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; Saëys 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) (Lucciardi 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; Saeyts 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; LeBris 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 (LeBris 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 Instrument1 (MSI) are described in Table 1 and the spectral response functions corresponding to these spectral bands are visualized in Figure 1.

1https://sentinel.esa.int/documents/ (last access: 22 February 2018)
an orthogonal transformation to transform a set of 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 $\hat{\rho}_{fc}$ and the average correlation between different features by $\hat{\rho}_{ff}$, the relevance $R$ of a feature subset comprising $n$ features is defined according to

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

whereby the involved correlation metric $\rho$ 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

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

<table>
<thead>
<tr>
<th>Spectral band</th>
<th>Central wavelength [nm]</th>
<th>Bandwidth [nm]</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1</td>
<td>433</td>
<td>20</td>
</tr>
<tr>
<td>B2</td>
<td>490</td>
<td>65</td>
</tr>
<tr>
<td>B3</td>
<td>560</td>
<td>35</td>
</tr>
<tr>
<td>B4</td>
<td>665</td>
<td>30</td>
</tr>
<tr>
<td>B5</td>
<td>705</td>
<td>15</td>
</tr>
<tr>
<td>B6</td>
<td>740</td>
<td>15</td>
</tr>
<tr>
<td>B7</td>
<td>783</td>
<td>20</td>
</tr>
<tr>
<td>B8</td>
<td>842</td>
<td>115</td>
</tr>
<tr>
<td>B8a</td>
<td>865</td>
<td>20</td>
</tr>
<tr>
<td>B9</td>
<td>945</td>
<td>20</td>
</tr>
<tr>
<td>B10</td>
<td>1375</td>
<td>20</td>
</tr>
<tr>
<td>B11</td>
<td>1610</td>
<td>90</td>
</tr>
<tr>
<td>B12</td>
<td>2190</td>
<td>180</td>
</tr>
</tbody>
</table>

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 $L_1, \ldots, L_D$, 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.

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).
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 \( \lambda_k \) of the \( k \)-th spectral band is given according to

\[
\lambda_k = \lambda_0 + (k - 1) \cdot \Delta \lambda \quad k = 1, \ldots, 115 \tag{2}
\]

and the spectral sampling is specified with \( \Delta \lambda = 4.0 \text{ nm} \) (Gege et al., 2008). The value of the constant \( \lambda_0 \) is approximated with \( \lambda_0 = 379.53 \text{ nm} \) following the aforementioned reference, so that the spectral bands reach from \( \lambda_1 = 379.53 \text{ nm} \) to \( \lambda_{115} = 835.53 \text{ nm} \) 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\(_2\)O) 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 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 × 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 × 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)
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 (OA = 95.46...96.72%), while the classification results achieved for the Salinas dataset are worse (OA = 84.77...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.

<table>
<thead>
<tr>
<th>Features</th>
<th>Pavia Centre</th>
<th>Salinas</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_i$</td>
<td>OA $\kappa$</td>
<td>$mF_1$</td>
</tr>
<tr>
<td>PCA($x_i$)</td>
<td>96.72 95.36 90.48</td>
<td>86.69 85.18 92.43</td>
</tr>
<tr>
<td>CFS($x_1$)</td>
<td>95.46 93.60 89.12</td>
<td>86.38 84.86 91.90</td>
</tr>
<tr>
<td>$S2(x_i)$</td>
<td>95.52 93.68 89.14</td>
<td>84.77 83.06 90.73</td>
</tr>
</tbody>
</table>

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 (OA = 95.46...96.72%) seems to be less challenging than the classification of the Salinas dataset (OA = 84.77...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 $mF_1$-score 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 PCA($x_i$) 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
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(x_i) leads to classification results of almost the same quality as obtained for the original hyperspectral data 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 $S_2(x_i)$ lead to classification results of a similar quality compared to $x_i$ and CFS($x_i$) 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 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


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

