Group Lasso-Based Band Selection for Hyperspectral Image Classification

Daiqin Yang, *Member, IEEE*, and Wentao Bao

***Abstract*— Band selection plays an important role in reducing the dimensionality of spectral response of hyperspectral images (HSIs) to avoid dimension disaster for land-cover classification. Compared with traditional dimension reduction methods, such as principal component analysis, independent component analysis, or linear discriminant analysis, band selection can help provide interpretability to later constructed models by preserving the physical meaning of selected features. In this letter, a group lasso-based band selection (GLBS) method is proposed for multilabel HSI classification. Using the group lasso algorithm, the two objectives of band selection and classification are implemented simultaneously. The performance of GLBS is fully investigated and compared with benchmark methods, and the experimental results demonstrate the superiority of GLBS.**

***Index Terms*— Band selection, group lasso, hyperspectral image (HSI), multinomial classification.**

I. Introduction

HYPERSPECTRAL remote sensing images can give a fine spectral description about target objects observed by remote sensors. Typical hyperspectral images (HSIs) provide a 5–10-nm spectral resolution and more than 200 spectral bands in a 0.4–2.5-μm wavelength range. Such detailed spectrum information can help researchers build delicate spectrum models for various land-cover types but also introduces the potential dimension disaster problem. Therefore, dimension reduction (DR) plays an important role in HSI-based land-cover classification. There are roughly two approaches to fulfill DR. One is feature extraction, including principal component analysis [1], independent component analysis [2], and linear discriminant analysis [3]; DR is conducted in a transformed feature space by selecting several most representative features for later model construction. The other approach of DR is feature (band) selection, which directly selects a subset of spectral bands from the continuous spectral input of the HSI. Comparing with feature extraction, band selection-based DR has the merit of preserving physical meanings of input data. It can thus give interpretable insights to the later constructed spectral models and help reveal physical meanings behind. The focus of this letter is band selection for HSI-based classification.

Manuscript received January 2, 2017; revised May 2, 2017 and August 9, 2017; accepted October 24, 2017. Date of publication November 16, 2017; date of current version December 4, 2017. This work was supported by the Natural Science Foundation of Hubei Province of China under Grant 2015CFA053. *(Corresponding author: Daiqin Yang.)*

The authors are with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China (e-mail: dqyang@whu.edu.cn; wtbao@whu.edu.cn).

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

Digital Object Identifier 10.1109/LGRS.2017.2768074

Band selection algorithms can be categorized into unsupervised and supervised approaches. Unsupervised band selection algorithms, including [4]–[6], select frequency bands according to the statistical distributions of the whole data set, and no training samples with known class labels are needed. As most of the unsupervised band selection algorithms are designed with clustering mechanisms, seldom of them can provide incremental selections of frequency bands. Without incremental band selection, any changes to the number of required frequency bands will result in a totally different selected subset of frequency bands compared with the previous one. Enhanced fast density-peak-based clustering (E-FDPC) [6] tries to support incremental band selection by providing ranking sequences to the clustering results; however, the ranking sequences are usually noisy due to the clustering nature of unsupervised band selections.

Band selection can also be fulfilled in supervised fashion, which takes into account the separation performance of training samples with known class labels. In [7], a rough set-based band selection (RSBS) algorithm ranks each spectral band according to its relevance and significance to the classification of training samples. The band selection results exhibit remarkable classification accuracies with limited number of selected bands, while as the number of selected bands increases, the later selected frequency bands become less accurate. In [8], a lasso-based band selection algorithm was proposed for HSIs. This algorithm has the merit of incremental band selection due to the underlying lasso property, while it can be used only for object detections due to its binary classification limitation. Sparse multinomial logistic regression algorithm with Bayesian regularization (SBMLR) [9], [10] uses a multinomial classification version of lasso to provide band selection for multilabel land-cover classifications. It exploits the Laplace prior and Bayesian evidence framework to derive a closed-form solution for the regularization parameter, and thus avoids the cross-validation phase for finding the optimal number of selected bands. However, as discussed later in Section II-B, its ungrouped sparsity to the coefficients will degrade the performance of classification accuracy.

In this letter, a group lasso-based band selection (GLBS) algorithm is proposed for multilabel land-cover classification problem of HSIs. The superiority of GLBS is demonstrated through substantial experiments comparing with existing band selection algorithms. The contributions of this letter can be summarized as the following twofolds. First, this letter is the first that introduces the group lasso algorithm for the band selection problem of multilabel land-cover classification for HSIs. Second, application issues are discussed relating

to the characteristics of HSIs and performances are studied through extensive experiments and analysis. The rest of this letter is organized as follows. Section II in detail depicts the algorithm of GLBS. Then experimental results are presented and analyzed in Section III. And finally, Section IV concludes this letter.

II. Group Lasso-Based Band Selection for Multilabel Classification

Lasso [11] has been a very famous algorithm for its capability of fulfilling the tasks of subset selection and linear regression simultaneously. In [12], a revised version of lasso was proposed to impose the sparsity constraint to groups of input dimensions, and in [13], this algorithm was generalized to solve binary classification problems with logistic regression. Based on these two pieces of research work, Vincent and Hansen [14] and Simon *et al.* [15] proposed a group lasso algorithm for multinomial classification, i.e., multiresponse version of logistic regression, to which the group-level sparsity constraint fits perfectly with the requirement that the same subset of input dimensions are selected for all the classes.

A. Multinomial Classification

Multilabel land-cover classification of HSI can be achieved by modeling the conditional probability density function (PDF) of each class with multinomial distributions. For HSIs with K land-cover classes and M spectral bands, given N training samples (HSI pixels), each with known spectral responses and class labels, the sample pairs can be denoted by $(\mathbf{x}_i, \mathbf{y}_i), i \in \{1, 2, \ldots, N\}$, where $\mathbf{x}_i \in R^M$ is an M dimension column vector representing the spectral responses of the ith sample, and its jth element x_i^j represents the response of the jth spectral band. $\mathbf{y}_i \in \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_K\}$ is a K dimension column vector depicting the class label of sample i, where $\mathbf{e}_k$ is the kth column vector of a $K \times K$ diagonal matrix $\mathbf{I}$. It should be noted here that, with multinomial distribution, the class label $\mathbf{y_i}$ is no longer a scale value, but a K dimension column vector with its kth element defined as

$$y_i^k = \begin{cases} 1 & \text{if sample } i \text{ belongs to class } k \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

For each sample label $\mathbf{y}_i$, the sum of its elements equals one, $\sum_{k=1}^{K} y_i^k = 1$. The conditional PDF of class label $\mathbf{y}$, given spectral response $\mathbf{x}$, can be defined as

$$P(\mathbf{y}|\mathbf{x}) = \prod_{k=1}^{K} [P(\mathbf{e}_k|\mathbf{x})]^{y^k} \tag{2}$$

where $P(\mathbf{e}_k|\mathbf{x})$ is the probability that a sample with spectrum response $\mathbf{x}$ belongs to class k. It can be modeled with the following multinomial distribution in matrix form:

$$P(\mathbf{e}_k|\mathbf{x}) = \frac{\exp(\boldsymbol{\beta}_k^T \mathbf{x})}{\sum_{p=1}^{K} \exp(\boldsymbol{\beta}_p^T \mathbf{x})} \tag{3}$$

where $\mathbf{x}$ is now an $M+1$ dimension column vector including a constant element of $x^0 = 1$ to accommodate the intercept, $\boldsymbol{\beta}_k$ is the kth column of coefficient matrix $\boldsymbol{\beta}$, and $\boldsymbol{\beta}_k^T$ is the transpose of $\boldsymbol{\beta}_k$. The coefficient matrix $\boldsymbol{\beta}$ is of $(M+1) \times K$ dimension, with its columns and rows relating to different classes and frequency bands, respectively.

With N sample pairs $(\mathbf{x}_i, \mathbf{y}_i)$, $i \in \{1, 2, \ldots, N\}$, the above multinomial classification problem can be resolved by maximum-likelihood estimation, which finds the optimal solution of $\hat{\boldsymbol{\beta}}$ that maximizes the log-likelihood function of the following:

$$\begin{aligned} l(\mathbf{x}_i, \mathbf{y}_i, \boldsymbol{\beta}) &= \log \prod_{i=1}^{N} P(\mathbf{y_i}|\mathbf{x_i}) \\ &= \sum_{i=1}^{N} \left[\mathbf{y}_i^T \boldsymbol{\beta}^T \mathbf{x}_i - \log \sum_{k=1}^{K} \exp\left(\boldsymbol{\beta}_k^T \mathbf{x}_i\right) \right]. \end{aligned} \tag{4}$$

B. Group Lasso-Based Multinomial Classification and Band Selection

The original lasso algorithm introduces an $l1$-norm penalty item to the ordinary least square estimates, as presented in the following, and brings sparsity to the solutions of linear regression problems:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{N} \frac{1}{2} \| y_i - \boldsymbol{\beta}^T \mathbf{x}_i \|_2^2 + \lambda \|\boldsymbol{\beta}\|_1 \tag{5}$$

where $y_i \in R$ is a scale-valued sample output, and coefficient $\boldsymbol{\beta}$ is an $M+1$ dimension column vector, where M is the original dimension of input $\mathbf{x}$. The $l1$-norm constraint of $\boldsymbol{\beta}$ is defined as $\|\boldsymbol{\beta}\|_1 = \sum_{j=0}^{M} |\beta^j|$. Sparsity is brought into the solution with the singularity of $l1$-norm at the vertexes along each basis. And the regularization parameter λ can be adjusted to control the sparsity level of solutions.

Similar to the original lasso algorithm, the GLBS algorithm is formulated as the minimization of the (minus) multinomial log-likelihood function of (4), with an $l1$-norm constraint of the coefficient matrix $\boldsymbol{\beta}$. It is formulated as follows:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} - \sum_{i=1}^{N} \left[\mathbf{y}_i^T \boldsymbol{\beta}^T \mathbf{x}_i - \log \sum_{k=1}^{K} \exp\left(\boldsymbol{\beta}_k^T \mathbf{x}_i\right) \right] + \lambda \|\boldsymbol{\alpha}\|_1. \tag{6}$$

The penalty item $\|\boldsymbol{\alpha}\|_1$ is the $l1$-norm of $\boldsymbol{\beta}$ in row level. It is defined as $\|\boldsymbol{\alpha}\|_1 = \sum_{j=0}^{M} |\alpha_j|$, $\alpha_j = \|\boldsymbol{\beta}^j\|_2$, where $\boldsymbol{\beta}^j$ is the jth row of the coefficient matrix $\boldsymbol{\beta}$. Although α_j is the $l2$-norm of $\boldsymbol{\beta}^j$, the direct summation of $|\alpha_j|$ makes $\|\boldsymbol{\alpha}\|_1$ an $l1$-norm. With GLBS, the band selection phase and the land-cover classification phase are combined into one optimization. As the value of λ decreases, the underlying lasso property of GLBS ensures that frequency bands are incrementally inserted into the final classification model.

Besides the substitution of the least square item by the (minus) log-likelihood function, the major difference between (5) and (6) is the form of sparsity constraint. In (5), the $l1$-norm constraint is defined at element level of coefficient vector $\boldsymbol{\beta}$. A similar generalization to the matrix version of $\boldsymbol{\beta}$ for multinomial classification should take the form of $\|\boldsymbol{\beta}\|_1 = \sum_{k=1}^{K} \sum_{j=0}^{M} |\beta_k^j|$, which is just the same as the SBMLR [10] algorithm. However, with group lasso, the sparsity constraint in (6) now takes the form of $\sum_{j=0}^{M} \|\boldsymbol{\beta}^j\|_2$, in which the

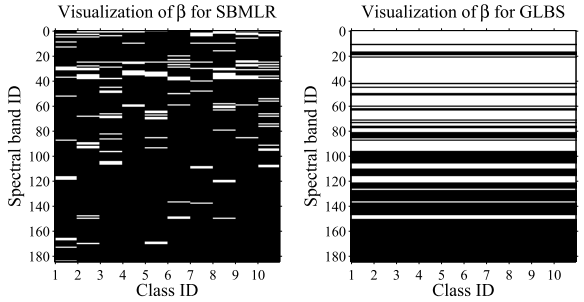

Fig. 1. Visualized sparsity of the coefficient matrix $\boldsymbol{\beta}$ for SBMLR and GLBS. The size of β is 185×10 for the Indian Pines data set (please refer to Section III for details), and nonzero coefficients are marked with white lines.

$l1$-norm is defined at row level of the coefficient matrix. This change is due to the drawback of element-wise $l1$-norm, with which the achieved sparsity exhibits violent variations across classes and gives very different set of nonzero coefficients for individual class.

Fig. 1 shows the sparsity of the coefficient matrix $\boldsymbol{\beta}$ achieved with SBMLR and GLBS, given the same constraint of 86 selected bands. Nonzero coefficients, which correspond to selected frequencies for certain land-cover classes, are marked with white lines. As shown in Fig. 1, with element-wise $l1$-norm of SBMLR, the selected frequency bands for one class could be totally different from another. For multinomial classification, when there are a large number of land-cover types to be classified, this inconsistent sparsity across classes may result in either a very large number of total selected bands and thus a complex classification model, or a very small number of selected bands for individual classes and thus degraded classification accuracy. With GLBS, the sparse penalty item changes from the $l1$-norm of the whole $\boldsymbol{\beta}$ matrix to the $l1$-norm of coefficient rows, which guarantees the row-level sparsity of the achieved $\boldsymbol{\beta}$. As $\boldsymbol{\beta}^j$ represents the impact of a certain spectral band j to all the K land-cover classes, row-level sparsity of $\boldsymbol{\beta}$ means that only a subset of spectral bands will be selected, and each selected band will contribute to the classification of all the classes. As shown in Fig. 1, the selected frequency bands for each land-cover class are the same with GLBS. Thus, with the same constraint of total selected bands, more frequencies will be involved in the multinomial classification model for each land-cover type.

C. GLBS and SVM

With GLBS, the band selection phase and the land-cover classification phase are combined into one optimization problem and solved simultaneously. When taking the band selection and land classification phases as a whole, the resolved sparse coefficients of $\boldsymbol{\beta}$ can provide indications about both which frequencies are selected, and how they are used in the multinomial classification model. As the value of λ in (6) decreases, more frequency bands will be selected into the classification model, and incremental band selection is guaranteed in the sense that each time the required number of selected bands increases, it is always new frequency bands to be inserted into the model and those already selected ones will not be removed.

GLBS can also be used for the selection of candidate frequency bands only, and leave the classification phase to other advanced algorithms, like the support vector machine (SVM) [16], as by other band selection algorithms [6], [7], [10]. In this case, the optimization problem of (6) can be solved only once with a very small λ, and the solution finding process will return a very long band list, representing the selection order of each band. Truncations to the list with different band numbers can thus be used as incremental band selection. Given the set of selected bands, the SVM algorithm can then be used to classify training samples into two classes with an optimal decision hyperplane. When used for multilabel classification, a one-against-all mechanism is always adopted to classify each land-cover type from others. SVM can also be extended to support nonlinear decision boundary using various kernels. In the performance evaluations of Section III, linear SVM is adopted for its simplicity in parameter optimization.

There is usually a standardization step before the implementation of lasso, so as to dispel unit and fluctuation range discrepancies among input dimensions. For HIS-based land-cover classification, the input is the spectral reflectance of an image pixel, and thus the units of all the input dimensions are the same. Examples of HSI inputs can be found in Fig. 3. It demonstrates the spectral reflectance of HSI pixels of the Indian Pines data set, which is a commonly used Airborne Visible Infrared Imaging Spectrometer data set taken over the agricultural area of Northwest Indiana [17]. Obvious extent and fluctuation range differences across frequency bands can be found. Besides the spectral signature differences of different land-cover types, the major reason for these discrepancies can be attributed to the solar radiation pattern and atmospheric absorption of H_2O, CO_2, O_2, etc. [18]. Although a zero-mean and unit variance standardization can help enlarge the signature difference of those frequencies under severe absorptions, it will also amplify the noise levels of the signal with fixed measurement sensitivity over the entire frequency bands. This will bring in additional noises and thus result in degradations of classification accuracy. It is therefore a better choice not to go through the standardization phase and use the original data directly.

III. Experiments

To evaluate the performance of GLBS for HSI classification, a set of experiments are conducted based on the Indian Pines data set [17] and the Botswana data set [19]. The performances are evaluated from two aspects of views, i.e., the classification accuracy and the ranking of frequency band. Land-cover types with sample number less than 300 and 200 are removed from the experiment, respectively, for reasonable experimental settings. Performances of GLBS are compared with E-FDPC [6], SBMLR [9], and RSBS [7].

A. Classification Accuracy

Classification performances are measured in terms of overall accuracy (OA), kappa coefficient (Kappa), and average accuracy (AA), which are all derived from the confusion matrix. All the results are averages of 10 round random

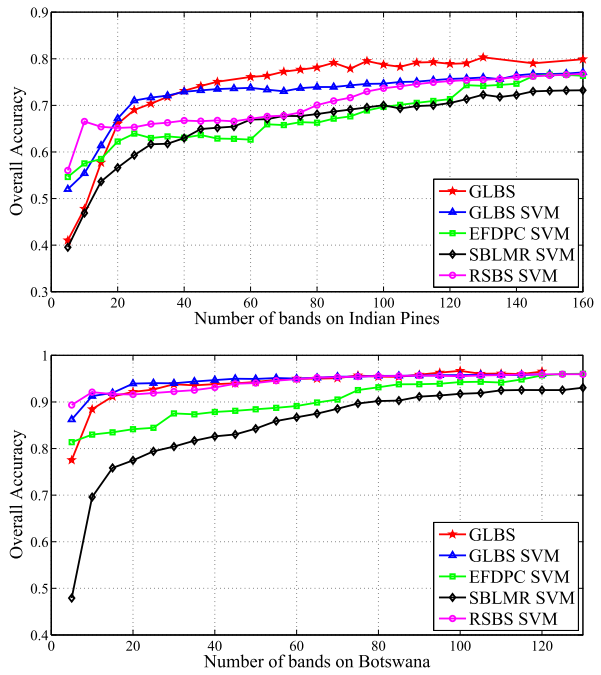

Fig. 2. OA with different numbers of selected bands.

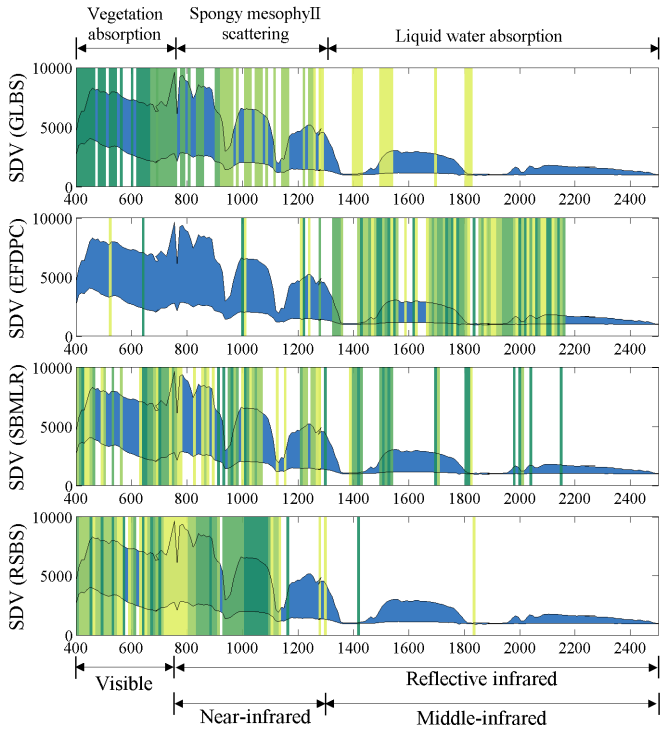

Fig. 3. Ranking sequences of 80 selected bands with GLBS, E-FDPC, SBMLR, and RSBS (from top to bottom). Selected bands are marked in graduated colors from green to yellow. Each color represents a 20 consecutive band selection. The blue background indicates the variation range of the scene data values versus HSI center wavelength (nm).

experiments, each with 150 (Indian Pines) or 100 (Botswana) randomly selected training samples per land-cover type. Fig. 2 shows the achieved overall accuracies of the two data sets. With the Indian Pines data set, the OAs of GLBS and SBMLR experience a slow start as the number of selected bands increases, and GLBS achieves the best saturation performance with more than 40 selected frequency bands. RSBS demonstrates a better performance with less than

TABLE I

KAPPA AND AA WITH RESPECT TO THE NUMBER OF SELECTED BANDS (N-BANDS) ON INDIAN PINES AND BOTSWANA DATA SET. (THE BEST RESULTS ARE SHOWN IN BOLDFACE)

Data	nBands	Evaluation	GLBS	GLBS(SVM)	E-FDPC	SBMLR	RSBS
Indian Pines	20	Kappa	0.608	**0.623**	0.567	0.500	0.500
		AA	0.698	**0.745**	0.700	0.603	0.725
	40	Kappa	**0.689**	0.687	0.576	0.575	0.617
		AA	0.785	**0.793**	0.702	0.688	0.745
	80	Kappa	**0.758**	0.700	0.613	0.634	0.655
		AA	**0.838**	0.811	0.747	0.762	0.781
Botswana	20	Kappa	0.914	**0.933**	0.825	0.752	0.907
		AA	0.930	**0.944**	0.849	0.783	0.922
	40	Kappa	0.932	**0.942**	0.866	0.808	0.924
		AA	0.943	**0.951**	0.885	0.833	0.936
	80	Kappa	0.950	**0.952**	0.925	0.892	0.951
		AA	0.957	**0.959**	0.937	0.908	0.959

TABLE II

ELAPSED TIME (SECONDS) IN BAND SELECTION STAGE

Data set	GLBS	SBMLR	RSBS	E-FDPC
Botswana	46.006	26.138	1032.922	34.440
Indian Pines	102.154	152.172	1529.279	2.232

20 selected bands. For the Botswana data set, GLBS and GLBS SVM achieve similar performances to RSBS, and outperform E-FDPC and SBMLR. For quantitative comparisons, the kappa coefficients and AAs are listed in Table I. The results demonstrate that GLBS and GLBS SVM achieve the best performances, and the AA of GLBS could be 12.2%, 10.0%, and 7.3% higher than that of E-FDPC, SBMLR, and RSBS, respectively, when 80 frequency bands are selected for the Indian Pines data set.

In summary, the achieved classification accuracies suggest the superiority of GLBS with relatively large numbers of selected frequency bands. This can be attributed to the row-level sparsity of GLBS, which guarantees the selection of those frequency bands that are most discriminative to all the land-cover types. Compared with results on Botswana data set, the superiority of GLBS is more significant on the Indian Pines. This, on the one hand, is due to the already saturated accuracy level of the Botswana data set (around 95% AA). On the other hand, the similarities among land-cover types of the Indiana Pines data make the classification a challenging task and thus vulnerable to the selection of frequency bands. It is also notable that RSBS and E-FDPC have better classification accuracies with less than 20 selected bands. This early saturated accuracy levels infer that their ranking results are more accurate for the first few selected bands and are noisy beyond.

The execution times of the four band selection algorithms are also compared. The average times to calculate the selection sequence of the whole frequency bands are listed in Table II. The results show that GLBS has a moderate speed, which is slower than E-FDPC but much faster than RSBS. Compared with SBMLR, GLBS is faster for data set with larger number of training samples (Indian Pines), while slower for data set with less training samples (Botswana).

B. Ranking of Frequency Bands

The ranking sequences of the selected bands with GLBS, E-FDPC, SBMLR, and RSBS are investigated and compared as shown in Fig. 3. As illustrated in Fig. 3, the wavelength of the first batch of 20 selected bands by GLBS, which is shown in the first row of Fig. 3, falls into the visible light range of about $0.4-0.6$ μm. This range is the vegetation absorption region, demonstrating dominant leaf reflecting factors, such as chlorophyll and β-carotene absorptions [18]. The higher ranking of these frequency bands conforms to the fact that most of the sample data are vegetation samples, as described in [17]. The wavelength of the second batch of 20 selected bands lies in the range of about $0.6-0.9$ μm, which is another sensitive region of leaf's spectral reflectance, including the most informative "red edge." After these two batches of 20 bands, the succedent 40 bands are then eventually spread toward longer wavelengths. This ranking sequence of the selected frequency bands conforms to the spectral reflectance properties of the studied land-cover types, and thus gives an interpretable classification model using those selected bands.

The other three rows of Fig. 3 show comparative illustrations of the ranking sequence of the selected bands with E-FDPC, SBMLR, and RSBS, respectively. Compared with GLBS, the ranking sequences of E-FDPC and SBMLR show a much more random pattern. There are no obvious clustering ranges for each 20-band batch. The first batch of the selected 20 bands with E-FDPC and SBMLR are widely spread over the entire spectrum, including bands with more than 2.0-μm wavelength. A large portion of the 80 selected bands with E-FDPC are in the range of $1.4-2.2$ μm, which is more related to the water content of the object. This random pattern and shifted focal region of selected bands with E-FDPC and SBMLR demonstrate again the unreliable ranking of frequency bands beyond their recommended optimal number of selected bands. With RSBS, the very informative "red edge" frequencies are given the least importance among the 80 bands' selection. It thus explains the reason why the performance of RSBS falls below GLBS with more than 20 selected bands.

C. Discussion

The comparison among GLBS, E-FDPC, SBMLR, and RSBS from the perspectives of classification accuracy and ranking of frequency bands reveals the fundamental difference among these band selection methods. If the aim of an application is to build a condensed model with very limited number of frequency bands, E-FDPC and RSBS could be good choices, which exhibit better classification accuracies with less than 20 frequency bands. However, if the application allows an acceptable larger number of selected bands for better classification accuracy and pays more attention to the physical explanation of the ranking sequence to the selected frequency bands in the final classification model, GLBS would be a better selection.

IV. Conclusion

In this letter, a GLBS method is proposed for multilabel land-cover classification with HSIs. With GLBS, a tradeoff between model complexity and classification accuracy can be achieved arbitrarily by choosing a proper number of selected bands, while maintaining the classification model interpretable. After a detailed elaboration of the group lasso algorithm, the performances of GLBS are studied through extensive experiments and compared with an unsupervised algorithm E-FDPC, and two supervised algorithms, SBMLR and RSBS. Results demonstrate that group lasso can achieve a more than 10% accuracy improvement and its ranking sequence to individual bands conforms to the property of the studied land-cover type, which helps the understanding of the final classification model.

References

[1] A. Agarwal, T. El-Ghazawi, H. El-Askary, and J. Le-Moigne, "Efficient hierarchical-PCA dimension reduction for hyperspectral imagery," in *Proc. IEEE ISSPIT*, Dec. 2007, pp. 353–356.

[2] J. Wang and C.-I. Chang, "Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1586–1600, Jun. 2006.

[3] W. Li, S. Prasad, J. E. Fowler, and L. M. Bruce, "Locality-preserving dimensionality reduction and classification for hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 4, pp. 1185–1198, Apr. 2012.

[4] Q. Du and H. Yang, "Similarity-based unsupervised band selection for hyperspectral image analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 4, pp. 564–568, Oct. 2008.

[5] Y.-Q. Zhao, L. Zhang, and S. G. Kong, "Band-subset-based clustering and fusion for hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 747–756, Feb. 2011.

[6] S. Jia, G. Tang, J. Zhu, and Q. Li, "A novel ranking-based clustering approach for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 88–102, Jan. 2016.

[7] S. Patra, P. Modi, and L. Bruzzone, "Hyperspectral band selection based on rough set," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 5495–5503, Oct. 2015.

[8] K. Sun, X. Geng, and L. Ji, "A new sparsity-based band selection method for target detection of hyperspectral image," *IEEE Trans. Geosci. Remote Sens. Lett.*, vol. 12, no. 2, pp. 329–333, Feb. 2015.

[9] G. C. Cawley, N. L. C. Talbot, and M. Girolami, "Sparse multinomial logistic regression via Bayesian L1 regularisation," in *Advances in Neural Information Processing Systems*. Cambridge, MA, USA: MIT Press, 2006, pp. 209–216.

[10] M. Pal, "Multinomial logistic regression-based feature selection for hyperspectral data," *Int. J. Appl. Earth Observat. Geoinf.*, vol. 14, no. 1, pp. 214–220, Feb. 2012.

[11] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Stat. Soc. B, Stat. Methodol.*, vol. 58, no. 1, pp. 267–288, 1996.

[12] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. Roy. Stat. Soc. B, Stat. Methodol.*, vol. 68, no. 1, pp. 49–67, 2006.

[13] L. Meier, S. Van De Geer, and P. Bühlmann, "Group lasso for logistic regression," *J. Roy. Stat. Soc. B, Stat. Methodol.*, vol. 70, no. 1, pp. 53–71, 2008.

[14] M. Vincent and N. R. Hansen, "Sparse group lasso and high dimensional multinomial classification," *Comput. Stat. Data Anal.*, vol. 71, pp. 771–786, Mar. 2012.

[15] N. Simon, J. Friedman, and T. Hastie, "A blockwise descent algorithm for group-penalized multiresponse and multinomial regression," unpublished paper, 2013. [Online]. Available: https://arxiv.org/abs/1311.6529

[16] V. Vapnik, *The Nature of Statistic Learning Theory*. New York, NY, USA: Springer-Verlag, 1995.

[17] D. Landgrebe. (1992). *AVIRIS NW Indianas Indian Pines 1992 Data Set*. [Online]. Available: http://ftp.ecn.purdue.eduDirectory:biehl/MultiSpec/File:92AV3C.lan

[18] J. R. Jensen, *Remote Sensing of the Environment: An Earth Resource Perspective*, 2nd ed. Bengaluru, India: Pearson Education, 2007.

[19] A. L. Neuenschwander, M. M. Crawford, and S. Ringrose, "Results from the EO-1 experiment—A comparative study of Earth observing-1 advanced land imager (ALI) and landsat ETM+ data for land cover mapping in the Okavango Delta, Botswana," *Int. J. Remote Sens.*, vol. 26, no. 19, pp. 4321–4337, 2005.