Relevance Assessment of Spectral Bands for Land Cover and Land Use Classification: A Case Study Involving Multispectral Sentinel-2-like and Hyperspectral Data

MARTIN WEINMANN[1] & UWE WEIDNER[1]

Abstract: In this paper, we comprehensively investigate the potential of both multispectral and hyperspectral data for land cover and land use classification and, in this context, we particularly focus on comparatively assessing the relevance of involved spectral bands for the considered classification task. We present a framework which comprises different data-driven techniques for assessing the relevance of spectral bands with respect to the given classification task. Based on the assessed feature relevance, our framework allows selecting relevant features as the basis for classification for which a Random Forest classifier is used. We also describe a transformation of given hyperspectral data to multispectral Sentinel-2-like data that are commonly used for large-scale land cover and land use classification. For performance evaluation, we provide classification results achieved for two standard benchmark datasets representing an urban area and an agricultural area, respectively.

1 Introduction

Land cover and land use classification is commonly performed on the basis of aerial or satellite imagery representing either multispectral or hyperspectral data. In the case of multispectral data, the number of spectral bands is relatively low and the spectral bands are wide. In contrast, the number of spectral bands is relatively high for the case of hyperspectral data, while the spectral bands themselves are narrow. Consequently, a more characteristic description of the spectral properties of the Earth's surface can be expected for the latter case in comparison to the use of multispectral data.

In contrast to multispectral data, where the neighboring spectral bands are well-separated by a sufficiently large margin across wavelengths, the spectral bands of hyperspectral data are directly next to each other so that the acquired reflectance values of neighboring spectral bands tend to be strongly correlated. This kind of redundancy typically has a negative impact on classification results, so that approaches for dimensionality reduction or band selection are commonly used. While dimensionality reduction techniques transform the given data to a new space, band selection techniques allow conclusions about relationships with respect to physical properties as they retain a subset of the original spectral bands which, in turn, can further be used for conclusions regarding a diversity of environmental applications.

In this paper, we comprehensively investigate the potential of both multispectral and hyperspectral data, and we thereby comparatively assess the relevance of involved spectral bands for the considered classification task. We present a framework which comprises different data-driven techniques for assessing the relevance of spectral bands with respect to the given classification

[1] Karlsruhe Institute of Technology, Institute of Photogrammetry and Remote Sensing, Englerstraße 7, D-76131 Karlsruhe, E-Mail: [Martin.Weinmann, Uwe.Weidner]@kit.edu

task. This includes both classifier-dependent techniques and classifier-independent techniques. The classifier-dependent techniques comprise 1) a sequential forward selection of spectral bands based on sequentially training a Linear Discriminant Analysis (LDA) classifier based on different sets of spectral bands and 2) the Mean Decrease in Permutation Accuracy (MDPA), a measure assessed during the training of a Random Forest classifier. The classifier-independent techniques comprise 1) a general relevance metric taking into account the relations between the given spectral bands and the defined classes to identify relevant spectral bands and 2) an approach taking into account both the relations between spectral bands and classes to identify relevant spectral bands and the relations among spectral bands to identify and discard redundant spectral bands. Based on the selected bands, we finally perform a pixel-based classification using a Random Forest classifier to quantify the effect on the achieved classification results.
As the available datasets contain either hyperspectral or multispectral data, we consider commonly used hyperspectral benchmark datasets for which a semantic labeling is available on a per-pixel basis, and we adequately transform these hyperspectral datasets to multispectral Sentinel-2-like data using the Sentinel-2 Spectral Response Functions (S2-SRFs). For performance evaluation, we consider two classification tasks, one focusing on the semantic interpretation of an urban area and one focusing on the semantic interpretation of an agricultural area. For both classification tasks, the data have been acquired from an airborne platform during a low-altitude flight campaign.
After briefly summarizing related work (Section 2), we present our framework in detail (Section 3) and we demonstrate its performance on two benchmark datasets (Section 4). This is followed by a discussion of the derived results (Section 5). Finally, we provide concluding remarks and suggestions for future work (Section 6).

2 Related Work

The semantic interpretation of hyperspectral imagery may easily be achieved via a pixel-wise classification relying on the reflectance values across all considered spectral bands. In this context, the reflectance values are defined as features and used as entries of a respective feature vector, while the classification may be based on well-known standard classifiers such as a Support Vector Machine (SVM) classifier (MELGANI & BRUZZONE 2004; CHI et al. 2008) or a Random Forest classifier (HAM et al. 2005; JOELSSON et al. 2005).
However, particularly for such high-dimensional classification tasks, it can often be observed that above a certain value a further increase of the number of involved features results in a (typically significant) decrease in predictive accuracy (MELGANI & BRUZZONE 2004; KELLER et al. 2016; BRADLEY et al. 2018). This effect is commonly referred to as the Hughes phenomenon (HUGHES 1968) and arises from the joint consideration of more or less relevant features and possibly even irrelevant or redundant features with respect to the considered classification task.
To address the Hughes phenomenon, either dimensionality reduction techniques or feature selection techniques are commonly applied. The latter have the advantage that they retain a meaningful subset of the original features with respect to the given task (i.e. the reflectance values corresponding to specific spectral bands with known wavelength, which may further be exploited for conclusions regarding feature engineering and for conclusions regarding environmental sciences), while discarding less relevant and/or redundant features. This, in turn, typically allows

gaining predictive accuracy and improving computational efficiency with respect to both time and memory consumption (GUYON & ELISSEEFF 2003; SAEYS et al. 2007; ZHAO et al. 2010). In the context of classifying hyperspectral data, such feature selection techniques resulting in the consideration of reflectance values across specific spectral bands have been used in several investigations (MELGANI & BRUZZONE 2004; LE BRIS et al. 2014; CHEHATA et al. 2014; KELLER et al. 2016; BRADLEY et al. 2018). Such techniques may also directly allow assessing the importance of single spectral bands for land cover and land use classification (LE BRIS et al. 2014; KELLER et al. 2016; WEINMANN & WEIDNER 2018). Typically, feature selection techniques are categorized with respect to filter-based methods, wrapper-based methods and embedded methods (GUYON & ELISSEEFF 2003; SAEYS et al. 2007; ZHAO et al. 2010; WEINMANN 2016).

The filter-based methods rely on evaluating relations between features and classes and possibly also among features. In this context, the relations are described via a score function which is directly applied to the given training data. When only focusing on relations between features and classes (univariate filter-based feature selection), the relations are quantified by comparing the values of a feature across all data points with the respective class labels, e.g. via the correlation coefficient, Gini index, Fisher score, or information gain. This allows ranking the features with respect to their relevance. When taking into account both feature–class relations and feature–feature relations (multivariate filter-based feature selection), it is also possible to remove redundancy to a certain degree. A respective approach is represented by Correlation-based Feature Selection (HALL 1999). In general, filter-based methods are classifier-independent and thus typically result in simplicity and efficiency.

The wrapper-based methods rely on the interaction with a classifier to select features based on their suitability for classification. This may be achieved via Sequential Forward Selection (SFS) where, beginning with an empty feature subset, it is successively tested which feature can be added to the feature subset so that the predictive accuracy of the classifier increases the most. Alternatively, a Sequential Backward Elimination (SBE) may be conducted where, beginning with the set of all features, it is successively tested which feature can be removed from the feature subset so that the predictive accuracy of the classifier is decreased the least. In general, however, the exhaustive interaction with a classifier tends to cause a high computational burden.

The embedded methods rely on the interaction with a classifier which provides the capability to internally select the most relevant features during the training phase, e.g. a Random Forest classifier (BREIMAN 2001). In contrast to wrapper-based methods, the involved classifier has to be trained only once to allow concluding about the relevance of involved features and the computational effort is therefore still acceptable in comparison to wrapper-based methods.

Besides such feature selection techniques, a transformation of hyperspectral data to multispectral data may be applied, where particularly Sentinel-2-like data are of interest as they are systematically acquired for Earth observation with short revisit times (SPOTO et al. 2012). Such a transformation has been introduced for simulating Sentinel-2 and other multispectral imagery in general (THONFELD et al. 2012), for assessing agricultural land use based on simulated Sentinel-2 data (ELBERTZHAGEN et al. 2012) and for simulating Sentinel-2 products that are relevant for geological and soil analyses (VAN DER MEER et al. 2014). Furthermore, such a transformation allows reasoning about the potential of multispectral data in comparison to hyperspectral data for land cover and land use classification (WEINMANN et al. 2018).

3 Methodology

An overview of the applied methodology is shown in Fig. 1. Given a hyperspectral data cube, we first derive the original data representation and its transformation to multispectral Sentinel-2-like data. Subsequently, we focus on feature relevance assessment to draw conclusions about the relevance of considered spectral bands with respect to the classification task. In this context, we involve 1) a wrapper-based feature selection method relying on a Linear Discriminant Analysis (LDA) classifier, 2) an embedded feature selection method relying on the Mean Decrease in Permutation Accuracy (MDPA) assessed with a Random Forest (RF) classifier (BREIMAN 2001; LIAW & WIENER 2002), 3) a univariate filter-based feature selection method relying on a general relevance metric (GRM) proposed in (WEINMANN 2016) and 4) a multivariate filter-based feature selection method relying on the Correlation-based Feature Selection (CFS) method (HALL 1999). Finally, a supervised classification based on a Random Forest (RF) classifier (BREIMAN 2001) is performed. In the following, we first focus on the two options for the data representation (Section 3.1). Subsequently, we explain the four applied data-driven techniques for feature relevance assessment (Section 3.2) and the used approach for supervised classification (Section 3.3) with more details.

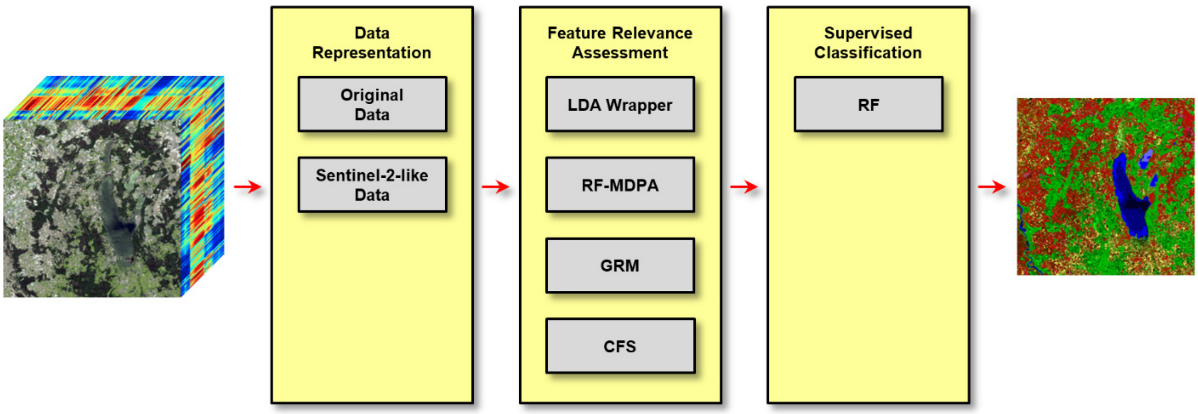

Fig. 1: Overview of our framework with different components for data representation, feature relevance assessment and supervised classification.

3.1 Data Representation

In our work, we consider the original representation of hyperspectral data (Section 3.1.1) and its transformation to multispectral data in the form of Sentinel-2-like data (Section 3.1.2) as input for subsequent processing steps.

3.1.1 Original Data Representation

The straightforward approach consists in using the given representation of the data, where we simply concatenate the reflectance values across all given spectral bands to obtain the respective feature vectors on a per-pixel basis.

3.1.2 Transformation of Hyperspectral Data to Sentinel-2-like Data

We also transform the original representation of hyperspectral data to multispectral Sentinel-2-like data (WEINMANN et al. 2018). In this context, we first identify those spectral bands of the hyperspectral data which are within the 13 spectral bands corresponding to Sentinel-2 data. The reflectance values corresponding to these hyperspectral bands, in turn, are used to derive a reflectance value for each considered spectral band of the Sentinel-2 data. More specifically, we consider the weighted mean of the reflectance values, where the weights are determined via linear interpolation based on the Sentinel-2 Spectral Response Functions (S2-SRFs) which are depicted in Fig. 2.

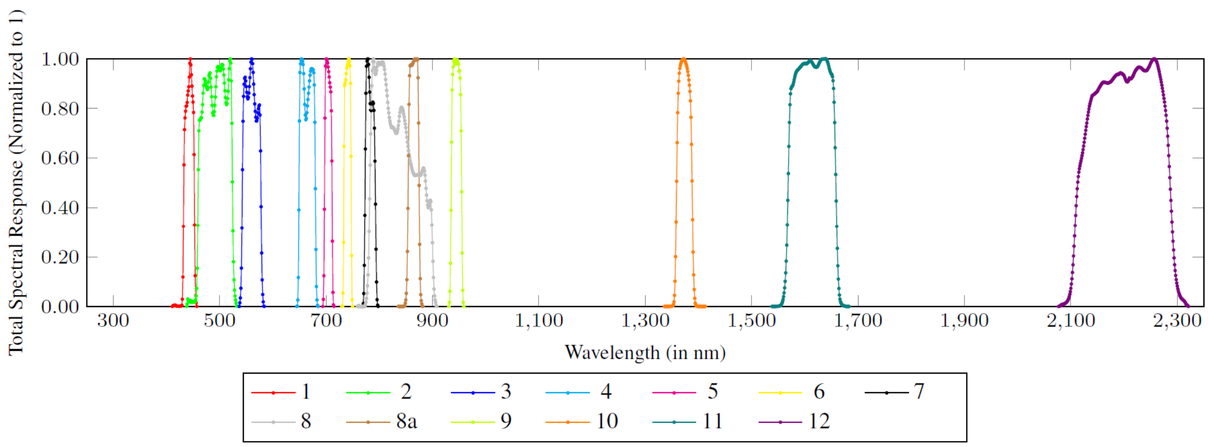

Fig. 2: Visualization of the Sentinel-2 Spectral Response Functions (S2-SRFs), i.e. the measured spectral responses for each band of the Sentinel-2 MultiSpectral Instrument (MSI) (WEINMANN et al. 2018).

However, for a classification task focusing on the use of Sentinel-2 data to derive thematic maps with respect to a variety of land cover and land use classes, not all of the 13 given spectral bands contain valuable information (WEINMANN & WEIDNER 2018). The reflectance values corresponding to the spectral bands B_1 (center wavelength of 443 nm), B_9 (945 nm) and B_{10} (1375 nm) are not considered, because they correspond to parts of the spectrum where the atmospheric transmission is low, e.g. due to ozone (O_3), oxygen (O_2) or water vapor (H_2O) which strongly affect the atmospheric transmissivity at specific wavelengths. Furthermore, the reflectance value corresponding to the spectral band B_8 (842 nm) is not considered, because it is overlapping with the spectral band B_{8a} but much wider and less characteristic. Consequently, we only use those reflectance values corresponding to the spectral bands B_2 (490 nm), B_3 (560 nm), B_4 (665 nm), B_5 (705 nm), B_6 (740 nm), B_7 (783 nm), B_{8a} (865 nm), B_{11} (1610 nm) and B_{12} (2190 nm).

3.2 Feature Relevance Assessment

Given the different data representations, we proceed with assessing the relevance of features (i.e. spectral bands) via an exhaustive interaction with a classifier (Section 3.2.1), via a classifier-internal metric (Section 3.2.2), and via classifier-independent schemes (Sections 3.2.3 and 3.2.4).

3.2.1 Wrapper-Based Sequential Forward Selection of Spectral Bands

A classifier-dependent feature relevance assessment may be achieved by sequential forward selection via interaction with a classifier. Thus, we derive a feature ranking by starting with an empty set of features and successively adding the feature that improves the predictive accuracy of the involved classifier the most when added to the given feature set. This procedure is repeated until all features have been added. Consequently, we obtain the rank r of a feature f_i with $i = 1, \dots, N_f$ and we define the relevance $R \in [0,1]$ of the feature f_i according to

$$R(f_i) = 1 - \frac{r(f_i) - 1}{N_f - 1}$$

so that relevant features are characterized by values close to 1, while irrelevant features are characterized by values close to 0. In this context, we take into account that exhaustive comparisons of feature vectors via the Euclidean distance as performed with a Nearest Neighbor (NN) classifier result in a high computational burden, particularly when considering high-dimensional feature vectors. Furthermore, we take into account that some classifiers such as a Support Vector Machine (SVM) classifier or a Random Forest (RF) classifier require an additional tuning of internal parameters during the training process (i.e. the kernel width and the cost parameter in case of the SVM classifier, or the number of decision trees, the maximum tree depth, etc. in case of the RF classifier), which also results in a high computational effort. Consequently, we involve a classifier which does neither require such a tuning nor rely on the evaluation of Euclidean distances between feature vectors. More specifically, the chosen classifier is represented by a Linear Discriminant Analysis (LDA) classifier relying on the principle of probabilistic learning. In the training stage, it is assumed that the instances of different classes follow a Gaussian distribution in the feature space. Accordingly, the training of the LDA classifier consists in fitting a multivariate Gaussian distribution to the given training data, where the parameters of a Gaussian distribution have to be estimated for each class. Due to a lack of knowledge about the behavior of single classes in the feature space, the same covariance matrix is assumed for each class so that only the means vary for the different classes. In the prediction stage, the class probabilities are evaluated for each feature vector to be classified and the label of the class with the maximum probability is selected.

3.2.2 Feature Ranking via an Embedded Method

To directly assess the relevance of the given spectral bands with respect to the considered classification task, we use a Random Forest classifier (BREIMAN 2001) which represents an embedded method for feature selection, since it allows assessing feature relevance via the consideration of the Mean Decrease in Permutation Accuracy (MDPA). In this regard, the main idea consists in training each decision tree of the Random Forest classifier on a randomly chosen subset of the training data, i.e. a bootstrap sample, and then performing the prediction for the data which are not in the bootstrap sample, i.e. the out-of-bag (OOB) data (LIAW & WIENER 2002). Subsequently, the OOB predictions of all trees are aggregated and an error rate is derived. To estimate the relevance of a feature, it is tested how much the prediction error increases if OOB data for that feature is randomly permuted, while all others are left unchanged. Using the MDPA as feature importance and sorting the features accordingly, we obtain the rank r of a feature f_i with

$i = 1, \dots, N_f$ and we define the relevance $R \in [0,1]$ of the feature f_i as done for the wrapper-based sequential forward selection method in the previous section.

3.2.3 Feature Ranking via a General Relevance Metric

To directly assess the relevance of the given spectral bands with respect to the considered classification task, we also apply a classifier-independent approach for feature ranking via a general relevance metric (GRM) (WEINMANN 2016). This metric is a compound metric defined on the basis of filter-based feature selection methods and makes use of the given training data. More specifically, each of the filter-based methods relies on a score function which evaluates relations between features and classes to distinguish between relevant and irrelevant features. For each feature, its values across all feature vectors given in the training data are concatenated to a vector whose "correlation" with the corresponding label vector is evaluated and represented by a real-valued score. In this context, different score functions may be used which address different intrinsic properties of the given training data (e.g. distance, information, dependency or consistency). To achieve a robust feature relevance assessment taking into account different intrinsic properties of the given training data, we involve seven score functions (Pearson correlation coefficient, Fisher score, Gini index, information gain, χ^2-test, *t*-test and ReliefF) and derive a separate ranking with respect to each score function. For more details about these score functions, we refer to (WEINMANN 2016) and references therein. As a result, we get the rank r of a feature f_i with $i = 1, \dots, N_f$ given the score function s_j. Averaging the ranks derived for the feature f_i across all $n_s = 7$ score functions yields the mean rank $\bar{r}$:

$$\bar{r}(f_i) = \frac{1}{n_s} \sum_{j=1}^{n_s} r(f_i|s_j)$$

On this basis, the relevance R of the feature f_i is defined according to

$$R(f_i) = 1 - \frac{\bar{r}(f_i) - 1}{N_f - 1}$$

with $R \in [0,1]$.

3.2.4 Filter-Based Selection of a Meaningful Subset of Spectral Bands

To directly derive a meaningful subset of the given spectral bands, we focus on feature subset selection. For this purpose, we use a filter-based feature selection method which is referred to as Correlation-based Feature Selection (CFS) (HALL 1999). This method takes into account 1) the correlation between features and classes to identify relevant features and 2) the correlation among features to identify and discard redundant features. More specifically, CFS exploits the average correlation $\bar{\rho}_{FC}$ between features and classes as well as the average correlation $\bar{\rho}_{FF}$ among classes to evaluate the relevance R of a feature subset comprising n features:

$$R = \frac{n\bar{\rho}_{FC}}{\sqrt{n + n(n-1)\bar{\rho}_{FF}}}$$

Here, the correlation metric is defined via the symmetrical uncertainty SU (HALL 1999) with

$$SU(X, C) = \frac{2\, MI(X, C)}{E(X) + E(C)}$$

where X and C represent random variables for the features and classes, respectively. The term $MI(X, C)$ represents the mutual information between X and C, while the terms $E(X)$ and $E(C)$ represent Shannon entropies indicating the distributions of feature values and classes, respectively. Deriving a suitable feature subset thus corresponds to maximizing the relevance R over the feature subset space in an iterative manner. This means that, given an initial feature subset, either a feature is added to the feature subset (forward selection) or a feature is removed from the feature subset (backward elimination) until the relevance R converges to a stable maximum.

3.3 Supervised Classification

For classification, we focus on a standard supervised classification based on a Random Forest (RF) classifier (BREIMAN 2001) which relies on the principle of ensemble learning. In the training stage, an ensemble of randomly trained decision trees is generated via bootstrap aggregating ("bagging") (BREIMAN 1996), i.e. various subsets of the training data are randomly drawn with replacement and an individual decision tree is trained for each subset. In the prediction stage, for a new feature vector to be classified, each decision tree casts a vote for one of the defined classes and the majority vote is selected to obtain the most probable class label. To select suitable values for the internal parameters of the Random Forest classifier (i.e. for the number of involved decision trees, the maximum tree depth, etc.), we conduct a grid search on a suitable subspace.

4 Experimental Results

In the following, we consider two classification tasks, one focusing on the semantic interpretation of an urban area and one focusing on the semantic interpretation of an agricultural area. First, we describe the used datasets (Section 4.1). Subsequently, we explain the conducted experiments and present the achieved results (Section 4.2).

4.1 Datasets

For our experiments, we use two benchmark datasets representing an urban area and an agricultural area. Both datasets are publicly available in a repository of hyperspectral remote sensing scenes: http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes

4.1.1 Pavia Centre Dataset

The Pavia Centre Dataset was acquired with the Reflective Optics System Imaging Spectrometer (ROSIS) during a low-altitude flight campaign over the city of Pavia, Italy. The considered scene corresponds to an urban area and the acquired data are represented in the form of two images with a size of 1096×223 pixels and 1096×492 pixels, respectively. Each pixel corresponds to an area of 1.3 m×1.3 m and comprises hyperspectral information on 102 spectral bands. In total, the Pavia Centre Dataset consists of about 784k pixels of which 7,456 have been labeled with respect to 9 semantic classes as shown in Fig. 3, while no labels are provided for the remaining pixels.

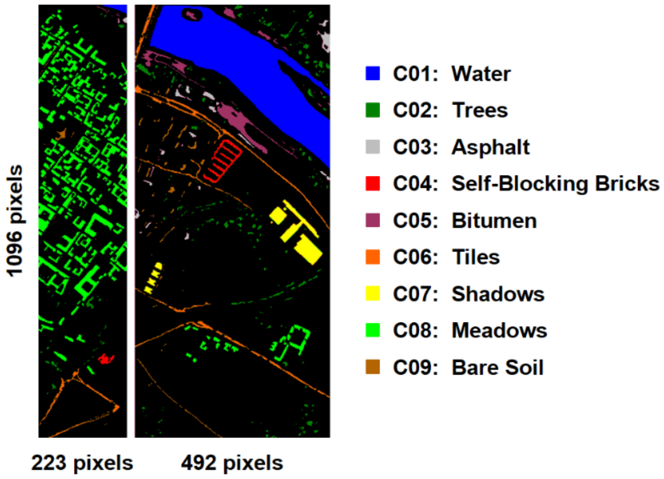

Fig. 3: Reference labels for the Pavia Centre Dataset: each pixel is characterized by reflectance values on 102 spectral bands. Unlabeled pixels are indicated in black.

4.1.2 Salinas Dataset

The Salinas Dataset was acquired with the Airborne Visible / Infrared Imaging Spectrometer (AVIRIS) during a low-altitude flight campaign over Salinas Valley in California, USA. The considered scene corresponds to an agricultural area which is mainly characterized by vegetables, corn, bare soils and vineyards, and the acquired data are represented as an image with a size of 512 × 217 pixels. Each pixel corresponds to an area of 3.7 m × 3.7 m and, after the removal of 20 water absorption bands, comprises hyperspectral information on 204 spectral bands. In total, the Salinas Dataset consists of about 111k pixels of which 54,129 have been labeled with respect to 16 semantic classes as shown in Fig. 4, while no labels are provided for the remaining pixels.

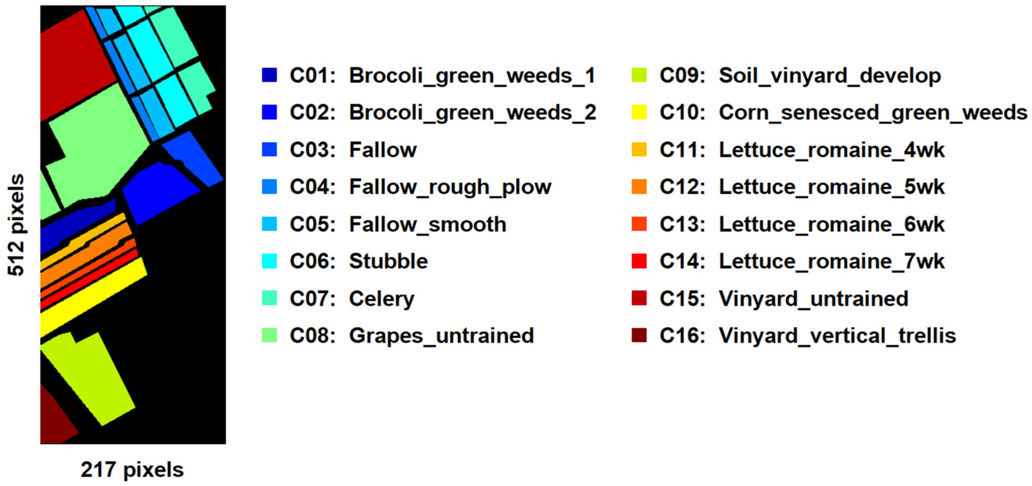

Fig. 4: Reference labels for the Salinas Dataset: each pixel is characterized by reflectance values on 204 spectral bands. Unlabeled pixels are indicated in black.

4.2 Experiments and Results

Our framework is tested on a standard laptop computer (Intel Core i7-6820HK, 2.7 GHz, 4 cores, 16 GB RAM, Matlab implementation). For the RF classifier, we use the implementation provided with (LIAW & WIENER 2002). This implementation also allows assessing the MDPA used for feature selection. For CFS and the score functions used for filter-based feature selection, we use implementations provided with (ZHAO et al. 2010), while we apply the GRM following the implementation used in (WEINMANN 2016).
First, we focus on the classification results derived with different configurations of our framework, i.e. with different feature (sub)sets provided as input to the RF classifier. In this context, we take into account that the wrapper-based feature selection method relying on a LDA classifier, the embedded feature selection method relying on the MDPA of a RF classifier and the GRM allow quantifying feature relevance and thus ranking the features according to their relevance with respect to the given classification task. However, unlike the CFS method, they do not retain a recommendation for the number of best-ranked features that should be used as the basis for classification. Hence, for the hyperspectral data, we follow (BRADLEY et al. 2018) and select the best-ranked features covering about 20% of all available features. For the Pavia Centre Dataset, we thus use subsets of the 20 best-ranked features, while we use subsets of the 40 best-ranked features for the Salinas Dataset. The quantitative classification results are provided in Tab. 1 in terms of overall accuracy (OA), κ-index (κ) and mean F_1-score across all classes (mF_1), and the qualitative classification results are provided in Fig. 5. With these results, we also provide results derived for the transformation of the hyperspectral data to multispectral Sentinel-2-like data.

Tab. 1: Classification results derived for different feature (sub)sets provided as input to a RF classifier.

Pavia Centre	OA (in %)	κ (in %)	mF_1 (in %)
All Features	**96.50**	**95.05**	90.50
LDA Wrapper	95.53	93.68	88.62
RF-MDPA	96.22	94.65	89.59
GRM	86.68	81.69	76.03
CFS	96.48	95.01	**90.60**
Sentinel-2	96.37	94.86	90.01

Salinas	OA (in %)	κ (in %)	mF_1 (in %)
All Features	87.13	85.69	92.19
LDA Wrapper	86.38	84.87	91.37
RF-MDPA	86.95	85.49	92.16
GRM	85.72	84.16	91.13
CFS	**87.14**	**85.71**	**92.32**
Sentinel-2	85.84	84.26	91.08

In the next step, we focus on feature relevance assessment using the described methods, i.e. 1) the wrapper-based feature selection method relying on a LDA classifier, 2) the embedded feature selection method relying on the MDPA of a RF classifier, 3) the GRM and 4) the CFS. For the latter, we take into account that it retains a subset of features without a ranking. Hence, we assign the selected features a relevance of 1 and all others a relevance of 0. CFS retains 21 features for the Pavia Centre Dataset and 36 features for the Salinas dataset. For all involved methods, the derived relevance of the considered spectral bands is visualized in Fig. 6 for the Pavia Centre Dataset and in Fig. 7 for the Salinas Dataset. Furthermore, the feature relevance assessment results for the transformation to multispectral Sentinel-2-like data are depicted in Fig. 8.

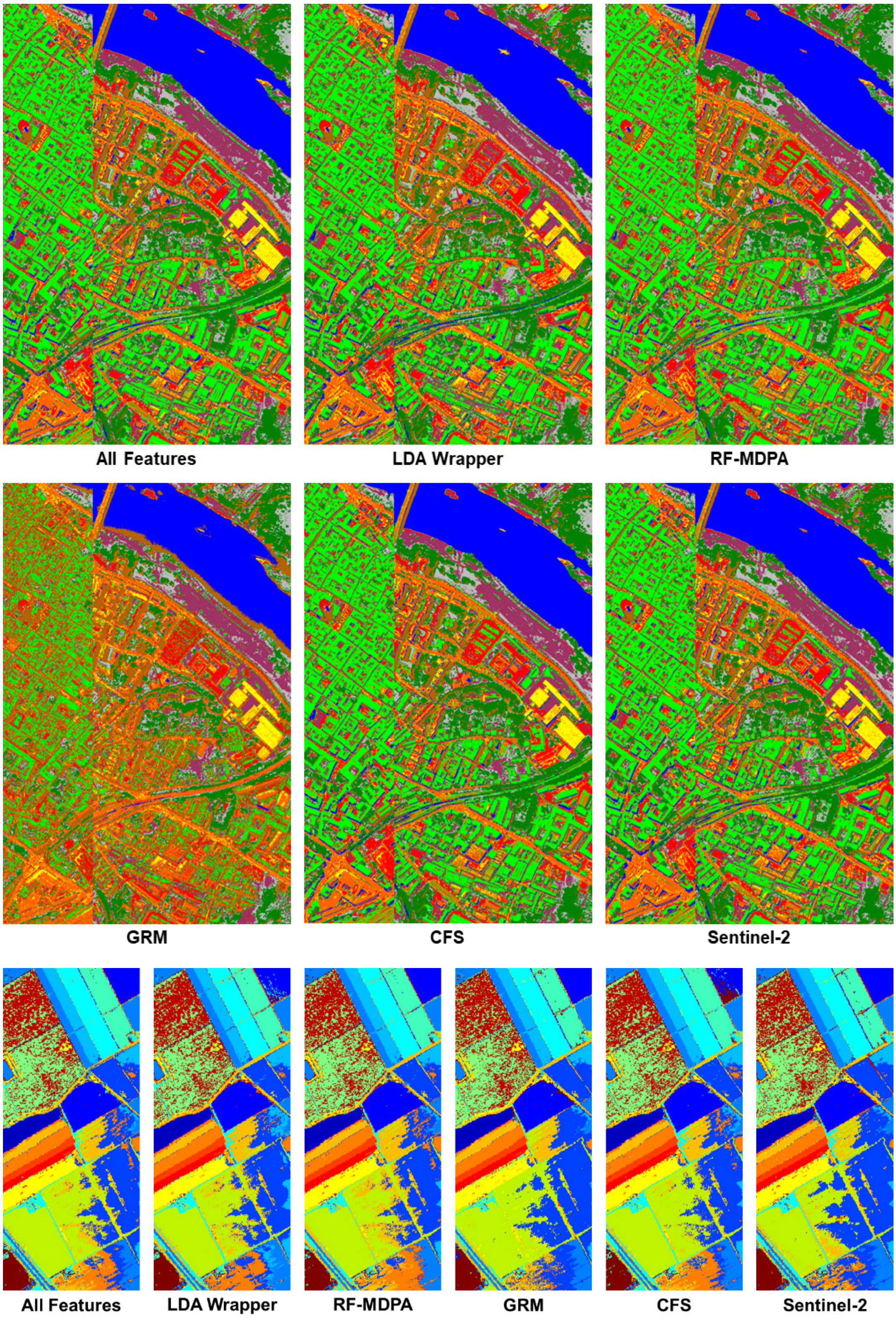

Fig. 5: Classification results achieved for the Pavia Centre Dataset (first row and second row) and for the Salinas Dataset (third row) when using different feature (sub)sets: the color encoding follows the definitions provided in Figs. 3 and 4.

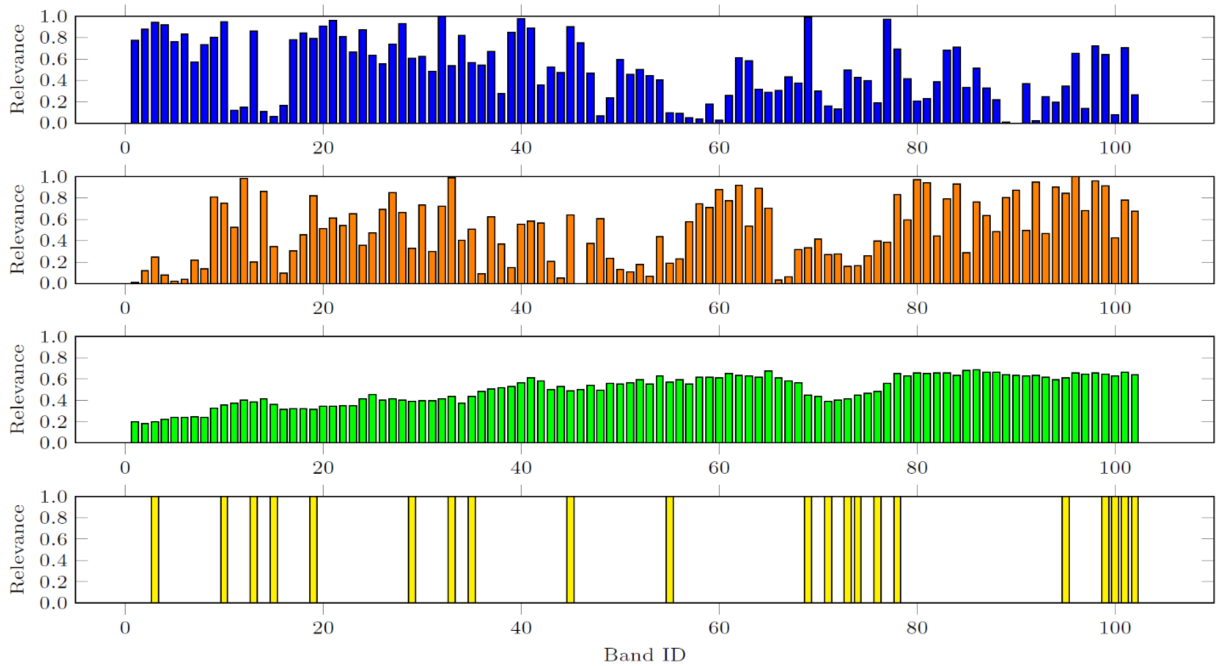

Fig. 6: Relevance of the spectral bands of the Pavia Centre Dataset when using the wrapper-based feature selection method relying on a LDA classifier (blue), the embedded feature selection method relying on the MDPA of a RF classifier (orange), the GRM (green) and the CFS (yellow). For the latter, selected features are assigned a relevance of 1 and all others a relevance of 0.

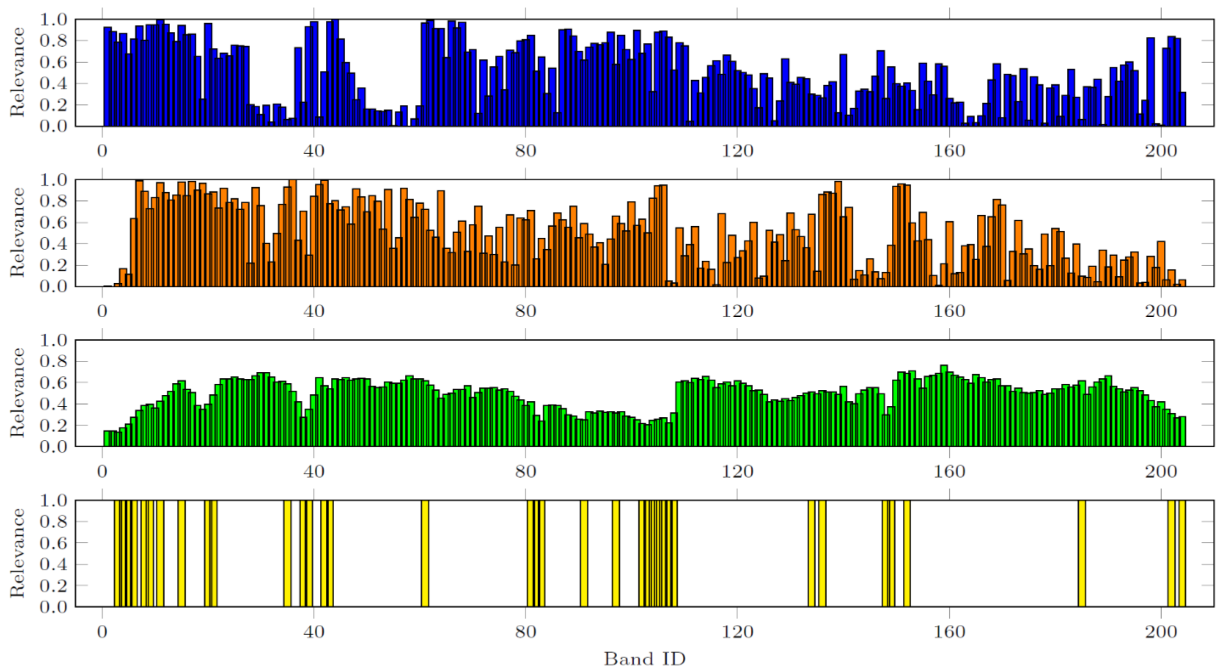

Fig. 7: Relevance of the spectral bands of the Salinas Dataset when using the wrapper-based feature selection method relying on a LDA classifier (blue), the embedded feature selection method relying on the MDPA of a RF classifier (orange), the GRM (green) and the CFS (yellow). For the latter, selected features are assigned a relevance of 1 and all others a relevance of 0.

Regarding feature relevance assessment, the involved wrapper-based method requires processing times of about 3 min 23 s and 41 min 24 s for the Pavia Centre Dataset and the Salinas Dataset, respectively. The MDPA is derived in 0.46 s and 2.20 s, the GRM in 4.95 s and 14.03 s, and the CFS in 0.69 s and 1.16 s for the Pavia Centre Dataset and the Salinas Dataset, respectively.

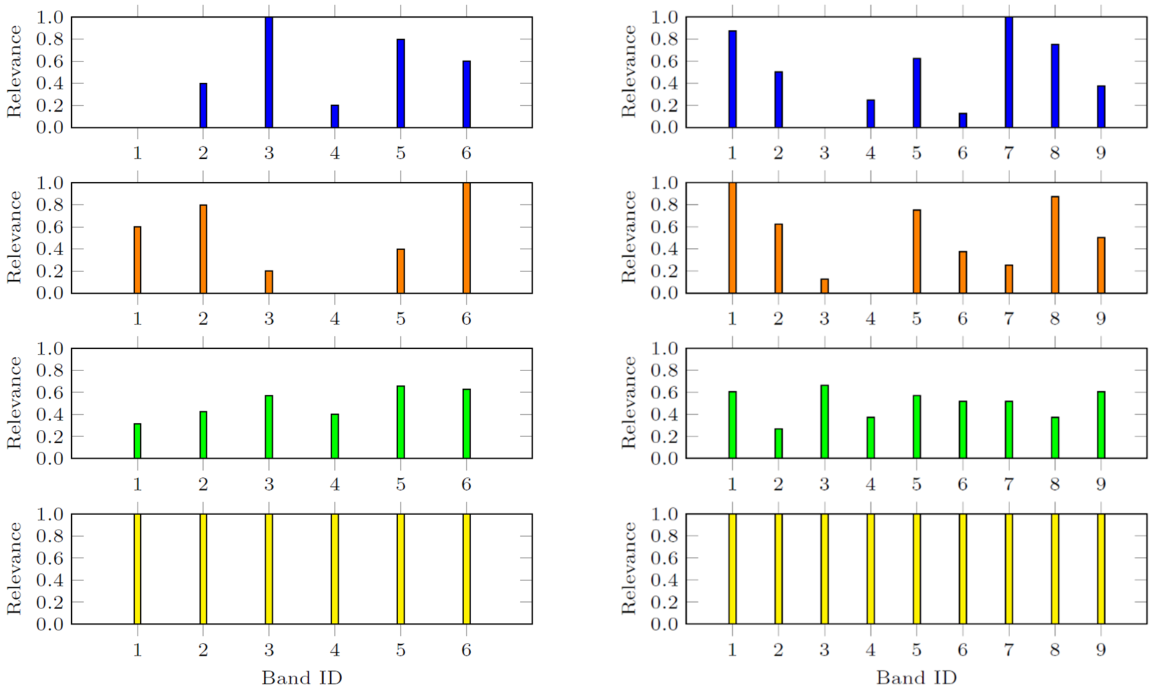

Fig. 8: Relevance of the Sentinel-2-like spectral bands derived from the Pavia Centre Dataset (left) and the Salinas Dataset (right) when using the wrapper-based feature selection method relying on a LDA classifier (blue), the embedded feature selection method relying on the MDPA of a RF classifier (orange), the GRM (green) and the CFS (yellow). For the latter, all features are selected and assigned a relevance of 1. The band IDs refer to the bands B_2, B_3, B_4, B_5, B_6 and B_7 for the Pavia Centre Dataset and to the bands B_2, B_3, B_4, B_5, B_6, B_7, B_{8a}, B_{11} and B_{12} for the Salinas Dataset.

5 Discussion

The derived classification results reveal that the classification of the Pavia Centre Dataset seems to be less challenging than the classification of the Salinas Dataset when considering the overall accuracy (Tab. 1). Intuitively, this might be due to the fact that there are only 9 classes of interest to be classified for the Pavia Centre Dataset and these classes are likely to be more distinctive, whereas the Salinas Dataset has 16 classes of interest which might have a higher similarity. However, the mean F_1-score across all classes reveals that instances of the different classes are on average better identified for the Salinas Dataset than for the Pavia Centre Dataset (Tab. 1). A look at the confusion matrices and class-wise evaluation metrics derived from these reveals that most of the classes of the Salinas Dataset are very well recognized and only the classes C08 ("Grapes_untrained") and C15 ("Vinyard_untrained") are not as well recognized. As instances of

these two classes occur frequently in the test data, the values for OA are worse. For the Pavia Centre Dataset, most of the classes can be well recognized and only the class C04 ("Self-Blocking Bricks") seems to be challenging due to the different materials which elements of this class might be composed of.
Furthermore, the derived classification results indicate that the different feature (sub)sets provided as input to the RF classifier tend to deliver results of similar quality (Tab. 1 and Fig. 5), even though very different strategies for feature selection are involved. In comparison to the use of all features, however, there is a significant reduction regarding the dimensionality of the considered feature vectors, as only about 20% of the given features are considered for all cases. A deeper analysis of the feature relevance assessment results (Figs. 6 and 7) reveals that the best-ranked and hence selected spectral bands are well-distributed across all available spectral bands. Interestingly, even the transformation of the given hyperspectral data to multispectral Sentinel-2-like data yields classification results of a similar, but slightly worse quality compared to the other approaches. This indicates that Sentinel-2-like data already provide a good source of information for land cover and land use classification, while the use of hyperspectral data seems to only slightly improve the results. Furthermore, the derived results indicate that multispectral Sentinel-2-like data seem to have a low degree of redundancy, so that the CFS method (which aims at reducing redundancy contained in the considered data) does not discard any of the given features (Fig. 8).

6 Conclusions

In this paper, we have comprehensively investigated the potential of both multispectral and hyperspectral data for land cover and land use classification and, in this context, we have also focused on comparatively assessing the relevance of involved spectral bands for the considered classification task. We have presented classification results achieved by using the original hyperspectral data as input to a Random Forest classifier, and we have demonstrated the potential of applying well-established approaches for feature selection: 1) a wrapper-based feature selection method relying on a Linear Discriminant Analysis classifier, 2) an embedded feature selection method relying on an internal metric of a Random Forest classifier, 3) a univariate filter-based feature selection method relying on a general relevance metric and 4) a multivariate filter-based feature selection method also allowing to reduce the degree of redundancy in the considered data. In addition, we have included a transformation of the given hyperspectral data to multispectral Sentinel-2-like data that are relevant for land cover and land use classification. The results derived for two benchmark datasets clearly reveal that Sentinel-2-like data already seem to provide a good source of information for land cover and land use classification, while the use of hyperspectral data (with or without well-established approaches for feature selection) seems to only provide a slight improvement for such a task. This is particularly interesting for large-scale land cover and land use mapping applications, where the use of Sentinel-2 data would correspond to a significantly less expensive data acquisition and a much faster repetition of data acquisition for the same area over time.

7 References

BRADLEY, P. E., KELLER, S. & WEINMANN, M., 2018: Unsupervised feature selection based on ultrametricity and sparse training data: a case study for the classification of high-dimensional hyperspectral data. Remote Sensing, **10**(10), Article 1564.

BREIMAN, L., 1996: Bagging predictors. Machine Learning, **24**(2), 123-140.

BREIMAN, L., 2001: Random forests. Machine Learning, **45**(1), 5-32.

CHEHATA, N., LE BRIS, A. & NAJJAR, S., 2014: Contribution of band selection and fusion for hyperspectral classification. Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 1-4.

CHI, M., FENG, R. & BRUZZONE, L., 2008: Classification of hyperspectral remote-sensing data with primal SVM for small-sized training dataset problem. Advances in Space Research, **41**(11), 1793-1799.

ELBERTZHAGEN, I., THONFELD, F. & MENZ, G., 2012: SVM-based agricultural land use assessment using Sentinel-2 simulation data. Proceedings of the Sentinel-2 Preparatory Symposium.

GUYON, I. & ELISSEEFF, A., 2003: An introduction to variable and feature selection. Journal of Machine Learning Research, **3**, 1157-1182.

HALL, M. A., 1999: Correlation-based feature subset selection for machine learning. PhD thesis, Department of Computer Science, University of Waikato, New Zealand.

HAM, J., CHENG, Y., CRAWFORD, M. M. & GHOSH, J., 2005: Investigation of the random forest framework for classification of hyperspectral data. IEEE Transactions on Geoscience and Remote Sensing, **43**(3), 492-501.

HUGHES, G. F., 1968: On the mean accuracy of statistical pattern recognizers. IEEE Transactions on Information Theory, **14**(1), 55-63.

JOELSSON, S. R., BENEDIKTSSON, J. A. & SVEINSSON, J. R., 2005: Random forest classifiers for hyperspectral data. IEEE Geoscience and Remote Sensing Symposium (IGARSS), 160-163.

KELLER, S., BRAUN, A. C., HINZ, S. & WEINMANN, M., 2016: Investigation of the impact of dimensionality reduction and feature selection on the classification of hyperspectral EnMAP data. 8th IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 1-5.

LE BRIS, A., CHEHATA, N., BRIOTTET, X. & PAPARODITIS, N., 2014: Use intermediate results of wrapper band selection methods: a first step toward the optimization of spectral configuration for land cover classifications. Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 1-4.

LIAW, A. & WIENER, M., 2002: Classification and regression by randomForest. R News, **2/3**, 18-22.

MELGANI, F. & BRUZZONE, L., 2004: Classification of hyperspectral remote sensing images with support vector machines. IEEE Transactions on Geoscience and Remote Sensing, **42**(8), 1778-1790.

SAEYS, Y., INZA, I. & LARRANAGA, P., 2007: A review of feature selection techniques in bioinformatics. Bioinformatics, **23**(19), 2507-2517.

SPOTO, F., SY, O., LABERINTI, P., MARTIMORT, P., FERNANDEZ, V., COLIN, O., HOERSCH, B. & MEYGRET, A., 2012: Overview of Sentinel-2. IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 1707-1710.

THONFELD, F., FEILHAUER, H. & MENZ, G., 2012: Simulation of Sentinel-2 images from hyperspectral data. Proceedings of the Sentinel-2 Preparatory Symposium.

VAN DER MEER, F. D., VAN DER WERFF, H. M. A. & VAN RUITENBEEK, F. J. A., 2014: Potential of ESA's Sentinel-2 for geological applications. Remote Sensing of Environment, **148**, 124-133.

WEINMANN, M., 2016: Reconstruction and analysis of 3D scenes – From irregularly distributed 3D points to object classes. Springer, Cham, Switzerland.

WEINMANN, M. & WEIDNER, U., 2018: Land-cover and land-use classification based on multitemporal Sentinel-2 data. IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 4946-4949.

WEINMANN, M., MAIER, P. M., FLORATH, J. & WEIDNER, U., 2018: Investigations on the potential of hyperspectral and Sentinel-2 data for land-cover / land-use classification. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, **IV-1**, 155-162.

ZHAO, Z., MORSTATTER, F., SHARMA, S., ALELYANI, S., ANAND, A. & LIU, H., 2010: Advancing Feature Selection Research – ASU Feature Selection Repository. Technical Report, School of Computing, Informatics, and Decision Systems Engineering, Arizona State University, Tempe, AZ, USA, 2010.