposed subspace-based projection feature vectors, and it incorporates the prior information from the known class proportions. Finally, Section IV concludes this letter with some remarks.

II. Class-Dependent Subspace-Based MLR ($\text{MLRsub}_{\text{mod}}$)

Let $\mathbf{x} \equiv \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ be the input hyperspectral image, where n is the number of pixels in $\mathbf{x}$; $\mathbf{x}_i \in \mathbb{R}^d$ denotes a spectral vector associated with an image pixel i, and d is the number of spectral bands. Let $\mathbf{y} \equiv (y_1, \ldots, y_n)$ denote an image of class labels, and $y_i \in \{1, \ldots, K\}$, where K is the number of classes. In [5], it was shown that the posterior class density $p(y_i|\mathbf{x}_i)$ can be computed in the MLR framework by using the nonlinear functions $\boldsymbol{\phi}^{(k)}(\mathbf{x}_i) = [\|\mathbf{x}_i\|^2, \|\mathbf{x}_i^T\mathbf{U}^{(k)}\|^2]^T$, where $\mathbf{U}^{(k)}$ is a set of $r^{(k)}$-dimensional orthonormal-basis vectors for the subspace associated with classes $k = 1, 2, \ldots, K$. Following [5], in this letter, $\mathbf{U}^{(k)}$ is computed as $\mathbf{U}^{(k)} = \{\mathbf{e}_1^{(k)}, \ldots, \mathbf{e}_{r^{(k)}}^{(k)}\}$, whereas $\mathbf{E}^{(k)} = \{\mathbf{e}_1^{(k)}, \ldots, \mathbf{e}_d^{(k)}\}$ is the eigenvector matrix computed from correlation matrix $\mathbf{R}^{(k)} = \mathbf{E}^{(k)}\boldsymbol{\Lambda}^{(k)}\mathbf{E}^{(k)^T}$. Here, $\boldsymbol{\Lambda}$ is the eigenvalue matrix with decreasing magnitude.

Manuscript received July 18, 2013; revised March 8, 2014; accepted April 4, 2014. Date of publication May 29, 2014; date of current version June 20, 2014. This work was supported by the Portuguese Science and Technology Foundation under Projects PEst-OE/EEI/LA0008/2013 and PTDC/EEI-PRO/1470/2012. (*Corresponding author: J. Li.*)

M. Khodadadzadeh and A. Plaza are with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, 10071 Cáceres, Spain.

J. Li is with the School of Geography and Planning, Sun Yat-Sen University, Guangzhou 510275, China (e-mail: jun@lx.it.pt).

J. M. Bioucas-Dias is with Instituto de Telecomunicações and Instituto Superior Técnico, Universidade de Lisboa, 1649-004 Lisboa, Portugal.

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.2014.2320258

In this letter, following [5], we use a subspace projection accounting for 99.9% of the original spectral information in order to determine the size of $\mathbf{U}^{(k)}$. As shown in [5], the MLRsub method aims to deal with the problems defined by the linear mixing model. However, nonlinear mixing is very common in real scenarios. We claim that a number of nonlinearities present in the hyperspectral mixing process are, approximately, linearly separable in the feature space defined by nonlinear functions $\boldsymbol{\phi}(\mathbf{x}_i) = [\|\mathbf{x}_i\|^2, \|\mathbf{x}_i^T\mathbf{U}^{(1)}\|^2, \ldots, \|\mathbf{x}_i^T\mathbf{U}^{(K)}\|^2]^T$, i.e., the vector features containing as components the energy of the projections on all class subspaces plus the energy of the original vector. This claim will be supported in Section III with the experimental results.

In this letter, we use a nonlinear vector of regression functions $\boldsymbol{\phi}(\mathbf{x}_i) = [\|\mathbf{x}_i\|^2, \|\mathbf{x}_i^T\mathbf{U}^{(1)}\|^2, \ldots, \|\mathbf{x}_i^T\mathbf{U}^{(K)}\|^2]^T$ to compute the posterior class density $p(y_i|\mathbf{x}_i)$ for a given class k as follows:

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

where, by assuming $p(y_i = k) = 1/K$, we exactly have an MLR classifier. However, in order to introduce the available prior knowledge, here, we include the estimation of the occurrence probabilities of each land-cover class from the training set. Let $N_k \in \{N_1, \ldots, N_K\}$ be the number of training samples for class k. The prior probability for class k may be computed as $p(y_i = k) = N_k / \sum_{l=1}^{K} N_l$ [10].

Notice that, if the data live in a class-dependent subspace defined by the linear mixing model, the proposed approach (1) can be recovered by the conventional MLRsub in [5] by a setting of regressing parameters $\boldsymbol{\omega}^{(k)} = [\omega_1, 0, \ldots, 0, \omega_{k+1}, 0, \ldots, 0]^T$. Another important aspect is that, if the data do not strictly live in a linear subspace and follow a nonlinear mixing model (which is a quite common scenario in practice), then as supported in Section III, the proposed $\text{MLRsub}_{\text{mod}}$ approach is able to separate the classes in the newly proposed feature space. Therefore, the proposed approach has the ability to handle both linear and nonlinear mixtures, which is the main contribution of this letter. However, further work should be conducted in order to fully analyze how the assumed dependence between the classes handles the nonlinearity of the mixtures.

Under the present setup, we compute $\boldsymbol{\omega}$ in (1) by calculating the maximum *a posteriori* estimate as follows:

$$\widehat{\boldsymbol{\omega}} = \arg\max_{\boldsymbol{\omega}} \quad \ell(\boldsymbol{\omega}) + \log p(\boldsymbol{\omega}) \quad (2)$$

where $\ell(\boldsymbol{\omega}) \equiv \log \prod_{i=1}^{N} p(y_i|\mathbf{x}_i, \boldsymbol{\omega})$ is the log-likelihood function ($N = \sum_{l=1}^{K} N_l$ is the total number of samples in the training set). Similar to the MLRsub algorithm in [5], $p(\boldsymbol{\omega}) \propto e^{-\beta/2\|\boldsymbol{\omega}\|^2}$ ($\beta \geq 0$ is a regularization parameter controlling the weight of the prior) is a quadratic prior on $\boldsymbol{\omega}$ that is intended to cope with difficulties in learning regression vector $\boldsymbol{\omega}$ associated with bad or ill conditioning of the underlying inverse problem.

The optimization problem in (2) is convex, although term $\ell(\boldsymbol{\omega})$ is nonquadratic. Following the previous work in [5], [11], and [12], we approximate this term by a quadratic lower bound, which leads to a sequence of quadratic problems that are easier to solve than the original problem.

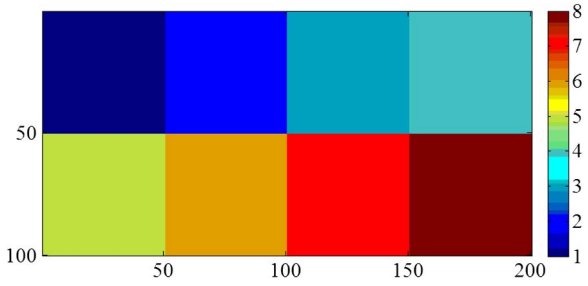

Fig. 1. Simulated hyperspectral image comprising eight different classes.

III. Experimental Results and Discussion

In this section, we evaluate the proposed class-subspace-based MLR by using both simulated and real hyperspectral images. For the parameter settings, we follow the indications given in [5] and include a comparison with the MLRsub in [5]. It should be noted that, in this letter, we only compare our $\text{MLRsub}_{\text{mod}}$ with MLRsub. The main reason is that the proposed subspace-based features yield better performance than those used in MLRsub. Another reason is that, in [5], there is already a comprehensive comparison with state-of-the-art methods.

A. Experiments With Simulated Data

In order to have an assessment in a fully controlled environment, we first used a simulated data set to evaluate the capability of the proposed approach for handling nonlinear mixtures. For this purpose, we generated a synthetic image with 50×50 samples for each of the eight classes simulated (see Fig. 1). We considered the following nonlinear mixture model for generating each simulated mixed pixel in class k:

$$\mathbf{x}_i^{(k)} = \sum_{j=0}^{M_l} \mathbf{m}^{(k+j)}\gamma_j + \alpha \prod_{j=0}^{M_{nl}} \mathbf{m}^{(k+j)} + \mathbf{n}_i \quad (3)$$

where $\mathbf{m}^{(l)}$, $l = 1, \ldots, 10$, are different spectral signatures that were randomly selected from the U.S. Geological Survey digital spectral library,[1] γ_j and α are the parameters controlling the impact of the linear and nonlinear terms, respectively, and $\sum_{j=0}^{M_l} \gamma_j = 1 - \alpha$. In our simulation, γ_0 is the abundance of the objective class, i.e.,. the class that received the maximum abundance value in the simulation and that will define the label for the considered class. To have a comprehensive comparison using both linear and nonlinear mixtures, we used $\alpha = 0$ for classes {1, 3, 5, 7}, which means that these four classes stay in a linear subspace, and we included nonlinear mixtures for classes {2, 4, 6, 8}. Furthermore, for each pixel, we randomly chose a value over {1, 2} for parameters M_l and M_{nl}, which means that we set the number of mixtures in each pixel to 2 or 3. Pure spectral signatures are considered for the first four

[1] http://speclab.cr.usgs.gov/spectral-lib.html

TABLE I
OVERALL CLASSIFICATION ACCURACY (IN PERCENTAGE) FOR THE DIFFERENT VALUES OF PARAMETER α (WITH THE NOISE STANDARD DEVIATION SET TO $\sigma = 0.4$) OBTAINED BY THE MLRsub AND MLRsub$_{\mathrm{mod}}$ METHODS FOR THE SIMULATED DATA SET IN FIG. 1

Simulated classes			$\alpha = 0$		$\alpha = 0.3$		$\alpha = 0.4$		$\alpha = 0.5$	
Class	Mixtures	Pure pixels	MLR*sub*	MLRsub_{mod}	MLR*sub*	MLRsub_{mod}	MLR*sub*	MLRsub_{mod}	MLR*sub*	MLRsub_{mod}
C1	Linear	Yes	79.78	86.85	75.77	81.06	75.37	79.43	72.08	74.71
C2	Nonlinear	Yes	74.70	83.80	75.59	82.83	74.56	80.15	71.82	77.31
C3	Linear	Yes	75.92	86.34	73.26	79.31	69.09	74.78	66.15	70.26
C4	Nonlinear	Yes	76.17	85.34	77.24	84.30	75.01	81.67	72.65	79.39
C5	Linear	No	75.41	83.91	70.34	78.15	65.58	71.97	60.58	66.13
C6	Nonlinear	No	74.86	83.77	69.25	79.05	67.66	76.33	65.44	73.92
C7	Linear	No	74.19	82.51	66.53	75.48	62.84	71.12	58.18	65.06
C8	Nonlinear	No	79.58	86.23	77.19	83.59	74.04	80.23	71.38	77.28
OA{1,2,3,4}	-	Yes	76.64	85.58	75.46	81.87	73.50	79.01	70.67	75.42
OA{5,6,7,8}	-	No	76.01	84.10	70.83	79.07	67.53	74.91	63.90	70.60
OA{1,3,5,7}	Linear	-	76.32	84.90	71.48	78.50	68.22	74.33	64.25	69.04
OA{2,4,6,8}	Nonlinear	-	76.33	84.78	74.82	82.44	72.82	79.59	70.32	76.97
Overall accuracy (OA)			76.33	84.84	73.15	80.47	70.52	76.96	67.28	73.01

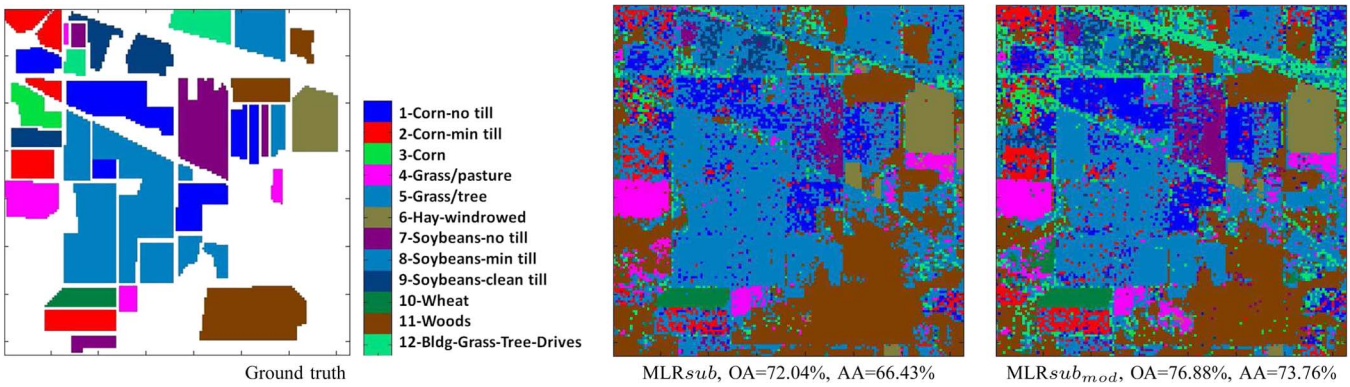

Fig. 2. Ground truth and classification results for the AVIRIS Indian Pines data set using 1076 training samples.

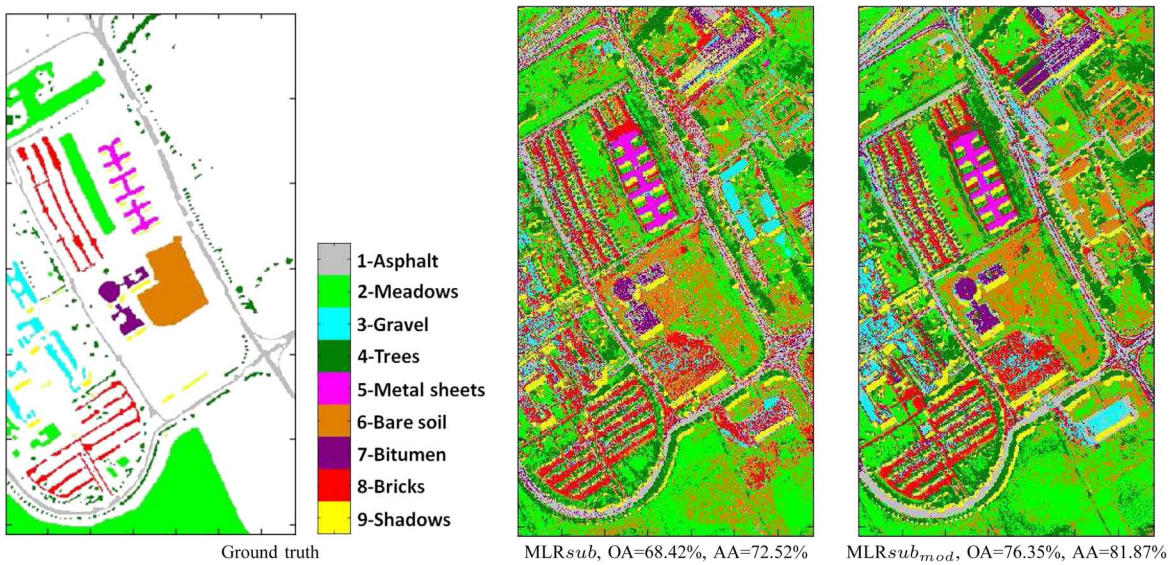

Fig. 3. Ground truth and classification results obtained for the ROSIS Pavia University data set with 781 training samples.

classes, whereas the remaining classes are entirely made up of mixed pixels (see Table I). Finally, zero-mean Gaussian noise with covariance $\sigma^2\mathbf{I}$, i.e., $\mathbf{n}_i \sim \mathcal{N}(0, \sigma^2\mathbf{I})$, was added to the generated synthetic image. Here, the noise standard deviation is $\sigma = 0.4$.

The classification experiments using the simulated data set have been conducted as follows. For each class, we randomly chose 250 samples from the available ground truth for training, and the remaining samples were used for testing. Table I tabulates the accuracy obtained by the proposed method as a

TABLE II
OA, AA, AND κ STATISTIC AND INDIVIDUAL CLASSIFICATION ACCURACY (IN PERCENTAGE) OBTAINED FOR THE AVIRIS INDIAN PINES IMAGE

Training set	100 samples per class (1200 in total)				10% of available samples per class (1011 in total)			
Class	Samples		Classification methods		Samples		Classification methods	
	Train	Test	MLRsub	MLRsub_{mod}	Train	Test	MLRsub	MLRsub_{mod}
Corn-no till	100	1334	60.63	64.90	143	1291	60.80	71.04
Corn-min till	100	734	65.56	65.40	83	751	39.61	60.04
Corn	100	134	89.07	86.89	23	211	24.44	46.57
Grass/trees	100	397	88.67	89.71	49	448	81.57	84.19
Grass/pasture	100	647	94.29	93.86	74	673	92.42	94.31
Hay-windrowed	100	389	99.38	99.07	48	441	99.24	98.34
Soybeans-notill	100	868	72.98	67.36	96	872	48.00	64.90
Soybeans-min till	100	2368	55.07	58.55	246	2222	89.68	78.33
Soybeans-clean till	100	514	84.01	79.26	61	553	58.09	64.73
Wheat	100	112	99.65	99.65	21	191	98.73	99.56
Woods	100	1194	91.47	91.63	129	1165	99.51	95.93
Bldg-Grass-Tree-Drives	100	280	66.53	70.54	38	342	27.49	41.70
OA			73.59	74.32	OA		73.51	76.71
AA			80.61	80.57	AA		68.30	74.97
κ			70.20	70.98	κ		68.87	73.23

function of the value of parameter α that controls the impact of the nonlinear features, in comparison with the conventional MLRsub, and all the values of the overall accuracy (OA) reported in this section correspond to the average of the accuracy values obtained after 100 Monte Carlo runs. From the results reported in Table I, we can conclude that the results achieved by the proposed MLRsub$_{\mathrm{mod}}$ algorithm are superior to those obtained by the MLRsub algorithm for all the considered values of parameter α. However, the improvement is more significant for low values of α. This is because, when parameter α increases, the value of γ_0 (i.e., the dominant class) decreases. If we compare the OA obtained for linear classes {1, 3, 5, 7} with regard to the OA obtained for nonlinear classes {2, 4, 6, 8}, we can observe that the proposed approach has very good improvements. This indicates that the proposed method can efficiently handle nonlinear mixtures. Furthermore, if we compare the OA obtained for classes {1, 2, 3, 4} (which contain pure pixels) with the OA obtained for classes {5, 6, 7, 8} (which contain mixed pixels), it is apparent that the improvement in OA is more significant for the classes without pure pixels. In other words, the proposed method can better manage mixed pixels instead of pure pixels as mixed pixels stay in the boundaries of the subspaces so that they are more difficult for subspace identification.

B. Experiments With Real Hyperspectral Data

Two different real hyperspectral images with different characteristics and contexts (one agricultural area and one urban area, with different spectral and spatial resolutions) were used in our experiments.

1) The AVIRIS Indian Pines scene was recorded over Northwestern Indiana in June 1992. The image has 145 $\times$ 145 pixels, with a spatial resolution of 20 m per pixel and 200 spectral channels. The original ground truth contains 16 classes of different types of crops, whereas in this letter, we discarded four classes, i.e., *Alfalfa*, *grass/pasture mowed*, *oats*, and *stone-steel towers*, which contain less than 100 labeled pixels. The considered ground-truth map is shown in Fig. 2(a).
2) The Pavia University image was recorded by the Reflective Optics System Imaging Spectrographic (ROSIS) sensor over the urban area of Pavia, Italy. The size of the image is 610 $\times$ 340 pixels, with 103 spectral channels and a spatial resolution of 1.3 m per pixel. The reference data contain nine classes of interest, which are shown in Fig. 3(a). The original training and test sets are composed of 3921 and 40 002 pixels, respectively.

In our experiments with real hyperspectral scenes, we designed two strategies to choose the training set. In our first strategy, we choose a constant number of training samples per class. In the second strategy, we choose a number of training samples per class that is proportional to the number of available labeled samples. The classification accuracy reported for the real scenes was obtained after 30 Monte Carlo runs.

Tables II and III summarize the OA, the average accuracy (AA), kappa coefficient κ, and the class-specific accuracy values for the two considered images, respectively. It is noticeable that the MLRsub$_{\mathrm{mod}}$, which includes class-dependent information and integrates the prior distribution of classes in the scene, significantly improves the classification accuracy provided by the MLRsub. For instance, Table II shows that the proposed MLRsub$_{\mathrm{mod}}$ approach obtained an OA of 76.71% and an AA of 74.97%, which contrast with an OA of 73.51% and an AA of 68.30% achieved by the MLRsub approach in the AVIRIS Indian Pines scene. Furthermore, the results reported in Table III reveal more significant improvements for the Pavia University image. Using the proposed method, the OA and the AA are improved by 5.57% and 8.09%, respectively, compared with the MLRsub.

TABLE III
OA, AA, AND κ STATISTIC AND INDIVIDUAL CLASSIFICATION ACCURACY (IN PERCENTAGE) OBTAINED FOR THE ROSIS PAVIA UNIVERSITY IMAGE

Training set	100 samples per class (900 in total)				20% of the 3921 training set (781 in total)			
Class	Samples		Classification methods		Samples		Classification methods	
	Train	Test	MLRsub	MLRsub_{mod}	Train	Test	MLRsub	MLRsub_{mod}
Asphalt	100	6531	44.61	67.00	109	6522	61.69	67.00
Meadows	100	18549	63.33	73.06	108	18541	72.89	76.67
Gravel	100	1999	65.91	66.94	78	2021	46.48	66.56
Trees	100	2964	76.68	94.36	104	2960	78.08	93.62
Metal sheets	100	1245	98.89	99.04	53	1292	98.45	98.87
Bare soil	100	4929	69.21	62.59	106	4923	65.37	64.25
Bitumen	100	1230	86.46	85.75	75	1255	66.60	85.56
Bricks	100	3582	64.22	78.07	102	3580	73.12	78.11
Shadows	100	847	99.29	99.71	46	901	94.73	99.57
OA			64.91	74.35	OA		70.46	76.03
AA			74.29	80.72	AA		73.04	81.13
κ			56.27	67.34	κ		62.20	69.31

If we focus on the results reported for the classes in which a constant number of training samples is selected for all classes, we can see that the class-specific accuracy values for MLRsub$_{\mathrm{mod}}$ are higher compared with those of MLRsub in most of the classes. This reveals that the proposed projection-based feature vectors provide a more consistent estimation of the posterior probability distributions. For illustrative purposes, some classification maps are shown in Figs. 2 and 3. These maps correspond to one of the 30 Monte Carlo runs conduced for each scene. Effective results can be observed in these figures.

IV. CONCLUSION

In this letter, we have developed a subspace-based MLR method for pixelwise hyperspectral classification. The proposed approach assumes that the observed vectors live in subspaces constructed by the classes and represents an extension of a previous methodology in which class independence was assumed. An important contribution of the proposed approach lies in its ability to deal with both linear and nonlinear mixtures. Our experimental results, which are conducted using both simulated and real hyperspectral data sets collected using NASA's AVIRIS and the ROSIS system, indicate that the proposed algorithm accurately performs in different hyperspectral image classification scenarios, particularly with limited training samples.

ACKNOWLEDGMENT

The authors would like to thank the Associate Editor who handled this manuscript and the anonymous reviewers for providing outstanding comments and suggestions that greatly helped us improve the technical quality and presentation of this manuscript. The authors would also like to thank Prof. D. Landgrebe for making the AVIRIS Indian Pines hyperspectral data set available to the community and Prof. P. Gamba for providing the ROSIS data over Pavia, Italy, along with the training and test sets.

REFERENCES

[1] C.-I. Chang, *Hyperspectral Data Exploitation: Theory and Applications*. New York, NY, USA: Wiley-Interscience, 2007.
[2] Q. Cheng, P. Varshney, and M. Arora, "Logistic regression for feature selection and soft classification of remote sensing data," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 4, pp. 491–494, Oct. 2006.
[3] 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, 2010.
[4] 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. 2011.
[5] J. Li, J. 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.
[6] J. Bioucas-Dias *et al.*, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.
[7] M. Pal and G. Foody, "Evaluation of SVM, RVM and SMLR for accurate image classification with limited ground data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 5, pp. 1344–1355, Oct. 2012.
[8] J. Borges, J. Bioucas-Dias, and A. R. S. Marcal, "Bayesian hyperspectral image segmentation with discriminative class learning," *IEEE Trans. Geosc. Remote Sens.*, vol. 49, no. 6, pp. 2151–2164, Jun. 2011.
[9] P. Zhong, P. Zhang, and R. Wang, "Dynamic learning of SMLR for feature selection and classification of hyperspectral data," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 280–284, Apr. 2008.
[10] M. A. T. Figueiredo, "Bayesian image segmentation using Gaussian field priors," in *Energy Minimization Methods in Computer Vision and Pattern Recognition*, vol. 3757, *Lecture Notes in Computer Science*, A. Rangarajan, B. Vemuri, and A. Yuille, Eds. Berlin, Germany: Springer-Verlag, 2005, pp. 74–89.
[11] D. Böhning, "Multinomial logistic regression algorithm," *Ann. Inst. Statist. Math.*, vol. 44, pp. 197–200, 1992.
[12] B. Krishnapuram, L. Carin, M. Figueiredo, and A. Hartemink, "Sparse multinomial logistic regression: Fast algorithms and generalization bounds," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 27, no. 6, pp. 957–968, Jun. 2005.