INVESTIGATIONS ON THE POTENTIAL OF HYPERSPECTRAL AND SENTINEL-2 DATA FOR LAND-COVER / LAND-USE CLASSIFICATION

M. Weinmann, P. M. Maier, J. Florath, U. Weidner

Institute of Photogrammetry and Remote Sensing, Karlsruhe Institute of Technology, Karlsruhe, Germany -
(martin.weinmann, philipp.maier, uwe.weidner)@kit.edu, janine.florath@student.kit.edu

Commission I, WG I/1

KEY WORDS: Classification, Land-cover/Land-use, Hyperspectral Data, Dimensionality Reduction, Feature Selection, Multispectral Data, Sentinel-2 Data

ABSTRACT:

The automated analysis of large areas with respect to land-cover and land-use is nowadays typically performed based on the use of hyperspectral or multispectral data acquired from airborne or spaceborne platforms. While hyperspectral data offer a more detailed description of the spectral properties of the Earth's surface and thus a great potential for a variety of applications, multispectral data are less expensive and available in shorter time intervals which allows for time series analyses. Particularly with the recent availability of multispectral Sentinel-2 data, it seems desirable to have a comparative assessment of the potential of both types of data for land-cover and land-use classification. In this paper, we focus on such a comparison and therefore involve both types of data. On the one hand, we focus on the potential of hyperspectral data and the commonly applied techniques for data-driven dimensionality reduction or feature selection based on these hyperspectral data. On the other hand, we aim to reason about the potential of Sentinel-2 data and therefore transform the acquired hyperspectral data to Sentinel-2-like data. For performance evaluation, we provide classification results achieved with the different types of data for two standard benchmark datasets representing an urban area and an agricultural area, respectively.

1. INTRODUCTION

Hyperspectral imagery acquired from airborne or spaceborne sensor platforms is meanwhile commonly used for scene interpretation in terms of land-cover and land-use classification (Plaza et al., 2009; Camps-Valls et al., 2014). In this context, the main objective is a classification on a per-pixel basis, whereby the classes of interest are typically defined with respect to a specific application or usecase.

To achieve a pixel-wise classification of hyperspectral imagery, different strategies can be followed. A rather intuitive strategy is to consider the reflectance values of each pixel across all spectral bands. These reflectance values are concatenated to a high-dimensional feature vector which, in turn, is provided as input to a classifier delivering a hypothesis about the respective class label. Thereby, a supervised classification based on a standard classifier such as a Support Vector Machine (Melgani and Bruzzone, 2004; Chi et al., 2008) or a Random Forest (Ham et al., 2005; Joelsson et al., 2005) is often applied, where the involved classifier needs to be trained before on representative training data.

Regarding the classification of high-dimensional data, however, the *Hughes phenomenon* (Hughes, 1968) has to be taken into account, according to which an increase of the number of considered features over a certain threshold typically results in a decrease in classification accuracy, given a constant number of training examples (Melgani and Bruzzone, 2004; Keller et al., 2016). For hyperspectral data, this curse of dimensionality mainly arises from two sources. On the one hand, there is the high degree of redundancy which can be expected in hyperspectral data due to the fact that reflectance values of adjacent spectral bands tend to be strongly correlated. On the other hand, the involved classifier has to deal with more or less relevant features to

solve the given classification task and possibly even with irrelevant features. To address the Hughes phenomenon, a variety of dimensionality reduction or feature selection techniques are commonly used (van der Maaten et al., 2009; Saeys et al., 2007).

In this paper, we focus on the potential of both multispectral data and hyperspectral data (and the commonly applied techniques for data-driven dimensionality reduction or feature selection based on these hyperspectral data) for land-cover and land-use classification. This is motivated by the fact that, with the launch of the Sentinel-2 satellites, a systematic acquisition of observations over land and coastal areas is pursued (Spoto et al., 2012), where the observations consist of multispectral imagery that is well-suited to derive thematic maps indicating land-cover and land-use (Weinmann and Weidner, 2018). Introducing a transformation of acquired hyperspectral data to Sentinel-2-like data thus allows directly analyzing the potential of Sentinel-2 data for land-cover and land-use classification in comparison to hyperspectral data as e.g. expected to be acquired with the Environmental Mapping and Analysis Program (EnMAP) mission (Kaufmann et al., 2006, 2008; Segl et al., 2010), a hyperspectral satellite mission which is supposed to be launched into space in the near future. In summary, this paper has several contributions:

- We focus on the use of multispectral and hyperspectral data for land-cover and land-use classification, and we thereby also consider the effectiveness of standard dimensionality reduction and feature selection techniques for comparison.

- We present a method to transform acquired hyperspectral data to Sentinel-2-like data.

- We demonstrate the potential of Sentinel-2-like data in comparison to hyperspectral data for land-cover and land-use classification on two standard benchmark datasets representing an urban area and an agricultural area, respectively.

After briefly summarizing related work in Section 2, we explain the applied methodology in Section 3. To demonstrate the performance of our methodology, we present results achieved on two commonly used benchmark datasets in Section 4. These results are discussed in detail in Section 5. Finally, in Section 6, we provide concluding remarks as well as suggestions for future work.

2. RELATED WORK

The classification of hyperspectral imagery can easily be achieved in the form of a pixel-wise classification based on the reflectance values across all spectral bands. Thereby, the reflectance values are used to define the entries of a respective feature vector, while the classification itself may be based on well-known standard classifiers such as kernel-based methods (Camps-Valls and Bruzzone, 2005), Support Vector Machines (Melgani and Bruzzone, 2004; Chi et al., 2008), or Random Forests (Ham et al., 2005; Joelsson et al., 2005).

Despite the classifier, the data representation plays an important role when classifying hyperspectral data. More specifically, the use of reflectance values across all available spectral bands to define the entries of a feature vector leads to more or less relevant features and possibly even irrelevant or redundant features with respect to the considered classification task. Depending on the defined classes, only considering the spectral bands within certain wavelength intervals might be sufficient to achieve appropriate classification results, i.e. not all spectral bands are relevant with respect to the classification task. Furthermore, the reflectance values of adjacent spectral bands tend to be strongly correlated which results in a high degree of redundancy. While the use of all available data as input to a classifier seems appropriate to compensate a lack of knowledge about the scene and/or the data, it has been proven in practice that the predictive accuracy of classifiers is negatively influenced for such high-dimensional classification tasks. Typically, the Hughes phenomenon (Hughes, 1968) can be observed, i.e. above a certain value a further increase of the number of involved features results in a (typically significant) decrease in predictive accuracy.

The Hughes phenomenon is typically addressed by transferring the given high-dimensional data into a new and more compact data representation encoding almost the same information with respect to the classification task. On the one hand, this can be achieved via dimensionality reduction (van der Maaten et al., 2009) focusing on the transformation of feature vectors from the original feature space onto a specific subspace spanned by meta-features. In the context of classifying hyperspectral data, dimensionality reduction is typically addressed by applying variants of the Principal Component Analysis (PCA) (Licciardi et al., 2012; Keller et al., 2016) or the Independent Component Analysis (ICA) (Wang and Chang, 2006; Villa et al., 2011), or by taking into account the Maximum Noise Fraction (MNF) transformation (Green et al., 1988). In contrast to the standard PCA which focuses on the variance of linearly uncorrelated components derived from the given data, the MNF transformation relies on the signal-to-noise ratio (SNR) or the noise fraction of components derived via a linear transformation. On the other hand, feature selection techniques can be applied which allow gaining predictive accuracy and improving computational efficiency with respect to both time and memory consumption, while retaining meaningful features (Guyon and Elisseeff, 2003; Saeys et al., 2007; Zhao et al., 2010). In the context of classifying hyperspectral data, feature selection techniques resulting in only the consideration of reflectance values across specific spectral bands have been used in several investigations (Melgani and Bruzzone, 2004; Le Bris et al., 2014; Chehata et al., 2014; Keller et al., 2016, 2017). Such techniques may also allow assessing the importance of single spectral bands for land-cover and land-use classification (Le Bris et al., 2014; Keller et al., 2016).

Despite a pixel-wise consideration of hyperspectral data, there are also approaches exploiting spatial-spectral features and thus accounting for relations within a local image neighborhood in addition to the spectral information per pixel. In this regard, spectral information can for instance be sampled within the local surrounding of a pixel, e.g. within adaptive pixel neighborhoods (Fauvel et al., 2008) or within superpixels (Fang et al., 2015). Furthermore, it has been proposed to use the results of a pixel-wise classification in combination with watershed segments (Tarabalka et al., 2008) which, in turn, allows a majority voting within watershed segments to account for spatial information. A probabilistic pixel-wise classification may also provide the basis for a subsequent hierarchical optimization (Tarabalka and Tilton, 2011), a Markov Random Field (MRF) regularization (Tarabalka et al., 2010) or a Conditional Random Field (CRF) regularization (Roscher and Waske, 2014). In recent years, more and more attention has been paid to the use of modern deep learning techniques for spatial-spectral classification, e.g. in the form of a Convolutional Neural Network (CNN) consisting of several convolutional layers and pooling layers which allow the extraction of deep spatial-spectral features that are highly discriminant (Chen et al., 2016).

In our work, we take into account that the more sophisticated classifiers improve the classification results, but involve a huge number of parameters that have to be optimized during the training process. Accordingly, a large amount of representative training data would be required including a dense labeling to learn local context. In case of the standard labeled hyperspectral imagery, however, only a small amount of training data is typically available and these training data often correspond to isolated segments within an image. Hence, we focus on standard supervised classification techniques in combination with either dimensionality reduction or feature selection techniques. Thereby, particular focus is put on a transformation of hyperspectral data to Sentinel-2-like data. Such a transformation has already been proposed 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). In contrast, we aim at reasoning about the potential of Sentinel-2 data for land-cover and land-use classification in comparison to hyperspectral data as expected to be acquired with the future Environmental Mapping and Analysis Program (EnMAP) mission (Kaufmann et al., 2006, 2008; Segl et al., 2010). Among a diversity of systems that can be used for the acquisition of multispectral data, the Sentinel-2 satellites are nowadays used to systematically acquire observations over land and coastal areas (Spoto et al., 2012). These satellites provide acquisitions of multispectral imagery with high update rates, resulting in data which is well-suited to derive accurate maps indicating land-cover and land-use and to describe dynamic processes. The 13 spectral bands of the Sentinel-2 MultiSpectral Instrument[1] (MSI) are described in Table 1 and the spectral response functions corresponding to these spectral bands are visualized in Figure 1.

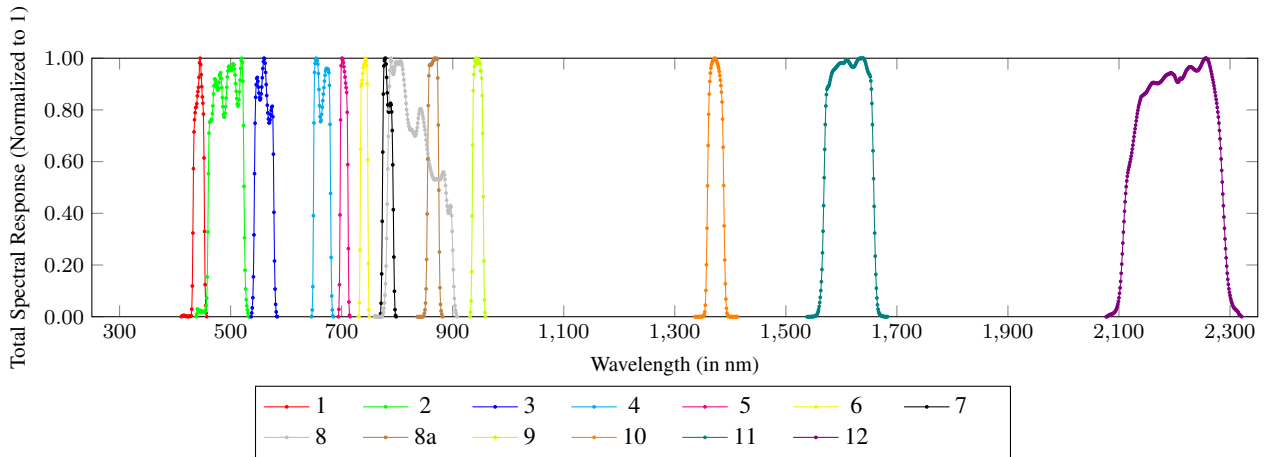[1] https://sentinel.esa.int/documents/ (last access: 22 February 2018)

Figure 1. 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).

Spectral band	Central wavelength [nm]	Bandwidth [nm]
B1	443	20
B2	490	65
B3	560	35
B4	665	30
B5	705	15
B6	740	15
B7	783	20
B8	842	115
B8a	865	20
B9	945	20
B10	1375	20
B11	1610	90
B12	2190	180

Table 1. Definition of the spectral bands given for Sentinel-2 data.

3. METHODOLOGY

The proposed methodology involves four different approaches to define feature vectors (Section 3.1) which serve as input to a standard supervised classification (Section 3.2).

3.1 Feature Extraction

The input to our framework is represented by hyperspectral imagery in the form of an image stack with a high number of layers $\mathcal{L}_1, \ldots, \mathcal{L}_D$, whereby each layer corresponds to a specific spectral band and the spectral information of each spectral band is represented by a reflectance value. Accordingly, each pixel is represented by D reflectance values. On the basis of this data representation, we focus on four options to derive feature vectors that serve as input to the subsequent classification procedure.

3.1.1 Original Data Representation The straightforward approach to define feature vectors consists in using all given information by simply concatenating the reflectance values for all spectral bands. This yields a D-dimensional feature vector.

3.1.2 Dimensionality Reduction To transform the data representation given by the concatenation of all defined features to a new data representation of lower dimensionality, we use a standard *Principal Component Analysis* (PCA) which has for instance been involved for the classification of simulated hyperspectral EnMAP data (Keller et al., 2016, 2017). The PCA uses an orthogonal transformation to transform a set of feature vectors from a high-dimensional feature space (which is spanned by possibly correlated features) to a new feature space which is spanned by linearly uncorrelated meta-features, the *principal components*, which account for the data variability along the respective dimension. The principal components are defined in a way that the first principal component covers the highest variability of the data, and that each subsequent principal component covers the highest possible variability under the constraint that it is orthogonal to all previous principal components. Thus, the most relevant information of the considered data is covered by the first few principal components, while the last few principal components do not significantly contribute to the variability of the given data and can hence be considered as almost irrelevant. Consequently, the PCA-based dimensionality reduction focuses on only considering the first few principal components, which cover a certain percentage of the variability of the given data. For our experiments, we use the first few principal components which cover 99.9 % of the variability of the given training data to define the new feature vectors of reduced dimensionality, and we expect no significant loss of relevant information with respect to the classification task when discarding all other principal components.

3.1.3 Feature Selection To retain a suitable subset of the original features, we conduct feature selection by applying *Correlation-based Feature Selection* (CFS) (Hall, 1999) which has for instance been used for the analysis of relevant spectral bands in simulated hyperspectral EnMAP data (Keller et al., 2017) and for the analysis of relevant spectral bands in spectral libraries with differently labeled materials and land-cover types (Ilehag et al., 2017). In general, CFS takes into account (i) the correlation between features and classes to identify relevant features and (ii) the correlation among features to identify and discard redundant features. Denoting the average correlation between features and classes by $\bar{\rho}_{fc}$ and the average correlation between different features by $\bar{\rho}_{ff}$, the relevance R of a feature subset comprising n features is defined according to

$$R = \frac{n\bar{\rho}_{fc}}{\sqrt{n + n(n-1)\bar{\rho}_{ff}}} \qquad (1)$$

whereby the involved correlation metric ρ is given by the symmetrical uncertainty (Hall, 1999). Maximizing the relevance R over the set of all possible feature subsets finally yields a feature subset of relevant features with low redundancy. This maximization can be achieved by applying an iterative scheme where, for

each iteration, 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. The new feature vectors defined on the basis of the selected feature subset thus correspond to a selection of a few dimensions of the original data representation.

3.1.4 Transformation to Sentinel-2-like Data To analyze the potential of Sentinel-2 data for the classification task, we transform the data representation given by the concatenation of all defined features to multispectral data corresponding to Sentinel-2-like data. For each of the 13 spectral bands corresponding to Sentinel-2 data, we check which of the spectral bands of the hyperspectral data are within. The reflectance values corresponding to these spectral bands, in turn, are used to derive a reflectance value for the considered spectral band of the Sentinel-2 data. Thereby, we consider the weighted mean of the reflectance values and the weights are determined via a linear interpolation based on the total spectral response functions defined for the Sentinel-2 data, whereby these functions are normalized to 1 (cf. Figure 1).

In the scope of our work, we particularly focus on systems that have been proposed for the airborne acquisition of hyperspectral data, because several labeled benchmark datasets including the specification of the used spectral bands are available for this case. On the one hand, we focus on data acquired with the Reflective Optics System Imaging Spectrometer (ROSIS), where the central wavelength λ_k of the k-th spectral band is given according to

$$\lambda_k = \lambda_0 + (k-1) \cdot \Delta\lambda \qquad k = 1, \ldots, 115 \qquad (2)$$

and the spectral sampling is specified with $\Delta\lambda = 4.0$ nm (Gege et al., 2008). The value of the constant λ_0 is approximated with $\lambda_0 = 379.53$ nm following the aforementioned reference, so that the spectral bands reach from $\lambda_1 = 379.53$ nm to $\lambda_{115} = 835.53$ nm in our experiments. On the other hand, we focus on data acquired with the Airborne Visible / Infrared Imaging Spectrometer[2] (AVIRIS) for which calibration files are provided which include the central wavelength and bandwidth of each of the 224 spectral bands.

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, it should however be taken into account that not all spectral bands contain valuable information (Weinmann and Weidner, 2018). Accordingly, for our classification based on the Sentinel-2-like data, we only make use of those reflectance values corresponding to the spectral bands B2 (central wavelength of 490 nm), B3 (560 nm), B4 (665 nm), B5 (705 nm), B6 (740 nm), B7 (783 nm), B8a (865 nm), B11 (1610 nm) and B12 (2190 nm). The spectral bands B1 (443 nm), B9 (945 nm) and B10 (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 certain wavelengths. Furthermore, we discard the reflectance value corresponding to the spectral band B8 (842 nm), because it is overlapping with B8a but much wider and less characteristic.

3.2 Supervised Classification

For classification, we focus on a standard supervised classification based on a *Random Forest* classifier (Breiman, 2001) as representative of modern discriminative methods. Such a classifier is

---

[2]http://aviris.jpl.nasa.gov (last access: 22 November 2017)
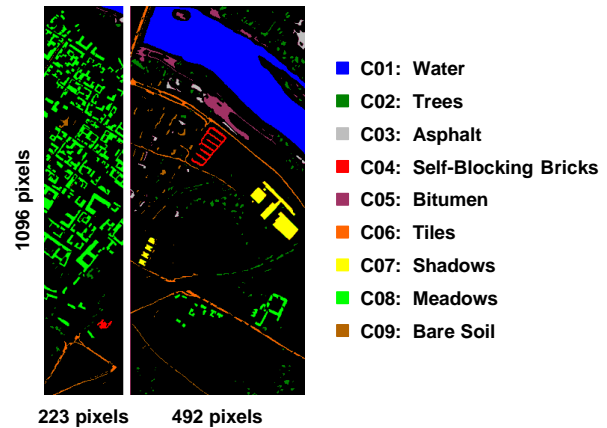

Figure 2. Reference labels for the *Pavia Centre* dataset: each pixel is characterized by reflectance values on 102 spectral bands. Unlabeled pixels are indicated in black.

composed of a number of decision trees, which are trained independently on randomly selected subsets of the training data via bootstrap aggregating ("bagging") (Breiman, 1996). Once the training is finished, a new feature vector can be classified by considering the majority vote across the individual votes of the involved decision trees. Suitable values for the number of involved decision trees and the maximum tree depth are selected by conducting a grid search on a suitable subspace.

4. EXPERIMENTAL RESULTS

In the following, we first describe the used benchmark datasets in Section 4.1 and implementation details in Section 4.2. Subsequently, we summarize the conducted experiments in Section 4.3. Finally, we present the derived results in Section 4.4.

4.1 Datasets

For our experiments, we use the *Pavia Centre* dataset representing an urban area and the *Salinas* dataset representing an agricultural area. Both datasets are available in a repository of hyperspectral remote sensing scenes.[3]

The *Pavia Centre* dataset was acquired with the Reflective Optics System Imaging Spectrometer (ROSIS) in a low-altitude flight campaign over the city of Pavia, Italy. The acquired data are represented in the form of two images of 1096×223 pixels and 1096×492 pixels, respectively, where each pixel corresponds to an area of 1.3 m $\times$ 1.3 m and comprises hyperspectral information on 102 spectral bands. The complete *Pavia Centre* dataset consists of about 784k pixels of which 7,456 have been labeled with respect to 9 classes as shown in Figure 2, while no reference labels are provided for the remaining pixels.

The *Salinas* dataset was acquired with the Airborne Visible / Infrared Imaging Spectrometer (AVIRIS) in a low-altitude flight campaign over Salinas Valley in California, USA. The acquired data are represented in the form of an image of 512×217 pixels, where each pixel corresponds to an area of 3.7 m $\times$ 3.7 m and, after the removal of 20 water absorption bands, comprises hyperspectral information on 204 spectral bands. The complete *Salinas* dataset consists of about 111k pixels of which 54,129 have been labeled with respect to 16 classes as shown in Figure 3, while no reference labels are provided for the remaining pixels.

---

[3]http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes (last access: 22 November 2017)

- ■ **C01: Brocoli_green_weeds_1**
- ■ **C02: Brocoli_green_weeds_2**
- ■ **C03: Fallow**
- ■ **C04: Fallow_rough_plow**
- ■ **C05: Fallow_smooth**
- ■ **C06: Stubble**
- ■ **C07: Celery**
- ■ **C08: Grapes_untrained**
- ■ **C09: Soil_vinyard_develop**
- ■ **C10: Corn_senesced_green_weeds**
- ■ **C11: Lettuce_romaine_4wk**
- ■ **C12: Lettuce_romaine_5wk**
- ■ **C13: Lettuce_romaine_6wk**
- ■ **C14: Lettuce_romaine_7wk**
- ■ **C15: Vinyard_untrained**
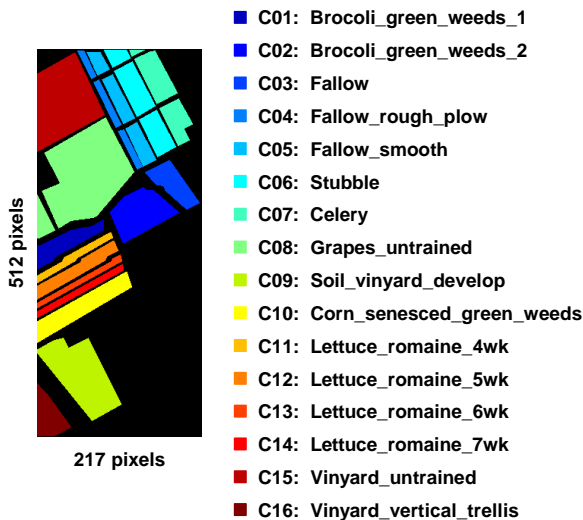- ■ **C16: Vinyard_vertical_trellis**

Figure 3. 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 Implementation

We implemented our framework in Matlab and used the CFS implementation provided with (Zhao et al., 2010) as well as the Random Forest implementation provided with (Liaw and Wiener, 2002).

4.3 Experiments

In our work, we consider the feature vectors $\mathbf{x}_i$ for the original hyperspectral data, $\text{PCA}(\mathbf{x}_i)$ for the PCA-based encoding of hyperspectral data, $\text{CFS}(\mathbf{x}_i)$ for the CFS-based selection of relevant spectral bands, and $\text{S2}(\mathbf{x}_i)$ for the Sentinel-2-like data. For each type of feature vector, classification is performed by using a Random Forest classifier. For performance evaluation, we use two benchmark datasets representing an urban area and an agricultural area, respectively. In total, this results in eight experiments.

For each experiment, the appropriate number of involved decision trees used to define the Random Forest classifier is determined individually during the training process by performing a grid search on a suitable subspace. Thereby, we also take into account that an imbalanced distribution of training examples across all classes might have a detrimental effect on the training process in terms of both feature selection and classification (Chen et al., 2004; Criminisi and Shotton, 2013; Weinmann, 2016), and we consequently use an identical number of training examples for all classes. In this regard, we randomly select 100 training examples per class and all remaining data is used for the subsequent prediction based on a trained Random Forest classifier.

For performance evaluation, we compare each derived labeling with the reference labeling on a per-point basis. Thereby, we consider the global evaluation metrics represented by the overall accuracy $\text{OA} \in [0, 1]$, the kappa value $\kappa \in [-1, 1]$ and the unweighted average $\text{mF}_1 \in [0, 1]$ of the F_1-scores across all classes. To also allow reasoning about the performance for single classes, we additionally consider the classwise evaluation metrics represented by recall, precision and F_1-score, each characterized by a value in $[0, 1]$.

4.4 Results

As mentioned before, we consider four different approaches to define feature vectors which serve as input to a Random Forest classifier. For each approach, the classification results derived for the two benchmark datasets are summarized in Table 2, while the corresponding visualizations of the labels predicted for the complete scene are given in Figures 4 and 5, respectively. The classification results achieved for the *Pavia Centre* dataset are generally of rather high quality ($\text{OA} = 95.46 \ldots 96.72\,\%$), while the classification results achieved for the *Salinas* dataset are worse ($\text{OA} = 84.77 \ldots 86.69\,\%$). The corresponding values for the classwise evaluation metrics of recall, precision and F_1-score are shown in Figures 6 and 7 for the *Pavia Centre* dataset and the *Salinas* dataset, respectively. These figures reveal that particularly class C04 ("Self-Blocking Bricks") of the *Pavia Centre* dataset and classes C08 ("Grapes_untrained") and C15 ("Vinyard_untrained") of the *Salinas* dataset seem to be problematic for the given classification task.

Features	Pavia Centre			Salinas		
	OA	κ	mF_1	OA	κ	mF_1
$\mathbf{x}_i$	95.61	93.81	89.40	86.55	85.05	91.98
$\text{PCA}(\mathbf{x}_i)$	96.72	95.36	90.48	86.69	85.18	92.43
$\text{CFS}(\mathbf{x}_i)$	95.46	93.60	89.12	86.38	84.86	91.90
$\text{S2}(\mathbf{x}_i)$	95.52	93.68	89.14	84.77	83.06	90.73

Table 2. Classification results (in %) achieved for the *Pavia Centre* dataset and the *Salinas* dataset when using different data representations as input to a Random Forest classifier.

5. DISCUSSION

At first sight, the derived classification results (cf. Table 2) indicate that the classification of the *Pavia Centre* dataset ($\text{OA} = 95.46 \ldots 96.72\,\%$) seems to be less challenging than the classification of the *Salinas* dataset ($\text{OA} = 84.77 \ldots 86.69\,\%$). This is rather intuitive, because the *Pavia Centre* dataset has only 9 classes of interest which are likely to be more distinctive, whereas the *Salinas* dataset has 16 classes of interest with all of them representing agricultural classes with a higher similarity. A closer look at the classification results however reveals a higher mF_1 which, in turn, means that instances of the different classes are on average better identified for the *Salinas* dataset. Indeed, most of the classes of the *Salinas* dataset are very well recognized (cf. Figure 7) and only the classes C08 ("Grapes_untrained") and C15 ("Vinyard_untrained") are not as well recognized (with $F_1 \approx 70\,\%$ and $F_1 \approx 60\,\%$ respectively). A look at the respective confusion matrices reveals that many misclassifications occur between these two classes. As many pixels corresponding to these classes are given in the prediction stage, the values for OA are worse. For the *Pavia Centre* dataset, most of the classes can be well recognized (cf. Figure 6). The classes C01 ("Water") and C09 ("Bare Soil") are extremely well recognized with almost no misclassifications ($F_1 \approx 100\,\%$), while the class C04 ("Self-Blocking Bricks") poses problems ($F_1 \approx 70\,\%$). This might be due to the different materials which elements of this class are composed of.

Among the different approaches to define the feature vectors serving as input to a Random Forest classifier, the PCA-based encoding of hyperspectral data $\text{PCA}(\mathbf{x}_i)$ yields the best results for both datasets. In case of the *Pavia Centre* dataset, the first 9 (of 102) principal components are selected to cover 99.9 % of
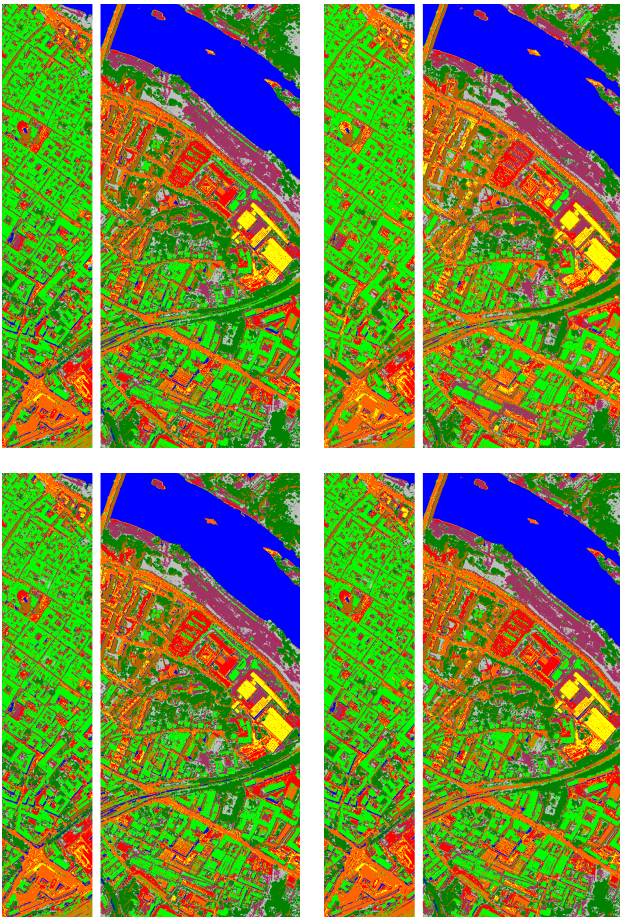
Figure 4. Classified scene for the *Pavia Centre* dataset when using the original hyperspectral data (top left), the PCA-based encoding of hyperspectral data (top right), the CFS-based selection of relevant spectral bands (bottom left) and the Sentinel-2-like data (bottom right) as input to a Random Forest classifier: the color encoding follows the definition in Figure 2.
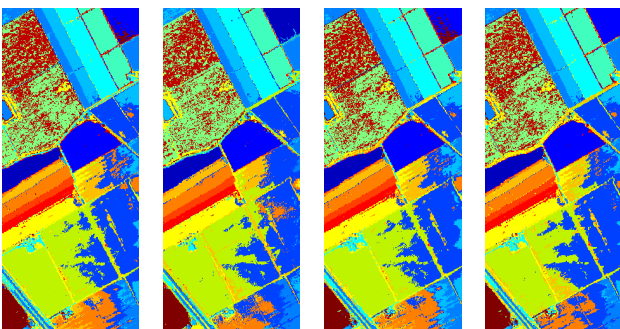

Figure 5. Classified scene for the *Salinas* dataset when using the original hyperspectral data (first column), the PCA-based encoding of hyperspectral data (second column), the CFS-based selection of relevant spectral bands (third column) and the Sentinel-2-like data (fourth column) as input to a Random Forest classifier: the color encoding follows the definition in Figure 3.

the variability of the given training data, while the first 6 (of 204) principal components are selected for the *Salinas* dataset. Thus, a significant reduction of dimensionality is achieved for both datasets when using the PCA. The CFS-based selection of relevant spectral bands $CFS(\mathbf{x}_i)$ leads to classification results of
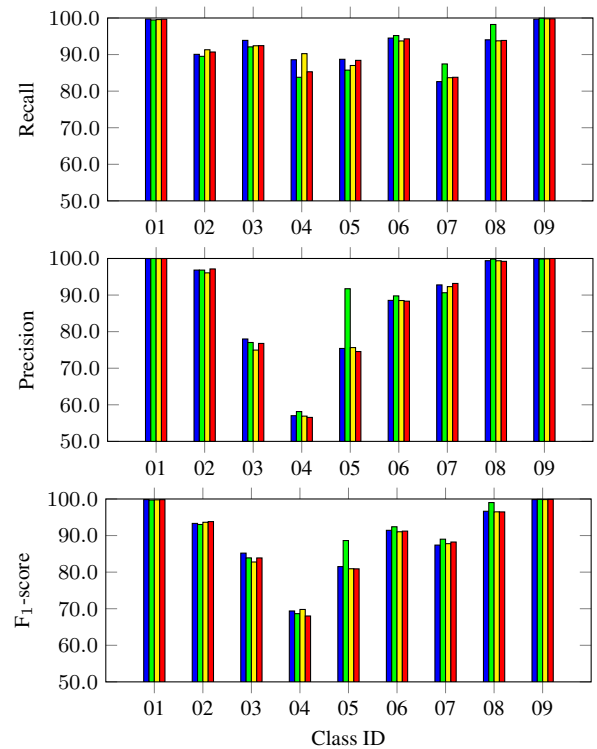

Figure 6. Recall (top), precision (center) and F_1-score (bottom) (in %) achieved for the *Pavia Centre* dataset when using different data representations as input to a Random Forest classifier.

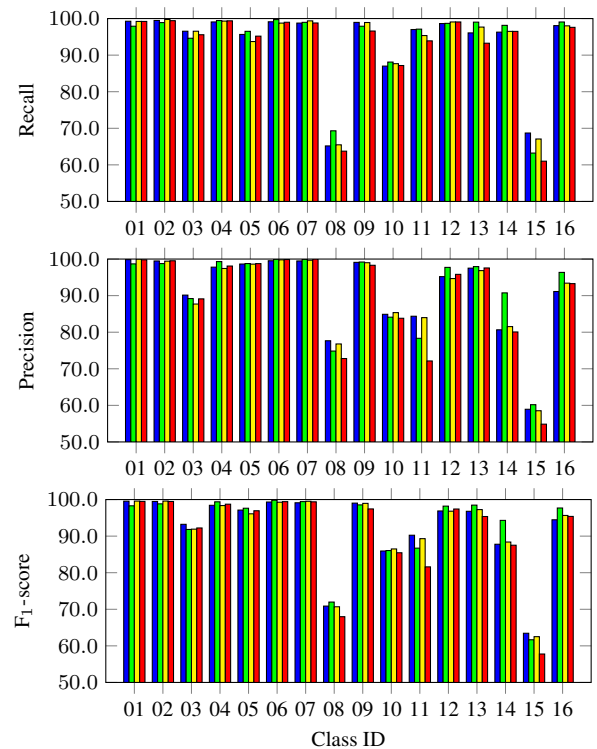Figure 7. Recall (top), precision (center) and F_1-score (bottom) (in %) achieved for the *Salinas* dataset when using different data representations as input to a Random Forest classifier.

almost the same quality as obtained for the original hyperspectral data $\mathbf{x}_i$. Thereby, the reflectance values corresponding to 25 (of 102) spectral bands are selected for the *Pavia Centre* dataset,

while the reflectance values corresponding to 43 (of 204) spectral bands are selected for the *Salinas* dataset. In both cases, the selected spectral bands are well-distributed across all available spectral bands. Interestingly, the Sentinel-2-like data $S2(\mathbf{x}_i)$ lead to classification results of a similar quality compared to $\mathbf{x}_i$ and $CFS(\mathbf{x}_i)$ for the *Pavia Centre* dataset, and to slightly worse classification results for the *Salinas* dataset. Yet, the difference in OA is so small that the visualizations of the corresponding classification results look rather similar (cf. Figures 4 and 5). This indicates that Sentinel-2 data already provide a good source of information for land-cover and land-use classification, while the use of hyperspectral data as can e.g. be expected to be provided with the EnMAP mission seems to only improve the results by a small margin. Such a conclusion is in accordance with other investigations where, based on the use of Landsat TM and AVIRIS data, it was found that the effect of seasonality on the results is more significant than the number of bands considered (Mannel and Price, 2012).

In the scope of our work, the focus is put on the spectral behavior of land-cover and land-use classes. For performance evaluation, we use two benchmark datasets for which hyperspectral data and a semantic labeling are available on a per-pixel basis. The data representations derived via dimensionality reduction, feature selection or the transfer to multispectral Sentinel-2-like data thus also refer to this image raster. To address effects arising from the spatial resolution of the data, a transfer of both the given hyperspectral data and the given labeling to a larger ground sampling distance of 20 m as given for Sentinel-2 data would be required. Such considerations are beyond the scope of this paper and remain subject of future work.

6. CONCLUSIONS

In this paper, we have focused on the task of land-cover and land-use classification based on hyperspectral and multispectral imagery. Thereby, we have demonstrated the potential of using the original hyperspectral data as well as the potential of applying well-established approaches for dimensionality reduction and feature selection. In addition, we have included a transfer of hyperspectral data to multispectral Sentinel-2-like data products that are relevant for the task of land-cover and land-use classification. As the basis for this transfer, we have used hyperspectral data acquired from airborne platforms with the ROSIS and AVIRIS devices, and we have applied a spectral resampling to Sentinel-2-like data products by using the spectral response functions given for the Sentinel-2 MultiSpectral Instrument. For each case, a Random Forest classifier has been used to assign semantic labels with respect to a set of user-defined classes. The achieved results 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 as e.g. available with the future EnMAP mission seems to only provide a slight improvement for such a task. This is quite important, because the use of Sentinel-2 data is particularly advantageous for large-scale land-cover and land-use mapping applications due to a significantly less expensive data acquisition and a much faster repetition of data acquisition for the same area.

In future work, we aim at deeper investigating the potential of multi-modal data for land-cover and land-use classification. This includes the combined use of hyperspectral and shape information, which has recently been proven to allow for a better analysis of urban areas in comparison to the use of only hyperspectral

data (Weinmann and Weinmann, 2018). Furthermore, we plan to investigate the potential of a time series analysis based on multitemporal Sentinel-2 data. This allows taking into account dynamic processes (e.g. caused by seasonal changes or growth cycles) as well as detecting changes in land-cover and land-use.

References

Breiman, L., 1996. Bagging predictors. *Machine Learning* 24(2), pp. 123–140.

Breiman, L., 2001. Random forests. *Machine Learning* 45(1), pp. 5–32.

Camps-Valls, G. and Bruzzone, L., 2005. Kernel-based methods for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing* 43(6), pp. 1351–1362.

Camps-Valls, G., Tuia, D., Bruzzone, L. and Benediktsson, J. A., 2014. Advances in hyperspectral image classification: Earth monitoring with statistical learning methods. *IEEE Signal Processing Magazine* 31(1), pp. 45–54.

Chehata, N., Le Bris, A. and Najjar, S., 2014. Contribution of band selection and fusion for hyperspectral classification. In: *Proceedings of the 7th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pp. 1–4.

Chen, C., Liaw, A. and Breiman, L., 2004. *Using random forest to learn imbalanced data*. Technical Report, University of California, Berkeley, USA.

Chen, Y., Jiang, H., Li, C., Jia, X. and Ghamisi, P., 2016. Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing* 54(10), pp. 6232–6251.

Chi, M., Feng, R. and Bruzzone, L., 2008. Classification of hyperspectral remote-sensing data with primal SVM for small-sized training dataset problem. *Advances in Space Research* 41(11), pp. 1793–1799.

Criminisi, A. and Shotton, J., 2013. *Decision forests for computer vision and medical image analysis*. Advances in Computer Vision and Pattern Recognition, Springer, London, UK.

Elbertzhagen, I., Thonfeld, F. and Menz, G., 2012. SVM-based agricultural land use assessment using Sentinel-2 simulation data. In: *Proceedings of the Sentinel-2 Preparatory Symposium*.

Fang, L., Li, S., Duan, W., Ren, J. and Benediktsson, J. A., 2015. Classification of hyperspectral images by exploiting spectral-spatial information of superpixel via multiple kernels. *IEEE Transactions on Geoscience and Remote Sensing* 53(12), pp. 6663–6674.

Fauvel, M., Chanussot, J. and Benediktsson, J. A., 2008. Adaptive pixel neighborhood definition for the classification of hyperspectral images with support vector machines and composite kernel. In: *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, pp. 1884–1887.

Gege, P., Bachmann, M. and Holzwarth, S., 2008. Airborne Remote Sensing – ROSIS Bonn 2008 Summary Report. Technical report, German Society for Photogrammetry, Remote Sensing and Geoinformation (DGPF).

Green, A. A., Berman, M., Switzer, P. and Craig, M. D., 1988. A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Transactions on Geoscience and Remote Sensing* 26(1), pp. 65–74.

Guyon, I. and Elisseeff, A., 2003. An introduction to variable and feature selection. *Journal of Machine Learning Research* 3, pp. 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., Chen, Y., Crawford, M. M. and Ghosh, J., 2005. Investigation of the random forest framework for classification of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing* 43(3), pp. 492–501.

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

Ilehag, R., Weinmann, M., Schenk, A., Keller, S., Jutzi, B. and Hinz, S., 2017. Revisiting existing classification approaches for building materials based on hyperspectral data. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Jyväskylä, Finland, Vol. XLII-3/W3, pp. 65–71.

Joelsson, S. R., Benediktsson, J. A. and Sveinsson, J. R., 2005. Random forest classifiers for hyperspectral data. In: *Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 160–163.

Kaufmann, H., Segl, K., Chabrillat, S., Hofer, S., Stuffler, T., Mueller, A., Richter, R., Schreier, G., Haydn, R. and Bach, H., 2006. EnMAP – A hyperspectral sensor for environmental mapping and analysis. In: *Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 1617–1619.

Kaufmann, H., Segl, K., Guanter, L., Hofer, S., Foerster, K.-P., Stuffler, T., Mueller, A., Richter, R., Bach, H., Hostert, P. and Chlebek, C., 2008. Environmental Mapping and Analysis Program (EnMAP) – Recent advances and status. In: *Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS)*, Vol. 4, pp. 109–112.

Keller, S., Braun, A. C., Hinz, S. and Weinmann, M., 2016. Investigation of the impact of dimensionality reduction and feature selection on the classification of hyperspectral EnMAP data. In: *Proceedings of the 8th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pp. 1–6.

Keller, S., Braun, A. C., Hinz, S. and Weinmann, M., 2017. Investigation of the potential of hyperspectral EnMAP data for land cover and land use classification. In: *Tagungsband der 37. Wissenschaftlich-Technischen Jahrestagung der DGPF*, pp. 110–121.

Le Bris, A., Chehata, N., Briottet, X. and Paparoditis, N., 2014. Use intermediate results of wrapper band selection methods: a first step toward the optimization of spectral configuration for land cover classifications. In: *Proceedings of the 7th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pp. 1–4.

Liaw, A. and Wiener, M., 2002. Classification and regression by randomForest. *R News* 2/3, pp. 18–22.

Licciardi, G., Marpu, P. R., Chanussot, J. and Benediktsson, J. A., 2012. Linear versus nonlinear PCA for the classification of hyperspectral data based on the extended morphological profiles. *IEEE Geoscience and Remote Sensing Letters* 9(3), pp. 447–451.

Mannel, S. and Price, M., 2012. Comparing classification results of multiseasonal TM against AVIRIS imagery – Seasonality more important than number of bands. *PFG – Photogrammetrie – Fernerkundung – Geoinformation* 5/2012, pp. 603–612.

Melgani, F. and Bruzzone, L., 2004. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing* 42(8), pp. 1778–1790.

Plaza, A., Benediktsson, J. A., Boardman, J. W., Brazile, J., Bruzzone, L., Camps-Valls, G., Chanussot, J., Fauvel, M., Gamba, P., Gualtieri, A., Marconcini, M., Tilton, J. C. and Trianni, G., 2009. Recent advances in techniques for hyperspectral image processing. *Remote Sensing of Environment* 113, pp. S110–S122.

Roscher, R. and Waske, B., 2014. Superpixel-based classification of hyperspectral data using sparse representation and conditional random fields. In: *Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 3674–3677.

Saeys, Y., Inza, I. and Larrañaga, P., 2007. A review of feature selection techniques in bioinformatics. *Bioinformatics* 23(19), pp. 2507–2517.

Segl, K., Guanter, L., Kaufmann, H., Schubert, J., Kaiser, S., Sang, B. and Hofer, S., 2010. Simulation of spatial sensor characteristics in the context of the EnMAP hyperspectral mission. *IEEE Transactions on Geoscience and Remote Sensing* 48(7), pp. 3046–3054.

Spoto, F., Sy, O., Laberinti, P., Martimort, P., Fernandez, V., Colin, O., Hoersch, B. and Meygret, A., 2012. Overview of Sentinel-2. In: *Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 1707–1710.

Tarabalka, Y. and Tilton, J. C., 2011. Spectral-spatial classification of hyperspectral images using hierarchical optimization. In: *Proceedings of the IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pp. 1–4.

Tarabalka, Y., Chanussot, J., Benediktsson, J. A., Angulo, J. and Fauvel, M., 2008. Segmentation and classification of hyperspectral data using watershed. In: *Proceedings of the IEEE Geoscience and Remote Sensing Symposium (IGARSS)*, pp. III:652–III:655.

Tarabalka, Y., Fauvel, M., Chanussot, J. and Benediktsson, J. A., 2010. SVM- and MRF-based method for accurate classification of hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing* 7(4), pp. 736–740.

Thonfeld, F., Feilhauer, H. and Menz, G., 2012. Simulation of Sentinel-2 images from hyperspectral data. In: *Proceedings of the Sentinel-2 Preparatory Symposium*.

van der Maaten, L. J. P., Postma, E. O. and van den Herik, H. J., 2009. Dimensionality reduction: a comparative review. Technical report, Tilburg University, Tilburg, The Netherlands.

van der Meer, F. D., van der Werff, H. M. A. and van Ruitenbeek, F. J. A., 2014. Potential of ESA's Sentinel-2 for geological applications. *Remote Sensing of Environment* 148, pp. 124–133.

Villa, A., Benediktsson, J. A., Chanussot, J. and Jutten, C., 2011. Hyperspectral image classification with independent component discriminant analysis. *IEEE Transactions on Geoscience and Remote Sensing* 49(12), pp. 4865–4876.

Wang, J. and Chang, C.-I., 2006. Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis. *IEEE Transactions on Geoscience and Remote Sensing* 44(6), pp. 1586–1600.

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

Weinmann, M. and Weidner, U., 2018. Erste Erfahrungen mit der Nutzung von Sentinel-2 Daten zur multispektralen und multitemporalen Analyse mit verschiedenen Klassifizierungsverfahren. In: *Tagungsband der 38. Wissenschaftlich-Technischen Jahrestagung der DGPF und PFGK18 Tagung*, pp. 58–70.

Weinmann, M. and Weinmann, M., 2018. Geospatial computer vision based on multi-modal data – How valuable is shape information for the extraction of semantic information? *Remote Sensing* 10(1), pp. 2:1–2:20.

Zhao, Z., Morstatter, F., Sharma, S., Alelyani, S., Anand, A. and 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.*