


Technical Note

Band Ranking via Extended Coefficient of Variation for Hyperspectral Band Selection

Peifeng Su ^{1,2,*}, Sasu Tarkoma ^{2,3} and Petri K. E. Pellikka ^{1,2} 

¹ Department of Geosciences and Geography, University of Helsinki, FI-00014 Helsinki, Finland; petri.pellikka@helsinki.fi

² Institute for Atmospheric and Earth System Research, Faculty of Science, University of Helsinki, FI-00014 Helsinki, Finland

³ Department of Computer Science, University of Helsinki, FI-00014 Helsinki, Finland; sasutarkoma@helsinki.fi

* Correspondence: peifeng.su@helsinki.fi

Received: 11 September 2020; Accepted: 8 October 2020; Published: 12 October 2020



Abstract: Hundreds of narrow bands over a continuous spectral range make hyperspectral imagery rich in information about objects, while at the same time causing the neighboring bands to be highly correlated. Band selection is a technique that provides clear physical-meaning results for hyperspectral dimensional reduction, alleviating the difficulty for transferring and processing hyperspectral images caused by a property of hyperspectral images: large data volumes. In this study, a simple and efficient band ranking via extended coefficient of variation (BRECV) is proposed for unsupervised hyperspectral band selection. The naive idea of the BRECV algorithm is to select bands with relatively smaller means and larger standard deviations compared to their adjacent bands. To make this simple idea into an algorithm, and inspired by coefficient of variation (CV), we constructed an extended CV matrix for every three adjacent bands to study the changes of means and standard deviations, and accordingly propose a criterion to allocate values to each band for ranking. A derived unsupervised band selection based on the same idea while using entropy is also presented. Though the underlying idea is quite simple, and both cluster and optimization methods are not used, the BRECV method acquires qualitatively the same level of classification accuracy, compared with some state-of-the-art band selection methods

Keywords: hyperspectral imagery; unsupervised band selection; coefficient of variation; band ranking; entropy; conditional entropy

1. Introduction

Hyperspectral images have a wide range of applications, such as change detection [1], target detection [2–4], semantic interpretation [5] and image classification [6–8]. The reason why hyperspectral images can identify and distinguish a variety of materials is that they have a large amount of narrow spectral bands [9]. On the other hand, while providing detailed spectral measurements, the large number of bands also make the hyperspectral images inconvenient to acquire, store, transmit, process, and also cause the curse of dimensionality [10,11]. However, a subset (i.e., a few bands) of an entire hyperspectral data set could be sufficient to identify and distinguish objects as the information in the nearby bands is highly correlated [12,13]. As a result, dimensional reduction should be applied to reduce the inter-band spectral redundancy without losing significant information in data exploitation [14–16] and to fulfill the requirements of following tasks, taking classification after dimensional reduction as an example.

Feature extraction and band selection (also known as feature selection) are the two main methods of dimensional reduction for hyperspectral images. Unlike band selection methods, which only retain the most informative observations without any transformation [17], the results of feature extraction methods are hard to explain physically. Hence, band selection methods are beneficial for hyperspectral data, not only for analysis but also for storage, transmission and processing. Meanwhile, they may also alleviate the effect of the curse of dimensionality.

According to whether prior knowledge, such as object information, participates in the band selection process and the degree of participation, band selection methods are divided into three categories: unsupervised band selection methods, semi-supervised band selection methods and supervised band selection methods [18]. As the acquisition of the prior knowledge is costly and usually time-consuming, unsupervised band selection methods are the most practical.

According to [9], there are mainly six types of hyperspectral band selection methods: ranking-based, searching-based, clustering-based, sparsity-based, embedding-based and hybrid scheme-based. The ranking-based methods rank each band according to some metrics (such as max information, dissimilarity, and/or correlation of bands) [19,20]. Searching-based methods convert the band selection problem to an optimization problem and select bands according to a given criterion function [21–23]. Clustering-based methods select the representation bands from the band clusters [9,24–26]. Sparsity-based methods transform the band selection problem to a sparsity-constrained optimization one [9,27,28]. Embedding-based methods use different learning criteria, such as the support vector machine (SVM) classifier or deep learning model, to select bands [29,30].

In this work, we propose a simple and efficient band ranking method termed BRECV for hyperspectral band selection. The basic idea is to select bands with relatively lower means and larger standard deviation compared with their adjacent bands. It is believed that these bands are more informative than their adjacent bands and should be chosen.

Coefficient of variation (CV), also known as relative standard deviation, is a dimensionless statistic composed of the mean and the standard deviation of a signal. To study the changes of means and standard deviations between every three nearby bands, we extend the original scalar CVs to a 3×3 CV matrix. Analyzing the changes of means and standard deviations becomes easier if the CV matrix is used. A criterion for assigning values for each band for ranking is proposed. Depending on the simple idea, some derived methods using entropy are also presented.

As we only deal with the mean and standard deviation of each band in a hyperspectral image, we do not need to face the large-volume problem, thus speeding up the band selection schedule. Meanwhile, though the idea is very simple and both cluster and optimization methods are not used, the BRECV method achieves qualitatively the same level of classification accuracy compared with some state-of-the-art band selection methods. In addition, the bands selected have a clear physical meaning: more informative than their adjacent bands.

2. Data Sets

Several real-world hyperspectral data sets were used to verify the effectiveness of the BRECV method. The details of these data sets are introduced in this section.

- Indian Pines

The Indian Pines data set was captured by AVIRIS sensor in Northwestern Indiana, USA, in 1992 [31]. This image consists of 145 × 145 pixels. There are 16 labeled classes, and the wavelengths range from 0.4 μm to 2.5 μm [32]. Four all-zero bands and 20 bands affected by water absorption are removed [31]. Two hundred bands are left for doing experiments.

- Kennedy Space Center (KSC)

This data set was also captured by AVIRIS sensor over Kennedy Space Center, Florida, USA, in 1996. The image consists of 512×614 pixels. There are 13 labeled classes, and 176 bands are left for doing experiments. The wavelengths of this image range from 0.4 μm to 2.5 μm .

- Pavia University

This data set was captured by the ROSIS sensor over Pavia, Northern Italy, in 1998. This image consists of 610 × 340 pixels, and the wavelengths range from 0.43 μm to 0.86 μm . The spatial resolution is 1.3 m. There are 103 bands and 9 classes of interest.

- Botswana

This data set was captured by the NASA EO-1 satellite over Okavango Delta, Botswana, in 2001. The image consists of 1476 × 256 pixels and 14 classes of interest. The wavelengths of this image range from 0.4 μm to 2.5 μm . One hundred and forty-five bands are left to do experiments.

- Salinas

The Salinas data set was also captured by AVIRIS sensor over Salinas Valley, California, USA, in 1998. The image consists of 512 × 217 pixels. There are 16 classes of interest, and the wavelengths range from 0.4 μm to 2.5 μm . Twenty bands are removed due to water absorption, and 204 bands are used in experiments.

- Taita Hills

The Taita Hills data set was captured by airborne AisaEAGLE (Specim Ltd., Finland) imaging spectrometer in Taita-Taveta District, Kenya, in 2012 [8]. The image consists of 586 × 701 pixels with 64 bands from 0.4 μm to 1.0 μm in 0.6 m ground resolution. The image is classified with field information to six agricultural classes, namely to “Acacia” (*Acacia* spp.), “Banana” (*Musa acuminata*), “Grevillea” (*Grevillea robusta*), “Maize” (*Zea mays*), “Mango” (*Mangifera indica*) and “Sugarcane” (*Saccharum officinarum*) [8]. Out of the classes, acacia, grevillea and mango are trees between 3 to 14 m tall, while banana, maize and sugarcane are tall grasses. Figure 1 shows the data in false-color image composition and the related ground truth investigated the following day after the airborne imagery acquisition.

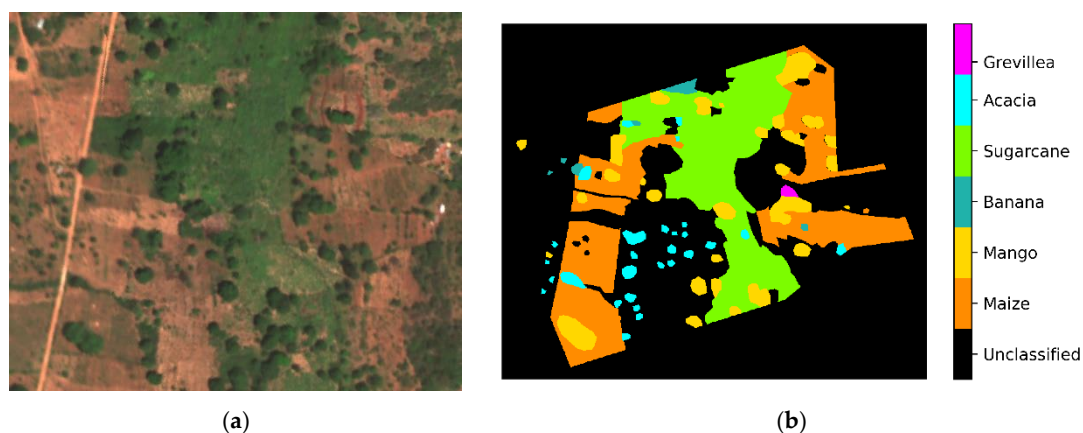


Figure 1. A false-color image over an agricultural field in the Taita Hills data set (a) and the related ground truth (b).

3. The BRECV Method

3.1. Underlying Impetus

We believe that for adjacent bands, a band with a relatively smaller mean and a relatively larger standard deviation means that this band is more informative than its nearby bands. On the contrary, if a band has an increased mean and a decreased standard deviation compared with its adjacent band, this band should not be selected. In our experiment, we directly dropped these bands. From the above analysis, the changes of means and standard deviations of nearby bands need to be investigated.

To this end, we constructed a 3×3 matrix extended from CV for every three adjacent bands. CV is dimensionless and calculates the standard deviation on the unit mean of a band:

$$CV = \sigma/\mu \quad (1)$$

σ is the standard deviation, and μ is the mean of a band.

3.2. Extend Scalar CVs to A 3×3 Matrix

Given three adjacent bands named b1, b2 and b3 with means of μ_1, μ_2, μ_3 and standard deviations of $\sigma_1, \sigma_2, \sigma_3$, we could construct a matrix:

$$M = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix} = \begin{bmatrix} \sigma_1/\mu_1 & \sigma_2/\mu_1 & \sigma_3/\mu_1 \\ \sigma_1/\mu_2 & \sigma_2/\mu_2 & \sigma_3/\mu_2 \\ \sigma_1/\mu_3 & \sigma_2/\mu_3 & \sigma_3/\mu_3 \end{bmatrix} \quad (2)$$

Clearly, we could evaluate the changes of standard deviations through each row in M and the changes of means through each column in M . For a band with a relatively larger standard deviation compared with its adjacent bands, m_{22} will be greater than m_{21} and m_{23} . For a band with a relatively smaller mean compared with its adjacent bands, m_{22} will also be greater than m_{12} and m_{32} , as mean is on the denominator. From this observation, it is easy to select bands with relatively smaller means and larger standard deviations. To evaluate the degree of increase in the standard deviation of b2 compared with b1 and b3, and the degree of decrease in the mean of b2 compared with b1 and b3, a criterion is proposed:

$$value(b2) = m_{22} - m_{21} - (m_{12} - m_{11}) + m_{22} - m_{23} - (m_{32} - m_{33}) \quad (3)$$

$m_{22} - m_{21} - (m_{12} - m_{11})$ is used to compare b2 with b1, and $m_{22} - m_{23} - (m_{32} - m_{33})$ is used to compare b2 with b3. The right-hand side of Equation (3) is equal to $(\sigma_2 - \sigma_1)(1/\mu_2 - 1/\mu_1)$, where $\sigma_2 - \sigma_1$ assesses the degree of how much bigger σ_2 is than σ_1 , and $1/\mu_2 - 1/\mu_1$ measures the degree of how much smaller μ_2 is than μ_1 . Therefore, according to Equation (3), b2 will have a large value if its mean is relatively smaller, and its standard deviation is relatively bigger. For every three adjacent bands, we could use Equation (3) to assign value to the middle band. Therefore, every band in a hyperspectral image will obtain a value for ranking. Ranking these values, we could get the order of each band for band selection.

3.3. Does Entropy Also Work?

In information theory, entropy is usually used to assess the “information” in a signal. The BRECV method tries to select bands that are more informative than their adjacent bands. It is reasonable to ask whether entropy could also be used based on the same idea as the BRECV method. To verify this, we use the conditional entropy to find the relationship between nearby bands. For every adjacent three bands, b1, b2 and b3, b2 will obtain a value based on:

$$value(b_2) = H(b_2|b_1) + H(b_2|b_3). \quad (4)$$

$H(b_2 | b_1)$ is the conditional entropy of b_2 given b_1 and $H(b_2 | b_3)$ is the conditional entropy of b_2 given b_3 . Then, the values are sorted for band selection. This method is termed as band ranking via entropy (BRE).

3.4. Drop Adjacent Bands

In some cases, nearby bands will have similar values, resulting in that some adjacent bands will be selected sequentially. To avoid this situation, a band will be dropped if its left band or right band has already been chosen. For instance, if b_2 has already been selected, b_1 and b_3 will be discarded even they have higher values than other bands. The BRECV and BRE methods with dropping adjacent bands are termed as BRECVD and BRED.

The CV values of each band could be directly sorted; this also could be used for band selection. This method is called as band ranking via CV (BRCV). Therefore, in this study, three band selection methods are proposed: BRCV, BRECV/BRECVD and BRE/BRED.

3.5. Time Complexity Analysis

Given a hyperspectral image $I \in \mathbb{R}^{h \times w \times c}$, the time complexity to calculate means of bands is $O(hwc)$, and the time complexity to calculate standard deviations of bands is $O(hwc)$. h means the height, w means the width, and c means the number of bands in a hyperspectral image. Operations on allocating values for bands, sorting values and dropping adjacent bands involve only means and standard deviations. Since c usually is far less than hwc , we could omit these operations. Hence, the time complexities of BRECV and BRECVD is linear.

Though we do not analyze the space complexity in detail, as only means and standard deviations of bands are used, the space complexity is also very low for BRECV and BRECVD. For the BRCV, BRE and BRED methods, their performances are not robust, which will be shown in the experimental results later. In this case, we do not analyze their computational complexity.

4. Results and Discussion

To verify the effectiveness of the proposed methods, classification experiments were implemented on the above-mentioned six different real-work hyperspectral images. Details of the experimental setup and results are shown in this section together with the discussion.

4.1. Comparison Methods

- Optimal neighborhood reconstruction (ONR) [32]

ONR selects bands by finding the optimal band combination to reconstruct the original data. A noise reducer was used to minimize the influence of noisy bands.

- Optimal clustering framework (OCF) [26]

OCF first finds the optimal clustering under some reasonable constraint and then ranks the clustering to effectively select bands on the clustering structure. OCF could also automatically determine the number of the required bands to choose.

- Enhanced fast density-peak-based clustering (EFDPC) [19,32]

EFDPC tries to find the cluster centers with properties as large local density and large intercluster distance. Large local density means a cluster should have as many points as possible. Large intercluster distance means different cluster centers should be far from each other. EFDPC ranks bands through weighting these two properties.

4.2. Classifiers

Support vector machine (SVM) and k-nearest neighborhood (KNN) classifiers are used to verify the classification performance for different band selection methods. We used the SVM and KNN classifiers provided by MATLAB R2019b. For SVM classifiers, the kernel function is “rbf”, and the coding method is “onevsall”. For KNN classifiers, k is 3. All the parameters were the same for each classification experiment.

For each data set, 10% of labeled samples of each class were randomly selected to train the classifiers, and the rest 90% of samples were used for testing. Each experiment was implemented 100 times individually, and these results were averaged to have a stable result. Overall accuracy (OA) curves were used to compare different band select methods. Similar to [26] and [32], we only used at most 30 bands in each experiment. Our code is available at <https://github.com/cvvsu/BRECV>.

4.3. Classification Results

- Indian Pines

From Figure 2, BRECV and BRECV D showed similar classification performance compared with OCF, ONR and EFDPC. EFDPC outperformed all other band selection methods on the Indian Pines data set. The BRCV, BRE and BRED methods did not show acceptable classification results on the Indian Pines data set. The dropping adjacent bands method did not show an obvious improvement for BRECV and improved the performance a little bit for BRE.

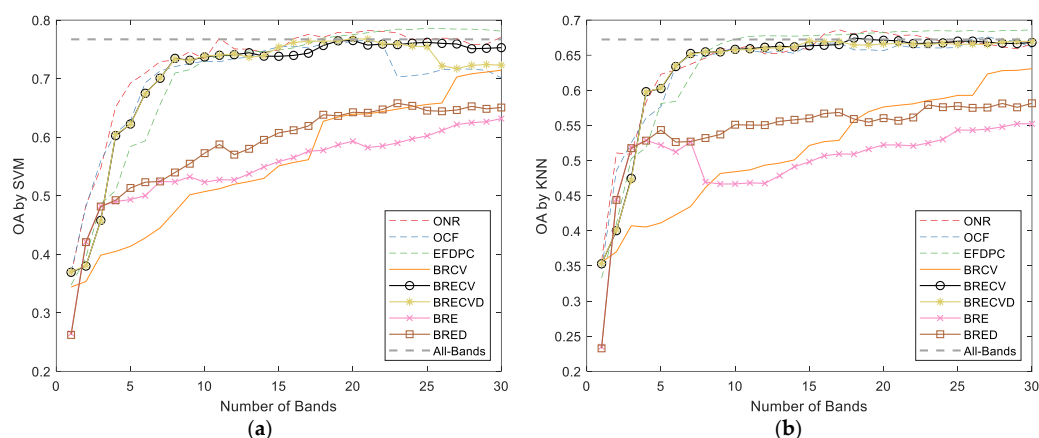


Figure 2. Overall accuracy (OA) curves produced by support vector machine (SVM) and k-nearest neighborhood (KNN) classifiers for the Indian Pines data set. (a) The OA curves produced by SVM classifier. (b) The OA curves produced by KNN classifier.

- KSC

From Figure 3, on the KSC data set, the classification results of the BRECV and BRECV D methods were quite similar compared with OCF and ONR and were better than EFDPC. The performances of BRE and BRED were also acceptable on the KSC data set. However, the BRCV method still had a lower performance on this data set. The dropping adjacent bands method improved the performance of the BRE method a lot.

- Pavia University

From Figure 4, on the Pavia University data set, the performance of BRECV D and BRED methods exceeded OCF methods when the number of selected bands greater than 25. BRECV outperformed the EFDPC method at first and was finally overtaken. BRCV still had the lowest performance compared with other methods.

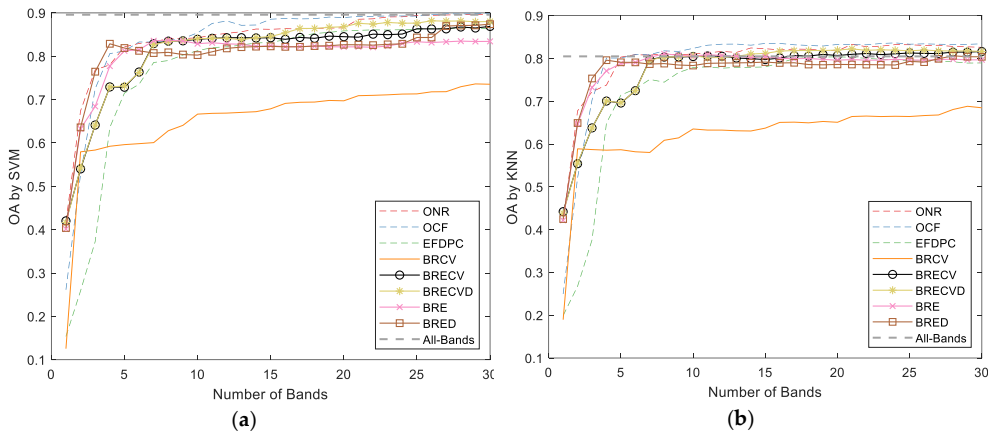


Figure 3. OA curves produced by SVM and KNN classifiers for the KSC data set. (a) The OA curves produced by SVM classifier. (b) The OA curves produced by KNN classifier.

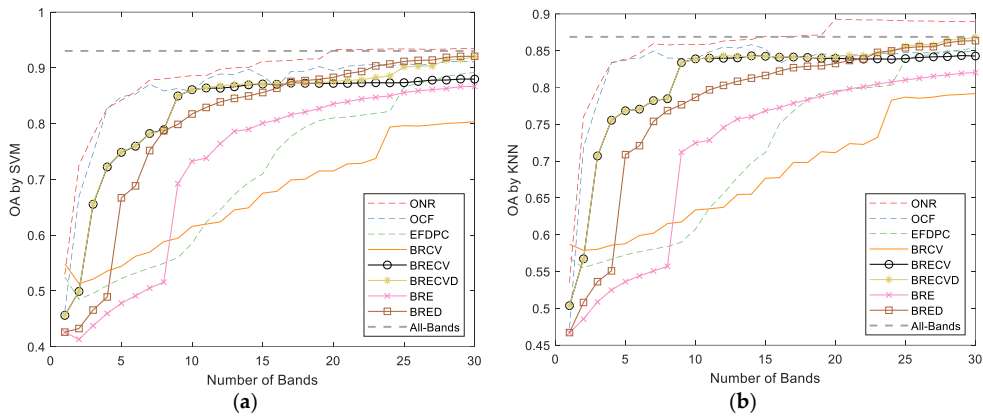


Figure 4. OA curves produced by SVM and KNN classifiers for the Pavia University data set. (a) The OA curves produced by SVM classifier. (b) The OA curves produced by KNN classifier.

- Botswana

From Figure 5, on the Botswana data set, except the BRCV method, all the other band selection methods showed similar classification performance.

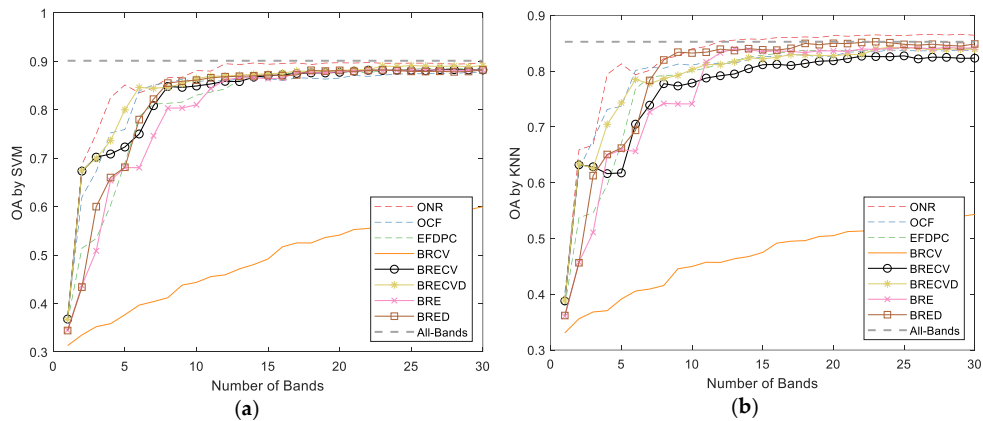


Figure 5. OA curves produced by SVM and KNN classifiers for the Botswana data set. (a) The OA curves produced by SVM classifier. (b) The OA curves produced by KNN classifier.

- Salinas

From Figure 6, on the Salinas data set, similar to the results on the Botswana data set, all methods except BRCV had qualitatively the same level of performance.

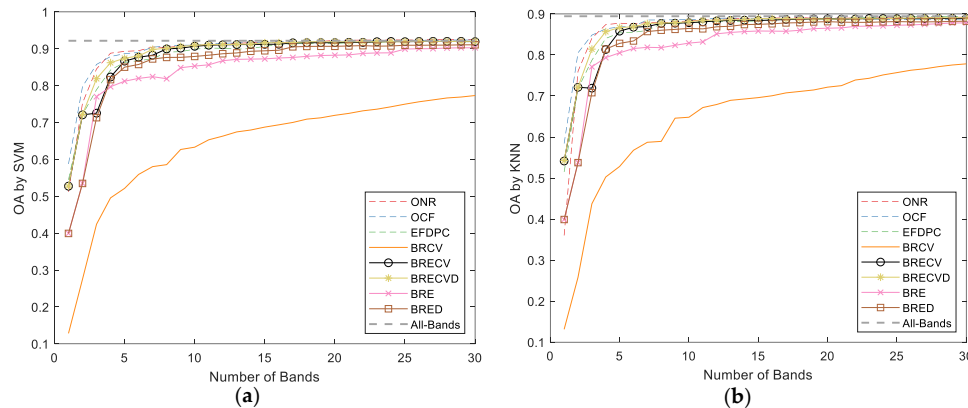


Figure 6. OA curves produced by SVM and KNN classifiers for the Salinas data set. (a) The OA curves produced by SVM classifier. (b) The OA curves produced by KNN classifier.

- Taita Hills

From Figure 7, on the Taita Hills data set, except the BRECV method, all the other methods showed similar classification performance. The classification result of the BRECV method was slightly worse compared to other methods. Interestingly, the BRCV method also verified itself on the Taita Hills data set. For the BRECV method, there were only 26 bands selected.

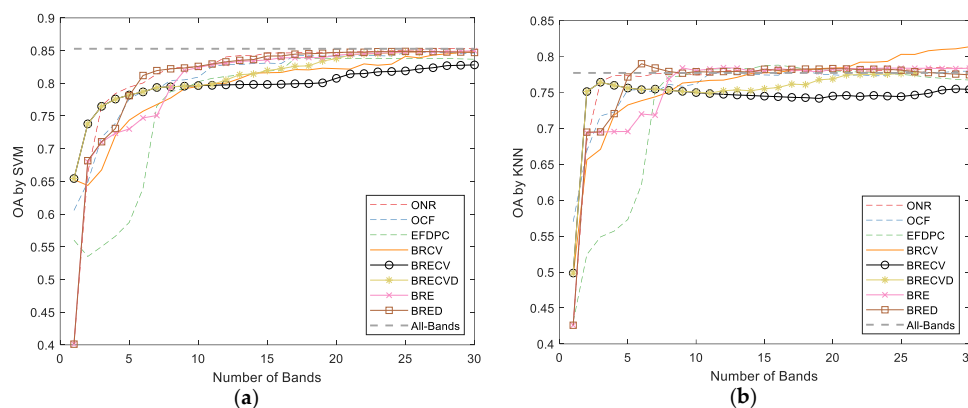


Figure 7. OA curves produced by SVM and KNN classifiers for the Taita Hills data set. (a) The OA curves produced by SVM classifier. (b) The OA curves produced by KNN classifier.

- Average OAs over different selected bands

Table 1 shows the indices of selected bands by BRECV on different data sets. Tables 2 and 3 show the average OAs by SVM classifier and KNN classifier over the 30 selected bands, respectively. Generally, BRECV and BRECV had better average OAs on different data sets compared with the EFDPC method. The performances of BRE and BRED methods were not robust, and their performances on the Indian Pines data set were quite bad. The BRCV method achieved relatively better performance on the Taita Hills data set, and on other data sets, the BRCV method did not have acceptable classification performance.

Table 1. Bands selected by band ranking via extended coefficient of variation with dropping adjacent bands (BRECVD) on different data sets.

Data Sets	Indexes of Selected 30 Bands
Indian Pines	13/18/32/90/26/65/117/163/190/181/161/168/193/178/128/49/173/51/170/70/186/88/183/84/72/145/196/131/176/165
KSC	79/81/60/73/70/49/23/28/54/43/67/51/30/20/18/15/143/41/45/87/12/56/9/58/47/6/75/65/39/90
Pavia University	3/12/63/26/20/60/39/47/100/88/42/31/90/53/49/93/5795/44/98/35/51/102/18/82/15/9/85/80/74
Botswana	121/43/51/7/97/16/41/65/4/18/92/129/100/70/56/87/21/137/53/34/123/132/26/95/126/58/24/90/119/106
Salinas	72/68/32/54/11/27/167/24/62/93/164/128/176/134/19/125/15/52/75/13/136/172/141/131/91/46/100/87/89/17
Taita Hills	(26 bands) 31/13/48/51/54/63/45/56/61/59/42/2/24/40/22/29/20/26/18/6/16/8/10/38/4/35

Table 2. OAs by SVM with standard deviation averaged over 30 selected bands.

	Indian Pines	Pavia University	Salinas	KSC	Botswana	Taita Hills
BRCV	0.5553 ± 0.1122	0.6699 ± 0.0967	0.6412 ± 0.1490	0.6547 ± 0.1110	0.4852 ± 0.0852	0.7944 ± 0.0567
BRE	0.5476 ± 0.0733	0.7229 ± 0.1641	0.8389 ± 0.1081	0.8001 ± 0.0873	0.7988 ± 0.1426	0.8005 ± 0.0899
BRED	0.5821 ± 0.0868	0.8005 ± 0.1526	0.8577 ± 0.1146	0.8061 ± 0.0866	0.8144 ± 0.1363	0.8096 ± 0.0881
BRECV	0.7018 ± 0.1093	0.8191 ± 0.1085	0.8818 ± 0.0841	0.8059 ± 0.1027	0.8261 ± 0.1067	0.7963 ± 0.0332
BRECVD	0.6984 ± 0.1082	0.8272 ± 0.1134	0.8883 ± 0.0792	0.8158 ± 0.1079	0.8393 ± 0.1051	0.8059 ± 0.0421
EFDPC	0.7037 ± 0.1216	0.7046 ± 0.1432	0.8815 ± 0.0769	0.7690 ± 0.1818	0.8086 ± 0.1325	0.7733 ± 0.1046
OCF	0.6974 ± 0.0886	0.8606 ± 0.0909	0.8974 ± 0.0640	0.8369 ± 0.1308	0.8275 ± 0.1053	0.8091 ± 0.0593
ONR	0.7242 ± 0.0954	0.8838 ± 0.0830	0.8939 ± 0.0795	0.8393 ± 0.0924	0.8559 ± 0.1038	0.8146 ± 0.0871

Table 3. OAs by KNN with standard deviation averaged over 30 selected bands.

	Indian Pines	Pavia University	Salinas	KSC	Botswana	Taita Hills
BRCV	0.5174 ± 0.0816	0.6813 ± 0.0754	0.6484 ± 0.1510	0.6229 ± 0.0881	0.4660 ± 0.0596	0.7603 ± 0.0623
BRE	0.5014 ± 0.0583	0.7132 ± 0.1217	0.8233 ± 0.1020	0.7788 ± 0.0741	0.7664 ± 0.1269	0.7542 ± 0.0700
BRED	0.5422 ± 0.0645	0.7760 ± 0.1117	0.8373 ± 0.1074	0.7723 ± 0.0710	0.7870 ± 0.1227	0.7606 ± 0.0676
BRECV	0.6341 ± 0.0799	0.8042 ± 0.0810	0.8587 ± 0.0739	0.7707 ± 0.0866	0.7656 ± 0.0984	0.7408 ± 0.0461
BRECVD	0.6332 ± 0.0795	0.8095 ± 0.0838	0.8650 ± 0.0691	0.7757 ± 0.0893	0.7883 ± 0.0939	0.7509 ± 0.0524
EFDPC	0.6393 ± 0.0880	0.7105 ± 0.1114	0.8610 ± 0.0748	0.7273 ± 0.1565	0.7780 ± 0.1145	0.7301 ± 0.0993
OCF	0.6330 ± 0.0705	0.8265 ± 0.0717	0.8710 ± 0.0568	0.7936 ± 0.1192	0.7933 ± 0.0891	0.7583 ± 0.0429
ONR	0.6401 ± 0.0690	0.8549 ± 0.0673	0.8636 ± 0.0983	0.7964 ± 0.0792	0.8210 ± 0.0967	0.7638 ± 0.0663

4.4. Discussion

From the above experiments, we found that the proposed BRECV and BRECVD methods achieved quite stable performances on all the data sets. The performances of entropy-based methods were not robust, and directly ranking the CV value of each band did not provide good results except on the Taita Hills data set. One possible reason is that mean and standard deviation provide two dimensions to investigate the relationships between nearby bands, while entropy and pure CV provide just one dimension.

Compared with the EFDPC method, the BRECV and BRECVD methods achieved better classification results. Only on the Indian Pines data set, both the tree methods had similar performance. Compared with the OCF, BRECV and BRECVD showed better performance on the Indian Pines data set and had a similar performance on the other five data sets. The ONR method outperformed all other methods on most data sets. In most cases, dropping adjacent bands for BRE and BRECV methods improved the classification performance.

Considering the fact that BRECV and BRECVD methods use neither clustering nor optimization methods, and select bands based only on means and standard deviations of bands, it is reasonable to believe these two methods are also useful for hyperspectral band selection. Moreover, the physical meaning is clear for bands selected by BRECV and BRECVD.

From the classification performance of the proposed methods, we could say that the bands selected by our methods were representative and informative since these selected bands from a hyperspectral image achieved qualitatively the same level of classification performance, compared with the results of some state-of-the-art band selection methods and the whole hyperspectral data set. Figure 8 illustrates the scatter plots of means and standard deviations of bands in each data set; the selected bands are also shown in filled red colors. Figure 8 illustrates a concrete sense of the relative locations of selected bands in each hyperspectral data set. For the KSC data set, the selected bands were mainly located in a small band region, which is similar to the results of ONR. In ONR, the selected bands from the KSC data set only covered $\frac{2}{5}$ of the whole spectrum.

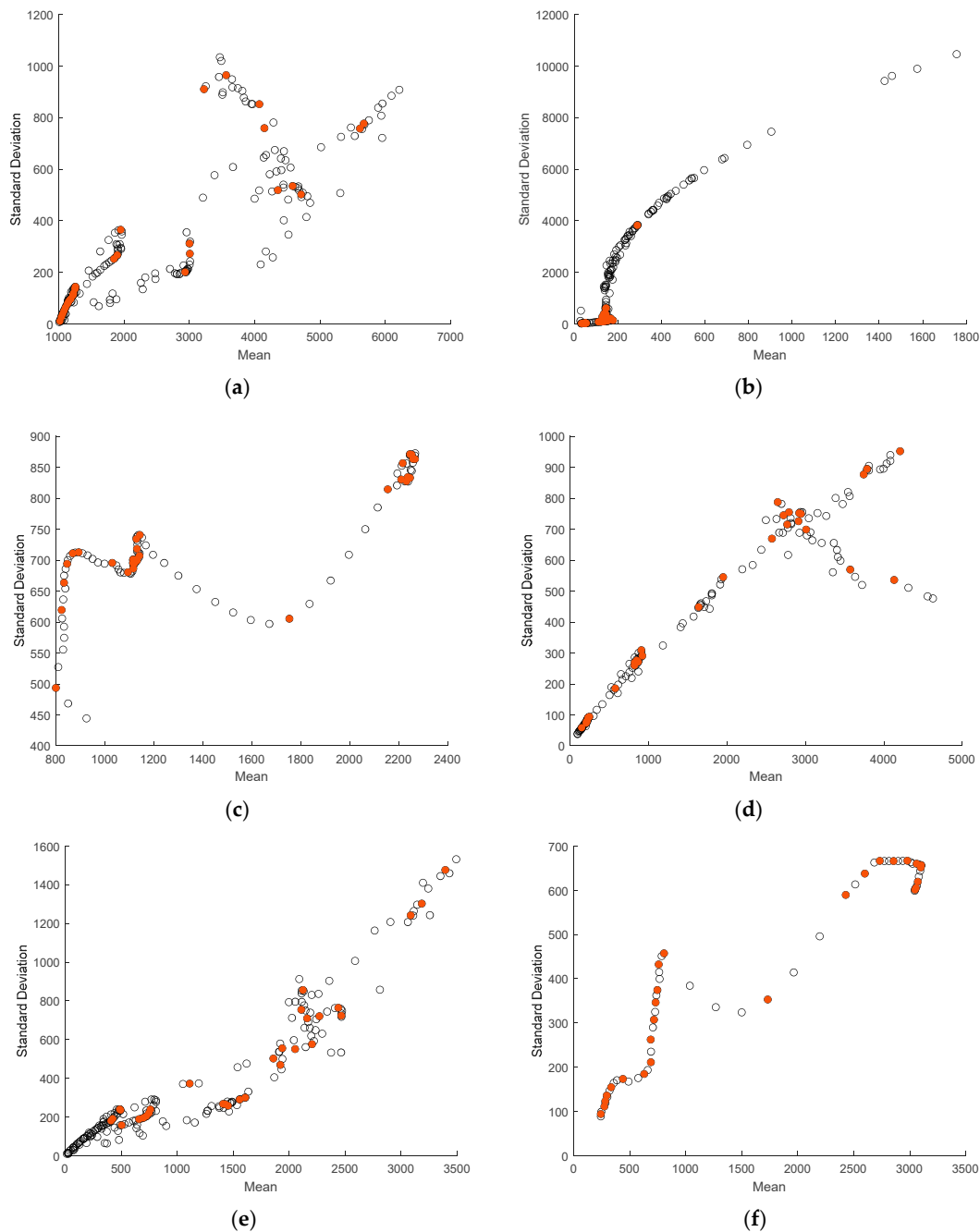


Figure 8. Scatter plot of means and standard deviations for each data set. Every circle in each figure indicates a band and the red circles represent the selected bands by the BRECV method. (a) Indian Pines. (b) KSC. (c) Pavia University. (d) Botswana. (e) Salinas. (f) Taita Hills.

5. Conclusions

This study investigated the relationship between nearby bands in a hyperspectral data set and proposes a criterion for band ranking. An extended matrix based on coefficient of variation was used to help study changes of means and standard deviations. Finally, several band ranking methods are presented for hyperspectral band selection according to the relationships between nearby bands. The proposed methods were quite efficient, as the methods did not need to face the large-volume problem. Compared with other band selection methods, the proposed methods obtained qualitatively the same level of classification performance.

Author Contributions: P.S. designed the research and analyzed the results. P.K.E.P. provided the Taita Hills data set and related field data. P.K.E.P. and S.T. provided advice for the preparation and revision of the paper. All authors have read and agreed to the published version of the manuscript.

Funding: The work is supported by the MegaSense research programme of the University of Helsinki, the City of Helsinki Innovation Fund, Business Finland, the European Commission through the Urban Innovative Action Healthy Outdoor Premises for Everyone (project No. UIA03-240). The capture of the airborne data over Taita Hills was funded by the Ministry for Foreign Affairs of Finland for the project, “Climate Change Impacts on Ecosystem Services and Food Security in Eastern Africa - Increasing Knowledge, Building Capacity and Developing Adaptation Strategies”. Open access funding provided by University of Helsinki.

Acknowledgments: The authors wish to acknowledge CSC–IT Center for Science, Finland, for computational resources and Taita Research Station of the University of Helsinki for logistics in the fieldwork in Kenya.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Marinelli, D.; Bovolo, F.; Bruzzone, L. A Novel Change Detection Method for Multitemporal Hyperspectral Images Based on Binary Hyperspectral Change Vectors. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 4913–4928. [[CrossRef](#)]
2. Geng, X.; Sun, K.; Ji, L. Band Selection for Target Detection in Hyperspectral Imagery Using Sparse CEM. *Remote Sens. Lett.* **2014**, *5*, 1022–1031. [[CrossRef](#)]
3. Piironen, R.; Fassnacht, F.E.; Heiskanen, J.; Maeda, E.; Mack, B.; Pellikka, P. Invasive Tree Species Detection in the Eastern Arc Mountains Biodiversity Hotspot Using One Class Classification. *Remote Sens. Environ.* **2018**, *218*, 119–131. [[CrossRef](#)]
4. Li, C.; Gao, L.; Wu, Y.; Zhang, B.; Plaza, J.; Plaza, A. A Real-Time Unsupervised Background Extraction-Based Target Detection Method for Hyperspectral Imagery. *J. Real Time Image Process.* **2018**, *15*, 597–615. [[CrossRef](#)]
5. Sellami, A.; Farah, M.; Farah, I.R.; Solaiman, B. Hyperspectral Imagery Semantic Interpretation Based on Adaptive Constrained Band Selection and Knowledge Extraction Techniques. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 1337–1347. [[CrossRef](#)]
6. Camps-Valls, G.; Tuia, D.; Bruzzone, L.; Benediktsson, J.A. Advances in Hyperspectral Image Classification: Earth Monitoring with Statistical Learning Methods. *IEEE Signal Process. Mag.* **2014**, *31*, 45–54. [[CrossRef](#)]
7. Cao, X.; Wei, C.; Ge, Y.; Feng, J.; Zhao, J.; Jiao, L. Semi-Supervised Hyperspectral Band Selection Based on Dynamic Classifier Selection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 1289–1298. [[CrossRef](#)]
8. Piironen, R.; Heiskanen, J.; Möttö, M.; Pellikka, P. Classification of Crops across Heterogeneous Agricultural Landscape in Kenya Using AisaEAGLE Imaging Spectroscopy Data. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *39*, 1–8. [[CrossRef](#)]
9. Sun, W.; Du, Q. Hyperspectral Band Selection: A Review. *IEEE Geosci. Remote Sens. Mag.* **2019**, *7*, 118–139. [[CrossRef](#)]
10. Wei, X.; Zhu, W.; Liao, B.; Cai, L. Scalable One-Pass Self-Representation Learning for Hyperspectral Band Selection. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 4360–4374. [[CrossRef](#)]
11. Zhai, H.; Zhang, H.; Zhang, L.; Li, P. Laplacian-Regularized Low-Rank Subspace Clustering for Hyperspectral Image Band Selection. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 1723–1740. [[CrossRef](#)]
12. Zheng, X.; Yuan, Y.; Lu, X. Hyperspectral Image Denoising by Fusing the Selected Related Bands. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 2596–2609. [[CrossRef](#)]

13. Das, S.; Bhattacharya, S.; Routray, A.; Deb, A.K. Band Selection of Hyperspectral Image by Sparse Manifold Clustering. *IET Image Process.* **2019**, *13*, 1625–1635. [[CrossRef](#)]
14. Chang, C.I.; Liu, K.H. Progressive Band Selection of Spectral Unmixing for Hyperspectral Imagery. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 2002–2017. [[CrossRef](#)]
15. Sumarsono, A.; Du, Q. Estimation of Number of Signal Subspaces in Hyperspectral Imagery Using Low-Rank Subspace Representation. In *Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote Sensing*; IEEE Computer Society: Lausanne, Switzerland, 2014; pp. 1–4. [[CrossRef](#)]
16. Jimenez, L.O.; Landgrebe, D.A. Supervised Classification in High-Dimensional Space: Geometrical, Statistical, and Asymptotical Properties of Multivariate Data. *IEEE Trans. Syst. Man Cybern. Part. C Appl. Rev.* **1998**, *28*, 39–54. [[CrossRef](#)]
17. Sun, W.; Du, Q. Graph-Regularized Fast and Robust Principal Component Analysis for Hyperspectral Band Selection. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 3185–3195. [[CrossRef](#)]
18. Zhang, W.; Li, X.; Zhao, L. Discovering the Representative Subset with Low Redundancy for Hyperspectral Feature Selection. *Remote Sens.* **2019**, *11*, 1341. [[CrossRef](#)]
19. Jia, S.; Tang, G.; Zhu, J.; Li, Q. A Novel Ranking-Based Clustering Approach for Hyperspectral Band Selection. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 88–102. [[CrossRef](#)]
20. Wang, Q.; Lin, J.; Yuan, Y. Salient Band Selection for Hyperspectral Image Classification via Manifold Ranking. *IEEE Trans. Neural Networks Learn. Syst.* **2016**, *27*, 1279–1289. [[CrossRef](#)]
21. Geng, X.; Sun, K.; Ji, L.; Tang, H.; Zhao, Y. Joint Skewness and Its Application in Unsupervised Band Selection for Small Target Detection. *Sci. Rep.* **2015**, *5*, 1–9. [[CrossRef](#)]
22. Geng, X.; Sun, K.; Ji, L.; Zhao, Y. A Fast Volume-Gradient-Based Band Selection Method for Hyperspectral Image. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 7111–7119. [[CrossRef](#)]
23. Du, Q.; Yang, H. Similarity-Based Unsupervised Band Selection for Hyperspectral Image Analysis. *IEEE Geosci. Remote Sens. Lett.* **2008**, *5*, 564–568. [[CrossRef](#)]
24. Dos Santos, L.C.B.; Guimaraes, S.J.F.; Dos Santos, J.A. Efficient Unsupervised Band Selection through Spectral Rhythms. *IEEE J. Sel. Top. Signal. Process.* **2015**, *9*, 1016–1025. [[CrossRef](#)]
25. Yuan, Y.; Lin, J.; Wang, Q. Dual-Clustering-Based Hyperspectral Band Selection by Contextual Analysis. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 1431–1445. [[CrossRef](#)]
26. Wang, Q.; Zhang, F.; Li, X. Optimal Clustering Framework for Hyperspectral Band Selection. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 5910–5922. [[CrossRef](#)]
27. Sun, K.; Geng, X.; Ji, L. A New Sparsity-Based Band Selection Method for Target Detection of Hyperspectral Image. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 329–333. [[CrossRef](#)]
28. Yuan, Y.; Zhu, G.; Wang, Q. Hyperspectral Band Selection by Multitask Sparsity Pursuit. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 631–644. [[CrossRef](#)]
29. Zhan, Y.; Hu, D.; Xing, H.; Yu, X. Hyperspectral Band Selection Based on Deep Convolutional Neural Network and Distance Density. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 2365–2369. [[CrossRef](#)]
30. Cai, Y.; Liu, X.; Cai, Z. BS-Nets: An End-to-End Framework for Band Selection of Hyperspectral Image. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 1969–1984. [[CrossRef](#)]
31. Baumgardner, M.F.; Biehl, L.L.; Landgrebe, D.A. 220 Band AVIRIS Hyperspectral Image Data Set: June 12, 1992 Indian Pine Test Site 3. *Purdue Univ. Res. Repos.* **2015**. [[CrossRef](#)]
32. Wang, Q.; Zhang, F.; Li, X. Hyperspectral Band Selection via Optimal Neighborhood Reconstruction. *IEEE Trans. Geosci. Remote Sens.* **2020**, 1–12. [[CrossRef](#)]

