Article

Crop Separability from Individual and Combined Airborne Imaging Spectroscopy and UAV Multispectral Data

Jonas E. Böhler *, Michael E. Schaepman and **Mathias Kneubühler**

Department of Geography, Remote Sensing Laboratories (RSL), University of Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland; michael.schaepman@geo.uzh.ch (M.E.S.); mathias.kneubuehler@geo.uzh.ch (M.K.)
* Correspondence: jonas.boehler@geo.uzh.ch; Tel.: +41-446-355-249

check for updates

Abstract: Crop species separation is essential for a wide range of agricultural applications—in particular, when seasonal information is needed. In general, remote sensing can provide such information with high accuracy, but in small structured agricultural areas, very high spatial resolution data (VHR) are required. We present a study involving spectral and textural features derived from near-infrared (NIR) Red Green Blue (NIR-RGB) band datasets, acquired using an unmanned aerial vehicle (UAV), and an imaging spectroscopy (IS) dataset acquired by the Airborne Prism EXperiment (APEX). Both the single usage and combination of these datasets were analyzed using a random forest-based method for crop separability. In addition, different band reduction methods based on feature factor loading were analyzed. The most accurate crop separation results were achieved using both the IS dataset and the two combined datasets with an average accuracy (AA) of >92%. In addition, we conclude that, in the case of a reduced number of IS features (i.e., wavelengths), the accuracy can be compensated by using additional NIR-RGB texture features (AA > 90%).

Keywords: crop separability; imaging spectroscopy; multispectral drone data; random forest; band reduction

1. Introduction

The accurate quantification of crop species in agricultural areas is crucial for various tasks such as decision-making and monitoring [1–3]. Many of these tasks require up-to-date information on current crop presences. However, official data are often only available at the end of or after the season, and when this is the case, they are usually aggregated by administrative units [4,5]. Remote sensing has therefore proved to be a viable alternative to human observations in the field [6,7].

Many studies of this kind are, however, conducted with satellite data [8,9], but the spatial resolution of the majority of these data is too low to provide accurate results at the field level in highly fragmented agricultural areas with small field plots [6,10,11]. Although satellite data with high spatial resolutions exist, they are often too expensive. Therefore, lower flying platforms are an option capable of collecting data at very high spatial resolutions. Using remote sensing, such studies for crop separation have been conducted with unmanned aerial vehicles (UAVs) [12–14] or airborne imaging spectrometers (IS) [15,16].

An UAV typically carries lightweight sensors, which often only contain a few spectral bands [17,18]. In addition, precise spectral calibration requires additional effort, which is subject to appropriate expertise [19]. Many UAVs carry a consumer-grade RGB camera to record data. These cameras may be modified to enable an additional near-infrared (NIR) channel to be acquired. NIR-RGB datasets with spatially very high resolutions (VHRs) can be generated from such platforms. However, additional

texture features are required to at least partially compensate for the lack of spectral information. Alternatively, imaging sensors with several bands are available, but they are costly [18].

Although IS sensors to be deployed on UAVs exist, more sophisticated sensors with very high signal-to-noise ratios are mainly operated from airborne platforms [19]. IS datasets are used in a range of applications in agriculture, i.e., biophysical properties [20,21], soil mapping [22], and, in particular, for crop separation [15,23]. Yet, the combined usage of IS and VHR datasets has so far been applied mainly in land cover studies in urban areas [24].

This study aims at evaluating the improvement of crop separation accuracy in very fragmented agricultural landscapes by combining VHR drone data and airborne imaging spectroscopy data. More specifically, the added value of the concurrent exploitation of spectral/textural features of VHR multispectral drone data and spectrally highly resolved imaging spectroscopy data is investigated and compared to single-usage scenarios of the two data sources.

Besides, this study assesses (i) the influence of feature reduction, as well as (ii) the effect of a model to fit the number of trees in the random forest (RF) model on the accuracy of crop separation.

2. Materials and Methods

2.1. Study Area

The study area is located in the Swiss Plateau next to Mönchaltorf (47.312 °N, 8.733 °E), and it is structured into numerous small plots (Figure 1). The fields have an average size of 1.3 ha (between 0.03 ha and 7.4 ha). Their lengths vary between 140 m and 200 m for a single field, and their widths between 23 m and 180 m. Cereals (winter barley, spelt, and winter wheat); rapeseed; grassland (permanent and temporary, including clover); maize; and sugar beet were included in the crop separation study. The field observations took place on 24 and 25 June 2015, corresponding to 1342 accumulated growing degree days (AGDD) and 1362 AGDD, respectively. Based on the scale of the Biologische Bundesanstalt, Bundessortenamt und Chemische Industrie (BBCH) [25], the phenological stage of a single reference field for maize (BBCH 33), rapeseed (BBCH 80), sugar beets (BBCH 39), and winter wheat (BBCH 75, representing cereal) were recorded at the time of observation.
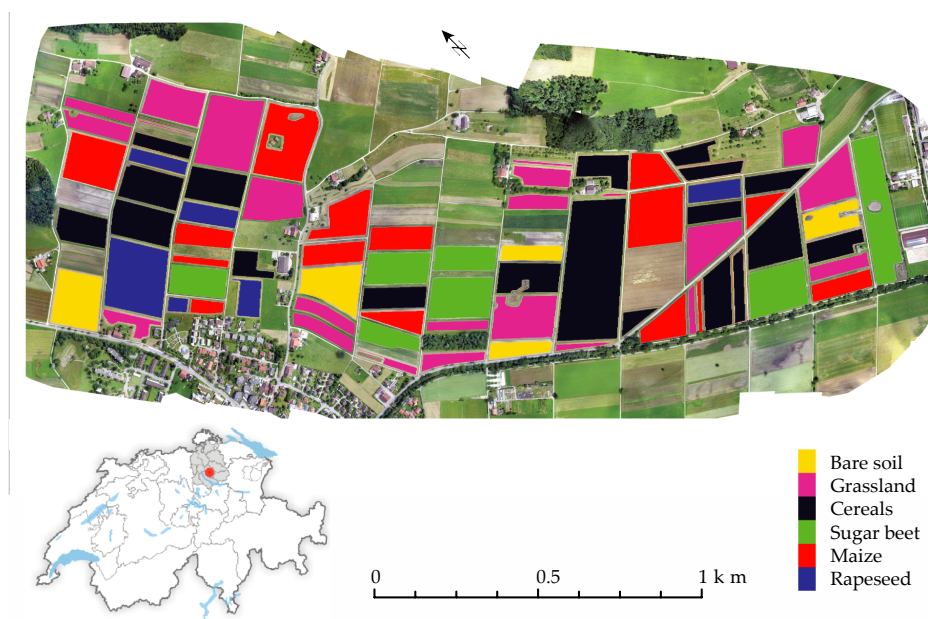
Figure 1. Mosaic of the study area (near Mönchaltorf, Switzerland; see overview map, **bottom left**) acquired on 25 June 2015 with an unmanned aerial vehicle (UAV)-mounted consumer-grade red green blue (RGB) camera. The crop classification as inventoried on the same date is superimposed (see legend, **bottom right**).

2.2. Data

The eBee dataset was based on uncalibrated VHR data acquired on 25 June 2015 with two consumer-grade cameras, i.e., an RGB and an NIR-GB, carried by an eBee-UAV (Sensefly, Cheseaux-Lausanne, Switzerland). The flight parameters were chosen in a way that the lateral overlap was 60%, and the longitudinal overlap 75%. The spatial resolution was 5 cm at a flight altitude of 150 m above the ground. Subsequently, an orthophotomosaic was built using a Pix4Dmapper Pro (version 4.2.27, Pix4D S.A., Prilly, Switzerland), and the NIR band was stacked to the RGB mosaic. This dataset was georeferenced to five ground control points (GCP), which were recorded with a differential GPS device (dGPS). Finally, the VHR dataset was resampled to a spatial resolution of 0.5 m, as this resolution proved to be the most promising spatial resolution for the investigated crops in our study area [26].

The APEX dataset was acquired the day before (24 June 2015) with the Airborne Prism EXperiment (APEX) sensor, an airborne imaging spectrometer (IS). A detailed description of the sensor properties and preprocessing chain can be found in [27,28]. The dataset was atmospherically corrected and orthorectified using a parametric geocoding approach [29]. The surface reflectance data cube contained 284 spectral bands in the range of 399–2431 nm at a spatial resolution of 2 m. Bands subject to atmospheric water vapor (i.e., the spectral ranges of 691–737 nm, 753–771 nm, 790–839 nm, 900–1008 nm, 1097–1174 nm, 1300–1513 nm, and 1753–2.050 nm) were interpolated during preprocessing and, subsequently, omitted, resulting in a dataset of 173 bands. The APEX dataset was co-registered to the eBee dataset of 25 June.

2.3. Methods

The chosen methodology of this study consists of seven steps (Figure 2). First, the input features for the random forest (RF) classifier were created on the basis of the two data sources. In the second step, the features were arranged in three settings (single and combined settings). In the third step, the dataset of each setting was split for training, validation, and testing. In the fourth step, the training split was used to calculate the feature importance and perform feature selection. In the fifth step, training of the RF model was performed. The free parameters for the RF model in the fifth step were determined on the basis of the validation split. In the sixth step, the test split was used to test the learned model from the fifth step. Finally, the crop separation accuracy was assessed. The individual steps are described in detail in the following sections.

2.3.1. Feature Extraction

The inclusion of texture information from VHR data on top of spectral information is known to improve classification results [26,30]. Therefore, first-order statistics (mean, standard deviation, range, and entropy) and mathematical morphology (dilatation/erosion, opening/closing, opening/closing top hat, opening/closing by reconstruction, and opening/closing by reconstruction top hat) were calculated from the eBee dataset, based on a disc-shaped structuring element (SE), to remain rotationally invariant. The SE has a diameter of 3 or 5 pixels. In order to form the final stack of features with the selected APEX bands, all features of the eBee dataset were resampled to a spatial resolution of 2 m by averaging.

2.3.2. Creation of Feature Settings

In order to analyze the influence of features from the two data sources (i.e., UAV-based VHR data and airborne IS data) on crop separation accuracy, three feature settings were created. The *eBee setting* contains only the eBee features, the *APEX setting* only the APEX features, and the *eBee & APEX setting* all eBee and APEX features.
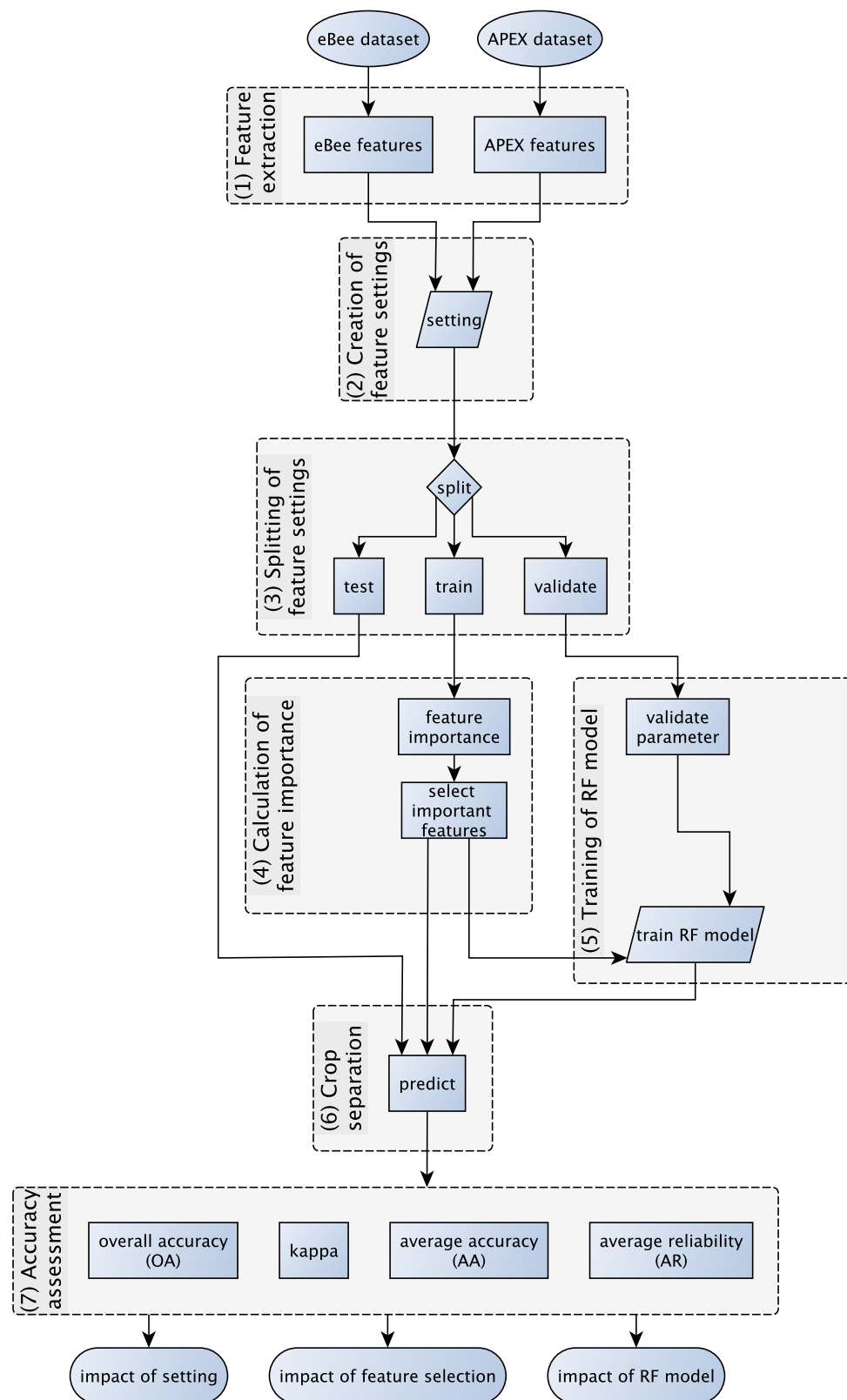
Figure 2. Data-processing scheme to assess crop separation accuracies from airborne imaging spectroscopy (IS) data, UAV-based multispectral data, and their combined usage. RF: random forest.

2.3.3. Splitting of Feature Settings

All three feature-setting datasets were divided class-wise into six splits. Two splits were selected for the RF feature selection and model training and two for validating the parameters in the RF model. In order to reduce the computational load, 1000 pixels were randomly selected for each crop class in the training and validation splits. The remaining two splits were retained for testing the learned model, resulting in a total of 15 permutations (so-called folds) [31].

2.3.4. Calculation of Feature Importance

In order to reduce the number of features in the RF model, the features were ordered into an ordered feature stack based on the sum of the factor loading of each crop class. Then, all features with higher correlation than a given correlation threshold between the first feature in the ordered feature stack and the remaining investigated features were excluded. This procedure was repeated with each remaining feature of the ordered feature stack until all features that correlate up to the threshold were excluded from the feature stack. Eleven evenly distributed threshold values between 0% and 100% were evaluated.

2.3.5. Training of RF Model

Two RF models were tested in our study. In the simple one, called *500-trees*, 500 trees were used to train the model with all data from the training and validation split that were selected from the feature stack based on the respective correlation threshold. In the second model, called *fitted-trees*, the appropriate number of trees was determined by building models with 20 logarithmically distributed grid points between 10 and 1000 trees. The trained models predicted the crop classes from the validation split. This procedure was repeated five times to obtain stable results. The resulting overall accuracy (OA) values were fitted by a curve, and the number of trees was determined, such that the loss in accuracy was less than 0.1% compared to the maximum accuracy of the fitted curve [26]. However, a minimum of 100 trees was set. In both models, the minimal leaf size was set to 3 to avoid overfitting. Default values were kept for the other parameters of the TreeBagger function in MATLAB (2018a).

2.3.6. Crop Separation

To test the previously trained model, the test data split was predicted for each of the 15 folds.

2.3.7. Accuracy Assessment

Based on the confusion matrix of the predicted test split, the average accuracy (AA) was calculated as the average AA over all folds. This allowed verifying the accuracy of crop separation of the different settings against each other for significant differences ($p < 0.05$). For this purpose, the Wilcoxon signed rank test was used. The results for the overall accuracy (OA), Kappa coefficient, and average reliability (AR) are contained in the Supplementary Materials (Tables S1–S3).

3. Results

The AA of crop separation depends on (i) the correlation threshold between the investigated features in the ordered feature stack, (ii) the RF model, and (iii) the feature setting. The results section is divided into three parts. First, the accuracy for crop separation is shown in relation to the correlation threshold. In the second part, the effects of the two evaluated RF models are highlighted, and in the third part, the differences between the three tested feature settings are presented. The exact p-values of the significance tests can be found in Tables S4–S9.

3.1. Feature Selection

All in all, crop separation in datasets with the *eBee setting* are less accurate than from settings also containing APEX data (Figure 3). The accuracy with the *eBee setting* ranges from an AA value of

80.8% for a correlation threshold between the investigated features of 90% to 43.5% for a correlation threshold of 0% in the *fitted-trees model* (Table 1). The individual AA values in this setting usually differ significantly (Table S4). Only the accuracies with a correlation threshold of 80% and 100% show no significant difference to each other, and the AA with a correlation threshold of 60% shows no significant difference to the ones with a correlation threshold of 30% or 40% (Table S4). With a correlation threshold of 90%, the amount of features can be reduced by more than 50% to an average of 56.3 features across all folds (Table 2), compared to a maximum of 116 features (correlation threshold of 100%). In addition, the reduced number of features leads to a significantly higher degree of crop separation. The complete list of selected features can be found in the Supplementary Materials (Table S10).
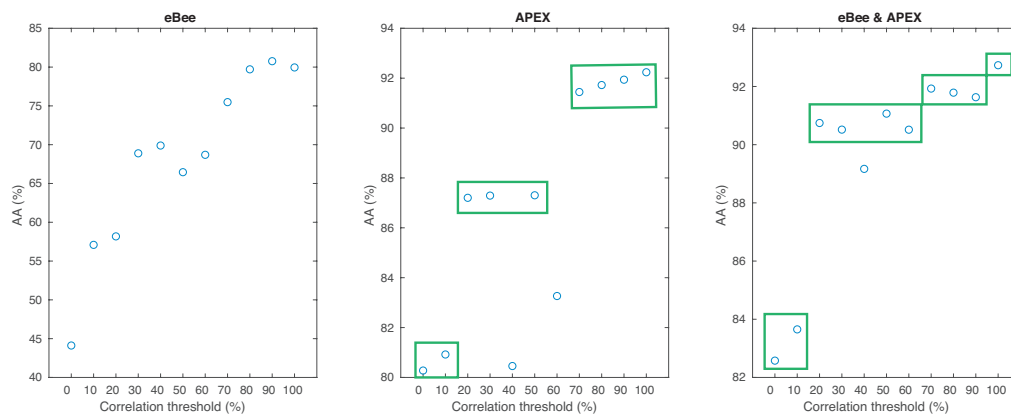
Figure 3. Average accuracy (AA) for the eBee, Airborne Prism EXperiment (APEX) and *eBee & APEX settings* for all evaluated correlation thresholds between the investigated features classified with the *fitted-trees model*. Clusters of similar values are marked in green.

Table 1. Average accuracy (AA) for all investigated settings, random forest (RF) models, and correlation thresholds between the investigated features. APEX: Airborne Prism Experiment.

Setting	Model	Correlation Threshold (%)										
		0	10	20	30	40	50	60	70	80	90	100
eBee	500-trees	44.1	57.1	58.2	68.9	69.9	66.4	68.7	75.5	79.6	80.9	79.9
eBee	fitted-trees	44.1	57.1	58.2	68.9	69.9	66.5	68.7	75.5	79.7	80.8	80.0
APEX	500-trees	80.3	81.0	87.2	87.3	80.4	87.3	83.3	91.5	91.8	91.9	92.2
APEX	fitted-trees	80.3	80.9	87.2	87.3	80.5	87.3	83.3	91.4	91.7	91.9	92.2
eBee & APEX	500-trees	82.6	83.6	90.8	90.6	89.3	91.2	90.6	92.0	91.8	91.8	92.8
eBee & APEX	fitted-trees	82.6	83.7	90.7	90.5	89.2	91.1	90.5	91.9	91.8	91.6	92.7

Table 2. Number of features for all investigated settings and correlation thresholds between the investigated features.

Setting	Correlation Threshold (%)										
	0	10	20	30	40	50	60	70	80	90	100
eBee	4.6	6.5	7.7	9.4	12.7	16.2	19.0	24.5	30.1	56.3	116
APEX	3.0	3.1	3.6	3.6	3.1	3.7	4.4	7.1	7.5	9.5	173
eBee & APEX	4.1	7.5	9.9	12.0	16.5	19.5	23.9	31.3	38.1	65.9	289

Concerning the *APEX setting*, AA values for different correlation thresholds between investigated features may be clustered into three distinct groups of AA, i.e., 0%–10%, 20%–50%, and 70%–100%, on the basis of significant differences between these groups and visual structures (Figure 3). Exceptions are the correlation thresholds of 40% and 60%, which do not fit into the pattern of a decreasing accuracy with a decreasing correlation threshold. The cluster of low AA values with correlation thresholds of 0%–10% differs significantly from the other two clusters. The middle cluster with correlation thresholds

of 20%–50% shows AA values about 4% lower than the high cluster with correlation thresholds of 70%–100%. However, the AA values in the middle cluster do not differ significantly from the AA values at a correlation threshold of 70% and 80% (Table S5). The AA values in the high cluster all lie between 91.4% and 92.2% (Table 1). Nevertheless, some of them differ significantly from each other, i.e., the AA values for a correlation threshold of 70% and 80%, 80% and 90%, and 90% and 100% differ significantly in each case from the other two values in the high cluster (Table S5). Finally, the AA for a correlation threshold of 40% fits into the first cluster from 0% to 10%, and the AA for the 60% threshold only exhibits insignificant differences with the AA values with a correlation threshold of 20% (Table S5).

The total amount of APEX features of 173 will be reduced to 9.5 on average, with a correlation threshold between investigated features of 90% (Table 2). The amount of features is not always an integer, as the number of features selected for different training splits may vary in different folds. With a correlation threshold of 0%, the average number of APEX features is reduced to three (although the AA remains 80.3% (Table 1)). Selected features in the case of APEX are wavelength bands (Table 3).

Table 3. Selected APEX features (i.e., wavelengths) for different correlation thresholds between the investigated features in the *APEX setting*. The numbers in the table correspond to the frequency at which a given feature was selected in a fold per the correlation threshold.

Wavelength (nm)	Correlation Threshold (%)									
	0	10	20	30	40	50	60	70	80	90
399						1		15		15
413							9		15	
427										15
490								1		2
509										2
553	15	15	15	15	15			15	15	
581										15
594									5	
681		15	15	15		15	15	15		
684			9			15		15	15	15
688				9	1		15	15	15	
785										1
1051	15				15	15	15		2	
1261							1			15
1521								1		
1549									1	
1567										1
1666	15	15	15	15	15	9	2	15	15	15
2057					1		9		15	15
2356										15
2388									15	
2419								15		15
2432										2

In the combined setting, i.e., eBee & APEX, four clusters of AA values could be defined based on a combined visual and statistical assessment (Figure 3). The first, with correlation thresholds between 0% and 10%, contains AA values that are significantly lower than the others (Table S6). The AA values in the second cluster with correlation thresholds of 20%–60% are between 90.5% and 91.1% (Table 1) and do not differ significantly from each other. In the third cluster, with correlation thresholds of 70%–90%, the AA values are between 91.6% and 91.9% (Table 1) and, again, do not differ significantly from each other. The highest AA value in this cluster is 91.9 % with a correlation threshold of 70%. The fourth cluster contains only the highest AA value (92.7%) with a correlation threshold of 100%, which also differs significantly from the other AA values. Again, the AA value for a correlation threshold of 40%

is significantly lower than the others in the second cluster of 20%–60% but significantly higher than in the first cluster of 0%–10%. Furthermore, there are no significant differences between the AA values for a correlation threshold of 20% and 80%, 20% and 90%, and 50% and 90%.

The number of features for a correlation threshold of 70% is reduced to 31.3 on average (Table 1). Compared to a total of 289 features (173 APEX features and 116 eBee features), this relates to a reduction by more than 90%. The number of APEX features will be reduced by up to three features (with a correlation threshold of 10%), with only one APEX feature being selected in that case (Table 2 and Table S12). Only with correlation thresholds of 60% and 90% will additional APEX features be chosen in comparison to the *APEX setting* (Table 2 and Table S12). In the case of correlation thresholds of 60%, this amounts to 1.3 features and, at 90%, to 7.8 features (Table 2 and Table S12). The complete list of selected features in the *eBee & APEX setting* can be found in the Supplementary Materials (Tables S11 and S12).

3.2. RF Model

The estimated number of trees with the *fitted-trees model* is about 500 trees for the *APEX setting* and the *eBee & APEX setting* and 770 for the *eBee setting* (Table 4). The difference in the number of trees is large between the 15 folds in all three settings, and for all correlation thresholds between the investigated features, and ranges from 100 trees (the set minimum) to about 900 trees (Figure S1).

Table 4. Number of trees in the *fitted-trees model* for all investigated settings and correlation thresholds between the investigated features.

Setting	Correlation Threshold (%)										
	0	10	20	30	40	50	60	70	80	90	100
eBee	776	893	749	886	824	754	794	791	732	490	785
APEX	350	436	413	630	406	467	403	414	414	559	440
eBee & APEX	578	334	502	654	545	466	625	494	529	500	567

There is no significant difference in the *eBee setting* between the mean AA values of the *500-trees model* and the *fitted-trees model* (Table S7). In general, the *fitted-trees model* leads to equal or higher AA values, except for a correlation threshold of 90%, where the AA value for the *500-trees model* is 0.14% higher.

There are no significant differences between the two RF models in the case of the *APEX setting*, either (Table S7). With correlation thresholds of 40% and 90%, the AA values are 0.04% and 0.05% higher for the *fitted-trees model*, respectively. For the *eBee & APEX setting*, significantly different AA values are observed for the correlation thresholds of 30%, 40%, and 90%, with AA differences amounting to 0.10%, 0.13%, and 0.14%, respectively (Table 1). Nevertheless, the differences in the AA values between the two RF models are less than 0.15%.

3.3. Sensor Data Combination

Since the AA values differ only slightly between the two RF models, only the results for the *fitted-trees model* are compared here. AA values of the *APEX setting* are significantly higher than the ones of the *eBee setting* for all corresponding correlation thresholds between the investigated features (Table S8). AA values of the *APEX setting* with low correlation thresholds, i.e., 0%, 10%, and 40%, are not significantly different from AA values of the *eBee setting* with high correlation thresholds, i.e., 80%–100%.

The AA values of the *eBee & APEX setting* are always significantly higher than those of the *eBee setting*. Compared to the *APEX setting*, AA values of the *eBee & APEX setting* are significantly higher, with corresponding correlation thresholds of 0% and 60% between the investigated features (Table S9).

4. Discussion

4.1. Feature Selection

Overall, we observe a decrease in AA values with decreasing correlation thresholds between the investigated features (Figure 3). Nevertheless, depending on the datasets, reduction of the amount of features, as investigated in our study, is a reasonable approach, even with the RF algorithm known to be resistant to the curse of dimensionality (Hughes' phenomenon) [32] and, therefore, able to handle large feature sets, as there is no significant decrease in AA values with decreased amounts of features.

In particular, reducing the number of features for the *eBee setting* leads to the most accurate crop separation, with a correlation threshold of 90% (AA significantly different from AA with a correlation threshold of 100%, Table 1). In a previous study, we examined in detail the effects of differing spatial resolutions of an eBee dataset. We concluded that a spatial resolution of 0.5 m leads to the highest accuracy in crop separation [26]. It was further found that average accuracies remained the same whether texture features were used that were calculated directly from the data with a spatial resolution of 0.5 m or aggregated to a 2-m resolution (see Section 2.3.1). In the case of the *APEX setting*, in contrast, crop separation does not improve with lower correlation thresholds (i.e., fewer features). However, the accuracy does not significantly decrease with a correlation threshold of 90% compared to 100%. Hence, the number of APEX features can be reduced by over 95%, i.e., corresponding to a reduction from 173 to 9.5 features (Table 2).

Elaborating on the physical background of the selected features in an RF model, the relevant APEX features (Table 3) are situated in spectral ranges which are sensitive to (i) pigments (413 nm, 553 nm, and 594 nm); (ii) biophysical traits (684 nm, 688 nm, and 1051 nm); (iii) plant water (2057 nm); or (iv) lignin; cellulose; or senescent material (1549 nm, 1666 nm, and 2388 nm) [20,21]. It is worth noting that selected features are considered important for low correlation thresholds (e.g., 0%–40% for a feature at 553 nm in Table 3), while they are not important for higher correlation thresholds (e.g., 50%–60%). Table 3 indicates that APEX features at a wavelength of 681 nm are important for separating the present crops. Therefore, a significant decrease of the AA value with a correlation threshold of 40% compared to 30% can be observed. At a correlation threshold of 60%, the excluded feature at a wavelength of 1666 nm leads most likely to a drop in the AA value compared to AA values with higher or lower correlation thresholds.

Three main categories of eBee features in our study are based on first-order statistics, morphology, or spectral bands (Table S10). Remarkably, all original spectral bands are excluded in the feature selection for all correlation thresholds from 0% to 90%. Among the top 20 features (most used over all correlation thresholds), morphological features are much more common than first-order statistical features. Spectral bands used to calculate textural features were most frequently the NIR, followed by the green and red bands. The blue band occurs to be the least relevant band (selected only once). Both structuring element (SE) sizes occur almost equally frequently.

4.2. RF Model

The proposed *fitted-trees model* to estimate the appropriate number of trees in the RF classifier leads only to small differences in the AA compared to the *500-trees model*. The average number of trees is close to 500 trees, or even considerably higher, in the case of the *eBee setting*. However, the accurate determination of the number of trees in a preliminary estimation with the *500-trees model* requires additional effort (user input), whereas the *fitted-trees model* can automatically determine the number of trees for the RF model.

4.3. Comparison of Sensor Data Settings

Crop separation based on the *eBee setting* results in the lowest observed AA values for the present crops. The methodology in its current state only includes the original bands and texture features, i.e., the eBee features. However, implementing additional information on field boundaries would

most likely increase the AA values by about 10% [26]. This result would be consistent with another multispectral study that achieves an OA of 91.5% (AA 90.7%) for a separation of 10 different classes based on an object-based RF approach [33].

Separation of crops based on the *APEX setting* results in significantly better AA values compared to the *eBee setting* (Table 1). Only the low cluster of AA values for the *APEX setting* with correlation thresholds of 0%–10% leads to similar AA values as the *eBee setting* with correlation thresholds of 80%–100%. Therefore, it is possible to obtain similar crop separation results on the basis of a multispectral dataset obtained with consumer-grade cameras, compared to an IS dataset being drastically reduced to a few bands.

A recent study based on Hyperion data, which used up to 30 best-performing bands to classify maize, cotton, rice, soybeans, and winter wheat in different images, achieved an OA of over 90% for distinguishing two individual crops using a support-vector machine (SVM) based approach [15]. However, for three or more crops, the OA fell below 89%. Therefore, our approach presented here with the *APEX setting* may yield a more precise separation of crops. Reasons for these differences may include differences in the investigated crops, as well as differing sensor characteristics. Besides a different number of bands and center wavelengths, the Hyperion sensor has two independent optical paths for the visible and near-infrared (VNIR) and short-wavelength infrared (SWIR) spectral ranges and a spatial resolution of 30 m, compared to a single optical path and a resolution of 2 m in the case of the APEX spectrometer.

The most accurate separation of crops can be achieved with the combined *eBee & APEX setting* (Table 1). However, using all available features, there is no significant difference to the *APEX setting* (with a correlation threshold of 100%), even if the AA value for the combined *eBee & APEX setting* is slightly higher. Nevertheless, the *eBee & APEX setting* with a reduced number of features leads to more accurate results than in the case of the *APEX setting*. In particular, crop separation with a correlation threshold of 20% yields better results in the combined setting (AA of 90.7%), with only 9.9 features being required (average over all folds), which is equivalent to the *APEX setting* with a correlation threshold of 60% (AA of 83.3%). The number of features obtained by imaging spectroscopy (APEX features) may be reduced to 1.7 on average (correlation threshold of 30%, Table S12), with an AA value still remaining over 90% (AA of 90.5%). Therefore, eBee texture features based on NIR-RGB bands can effectively compensate for a reduced number of APEX spectral features (narrow bands). This is particularly important in the case of multispectral camera systems that only allow the selection of a few, narrow preconfigured bands in addition to an RGB image. In addition, the operation of UAVs is usually more cost-effective than using an aircraft [34,35]. Similarly, data acquisition is more flexible in terms of weather conditions and flight planning [17,36], while fidelity and accuracy are usually superior from less-frequent airborne data acquisition.

5. Conclusions

This study presents a methodology for separating crops in a highly fragmented landscape with small structured plots, as in the Swiss Plateau, based on two different datasets. On the one hand, a multispectral VHR dataset with red (R), green (G), blue (B) and near-infrared (NIR) bands and texture features thereof, obtained with consumer-grade cameras, was investigated. On the other hand, an airborne imaging spectroscopy (IS) dataset of 2-m spatial resolution and 173 spectral bands between 399 nm to 2431 nm was used.

The highest AA values of over 92% could be achieved with the IS features (the *APEX setting*) and the combination of IS and VHR features (the *eBee & APEX setting*). Overall, we conclude that the reduction of features based on factor loading (decreasing correlation thresholds) results in significantly lower accuracies. Especially for the IS dataset (the *APEX setting*), the AA values will drop to almost 80%. For the combined dataset (the *eBee & APEX setting*), the accuracy with less features is also significantly lower but remains above 90%. With the VHR dataset (the *eBee setting*), the crops can be separated with

a maximum accuracy of around 80%. The proposed automatic selection of RF parameters is as good as the preliminary estimation.

In summary, this paper concludes that the accuracy of crop separation based on IS data exceeds the accuracy based on an NIR-RGB dataset and its texture features. Nevertheless, if the number of used IS features is reduced, additional NIR-RGB texture features (the morphological features of the NIR G and R bands are most important) can compensate a decrease in crop separation accuracy.

Supplementary Materials: The following are available online at http://www.mdpi.com/2072-4292/12/8/1256/s1, Figure S1. Boxplot of the number of trees in the *fitted-trees model* for the *eBee setting* (left), *APEX setting* (middle), and *eBee & APEX setting* (right), Table S1. Overall accuracies (OA) for all settings, models, and correlation thresholds, Table S2. Kappa coefficient for all settings, models, and correlation thresholds, Table S3. Average reliability (AR) for all settings, models, and correlation thresholds, Table S4. The p-values of the significance test ($p < 0.05$) for the *eBee setting* with all correlation thresholds, Table S5. The p-values of the significance test ($p < 0.05$) for the *APEX setting* with all correlation thresholds, Table S6. The *p*-values of the significance test ($p < 0.05$) for the *eBee & APEX setting* with all correlation thresholds, Table S7. The *p*-values of the significance test ($p < 0.05$) for the differences between the *500-trees model* and the *fitted-trees model* for each setting with all correlation thresholds, Table S8. The p-values of the significance test ($p < 0.05$) for the comparison of the *eBee setting* versus the *APEX setting* with all correlation thresholds, Table S9. The p-values of the significance test ($p < 0.05$) for the comparison of the *eBee & APEX setting* versus the *APEX setting* with all correlation thresholds, Table S10. Selected eBee features for the different correlation thresholds in the *eBee setting*. We report the frequency how many folds a feature was selected for each correlation threshold. A feature is defined by its size of the structuring element (SE), feature type, and spectral band of the eBee sensor, Table S11. Selected eBee features for different correlation thresholds in the *eBee & APEX setting*. The numbers in the table give the frequency in how many folds of a feature were selected per correlation threshold. A feature is defined through the size of the structuring element (SE), feature type, and band of the eBee sensor, Table S12. Selected APEX features and the mean number of features per fold for different correlation thresholds in the *eBee & APEX setting*. The numbers in the table give the frequency in how many folds of a feature were selected per correlation threshold.

Author Contributions: J.E.B. designed the research and analyzed the data with scientific advice from M.K. and M.E.S., J.E.B. wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

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

References

1. Azar, R.; Villa, P.; Stroppiana, D.; Crema, A.; Boschetti, M.; Brivio, P.A. Assessing in-season crop classification performance using satellite data: A test case in Northern Italy. *Eur. J. Remote Sens.* **2016**, *49*, 361–380. [CrossRef]
2. Ozdarici-Ok, A.; Ok, A.; Schindler, K. Mapping of Agricultural Crops from Single High-Resolution Multispectral Images–Data-Driven Smoothing vs. Parcel-Based Smoothing. *Remote Sens.* **2015**, *7*, 5611–5638. [CrossRef]
3. Inglada, J.; Arias, M.; Tardy, B.; Morin, D.; Valero, S.; Hagolle, O.; Dedieu, G.; Sepulcre, G.; Bontemps, S.; Defourny, P. Benchmarking of algorithms for crop type land-cover maps using Sentinel-2 image time series. In Proceedings of the 2015 IEEE International Geoscience and Remote Sensing Symposium, Milan, Italy, 26–31 July 2015; pp. 3993–3996. [CrossRef]
4. Waldhoff, G.; Lussem, U.; Bareth, G. Multi-Data Approach for remote sensing-based regional crop rotation mapping: A case study for the Rur catchment, Germany. *Int. J. Appl. Earth Obs. Geoinf.* **2017**, *61*, 55–69. [CrossRef]
5. Bundesrat, D.S. *Verordnung über die Direktzahlungen an die Landwirtschaft*; The Swiss Federal Council: Berne, Switzerland, 2017; Volume 2013, pp. 1–152.
6. Atzberger, C. Advances in Remote Sensing of Agriculture: Context Description, Existing Operational Monitoring Systems and Major Information Needs. *Remote Sens.* **2013**, *5*, 949–981. [CrossRef]
7. Burke, M.; Lobell, D.B. Satellite-based assessment of yield variation and its determinants in smallholder African systems. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 2189–2194. [CrossRef]

8. Zhong, L.; Hu, L.; Zhou, H. Deep learning based multi-temporal crop classification. *Remote Sens. Environ.* **2019**, *221*, 430–443. [CrossRef]

9. Ouzemou, J.-E.; El Harti, A.; Lhissou, R.; El Moujahid, A.; Bouch, N.; El Ouazzani, R.; Bachaoui, E.M.; El Ghmari, A. Crop type mapping from pansharpened Landsat 8 NDVI data: A case of a highly fragmented and intensive agricultural system. *Remote Sens. Appl. Soc. Environ.* **2018**, *11*, 94–103. [CrossRef]

10. Vuolo, F.; Neugebauer, N.; Bolognesi, S.; Atzberger, C.; D'Urso, G. Estimation of Leaf Area Index Using DEIMOS-1 Data: Application and Transferability of a Semi-Empirical Relationship between two Agricultural Areas. *Remote Sens.* **2013**, *5*, 1274–1291. [CrossRef]

11. Whitcraft, A.K.; Becker-Reshef, I.; Killough, B.D.; Justice, C.O. Meeting earth observation requirements for global agricultural monitoring: An evaluation of the revisit capabilities of current and planned moderate resolution optical earth observing missions. *Remote Sens.* **2015**, *7*, 1482–1503. [CrossRef]

12. Zhang, J.; Zhao, Y.; Abbott, A.L.; Wynne, R.H.; Hu, Z.; Zou, Y.; Tian, S. Automated mapping of typical cropland strips in the North China Plain using small Unmanned Aircraft Systems (sUAS) photogrammetry. *Remote Sens.* **2019**, *11*, 2343. [CrossRef]

13. Zhao, L.; Shi, Y.; Liu, B.; Hovis, C.; Duan, Y.; Shi, Z. Finer Classification of Crops by Fusing UAV Images and Sentinel-2A Data. *Remote Sens.* **2019**, *11*, 3012. [CrossRef]

14. Yan, Y.; Deng, D.; Liu, L.; Zhu, Z. Application of UAV-Based Multi-Angle Hyperspectral Remote Sensing in Fine Vegetation Classification. *Remote Sens.* **2019**, *11*, 2753. [CrossRef]

15. Aneece, I.; Thenkabail, P. Accuracies Achieved in Classifying Five Leading World Crop Types and their Growth Stages Using Optimal Earth Observing-1 Hyperion Hyperspectral Narrowbands on Google Earth Engine. *Remote Sens.* **2018**, *10*, 2027. [CrossRef]

16. Thenkabail, P.S.; Gumma, M.K.; Teluguntla, P.; Mohammed, I.A. Hyperspectral remote sensing of vegetation and agricultural crops. *Photogramm. Eng. Remote Sens.* **2014**, *80*, 697–709. [CrossRef]

17. Pajares, G. Overview and Current Status of Remote Sensing Applications Based on Unmanned Aerial Vehicles (UAVs). *Photogramm. Eng. Remote Sens.* **2015**, *81*, 281–330. [CrossRef]

18. Yao, H.; Qin, R.; Chen, X. Unmanned Aerial Vehicle for Remote Sensing Applications—A Review. *Remote Sens.* **2019**, *11*, 1443. [CrossRef]

19. Aasen, H.; Honkavaara, E.; Lucieer, A.; Zarco-Tejada, P. Quantitative Remote Sensing at Ultra-High Resolution with UAV Spectroscopy: A Review of Sensor Technology, Measurement Procedures, and Data Correction Workflows. *Remote Sens.* **2018**, *10*, 1091. [CrossRef]

20. Thenkabail, P.S.; Mariotto, I.; Gumma, M.K.; Middleton, E.M.; Landis, D.R.; Huemmrich, K.F. Selection of hyperspectral narrowbands (hnbs) and composition of hyperspectral twoband vegetation indices (HVIS) for biophysical characterization and discrimination of crop types using field reflectance and hyperion/EO-1 data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 427–439. [CrossRef]

21. Hank, T.B.; Berger, K.; Bach, H.; Clevers, J.G.P.W.; Gitelson, A.; Zarco-Tejada, P.; Mauser, W. Spaceborne Imaging Spectroscopy for Sustainable Agriculture: Contributions and Challenges. *Surv. Geophys.* **2019**, *40*, 515–551. [CrossRef]

22. Diek, S.; Chabrillat, S.; Nocita, M.; Schaepman, M.E.; de Jong, R. Minimizing soil moisture variations in multi-temporal airborne imaging spectrometer data for digital soil mapping. *Geoderma* **2019**, *337*, 607–621. [CrossRef]

23. Marshall, M.; Thenkabail, P. Advantage of hyperspectral EO-1 Hyperion over multispectral IKONOS, GeoEye-1, WorldView-2, Landsat ETM+, and MODIS vegetation indices in crop biomass estimation. *ISPRS J. Photogramm. Remote Sens.* **2015**, *108*, 205–218. [CrossRef]

24. Gamba, P.; Dell'Acqua, F. 18 Spectral Resolution in the Context of Very High Resolution Urban Remote Sensing. In *Urban Remote Sensing*; Weng, Q., Quattrochi, D., Gamba, P.E., Eds.; CRC Press: Boca Raton, FL, USA, 2018; p. 377.

25. Meier, U.; Bleiholder, H.; Buhr, L.; Feller, C.; Hack, H.; Heß, M.; Lancashire, P.D.; Schnock, U.; Stauß, R.; van den Boom, T.; et al. The BBCH system to coding the phenological growth stages of plants-history and publications-Das BBCH-System zur Codierung der phänologischen Entwicklungsstadien von Pflanzen-Geschichte und Veröffentlichungen. *J. Für Kult.* **2009**, *61*, 41–52. [CrossRef]

26. Böhler, J.; Schaepman, M.; Kneubühler, M. Crop Classification in a Heterogeneous Arable Landscape Using Uncalibrated UAV Data. *Remote Sens.* **2018**, *10*, 1282. [CrossRef]

27.   Hueni, A.; Lenhard, K.; Baumgartner, A.; Schaepman, M.E. Airborne Prism Experiment Calibration Information System. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 5169–5180. [CrossRef]

28.   Schaepman, M.E.; Jehle, M.; Hueni, A.; D'Odorico, P.; Damm, A.; Weyermann, J.J.; Schneider, F.D.; Laurent, V.; Popp, C.; Seidel, F.C.; et al. Advanced radiometry measurements and Earth science applications with the Airborne Prism Experiment (APEX). *Remote Sens. Environ.* **2015**, *158*, 207–219. [CrossRef]

29.   Hueni, A.; Biesemans, J.; Meuleman, K.; Dell'Endice, F.; Schlapfer, D.; Odermatt, D.; Kneubuehler, M.; Adriaensen, S.; Kempenaers, S.; Nieke, J.; et al. Structure, Components, and Interfaces of the Airborne Prism Experiment (APEX) Processing and Archiving Facility. *Geosci. Remote Sens. IEEE Trans.* **2009**, *47*, 29–43. [CrossRef]

30.   Laliberte, A.S.; Rango, A. Texture and Scale in Object-Based Analysis of Subdecimeter Resolution Unmanned Aerial Vehicle (UAV) Imagery. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 761–770. [CrossRef]

31.   Böhler, J.E.; Schaepman, M.E.; Kneubühler, M. Optimal Timing Assessment for Crop Separation Using Multispectral Unmanned Aerial Vehicle (UAV) Data and Textural Features. *Remote Sens.* **2019**, *11*, 1780. [CrossRef]

32.   Belgiu, M.; Drăgu, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [CrossRef]

33.   Lee, R.-Y.; Chang, K.-C.; Ou, D.-Y.; Hsu, C.-H. Evaluation of crop mapping on fragmented and complex slope farmlands through random forest and object-oriented analysis using unmanned aerial vehicles. *Geocarto Int.* **2019**, 1–18. [CrossRef]

34.   Colomina, I.; Molina, P. Unmanned aerial systems for photogrammetry and remote sensing: A review. *ISPRS J. Photogramm. Remote Sens.* **2014**, *92*, 79–97. [CrossRef]

35.   Hugenholtz, C.H.; Moorman, B.J.; Riddell, K.; Whitehead, K. Small unmanned aircraft systems for remote sensing and Earth science research. *EOS Trans. Am. Geophys. Union* **2012**, *93*, 236. [CrossRef]

36.   Whitehead, K.; Hugenholtz, C.H.; Myshak, S.; Brown, O.; LeClair, A.; Tamminga, A.; Barchyn, T.E.; Moorman, B.; Eaton, B. Remote sensing of the environment with small unmanned aircraft systems (UASs), part 2: Scientific and commercial applications 1. *J. Unmanned Veh. Syst.* **2014**, *2*, 86–102. [CrossRef]