Wickramarathna S *et al.* (2023)
Notulae Botanicae Horti Agrobotanici Cluj-Napoca
Volume 51, Issue 3, Article number 13227
DOI:10.15835/nbha51313227
Research Article

Assessment of herbaceous vegetation classification using orthophotos produced from the image acquired with unmanned aerial systems

Sudeera WICKRAMARATHNA[1], John GOETZ III[2], Jon SOUDER[1], Benjamin PROTZMAN[2], Brian SHEPARD[2], Sorin HERBAN[3], Francisco MAURO[4], Temesgen HAILEMARIAM[1], Bogdan M. STRIMBU[1,5]*

[1]*Oregon State University, College of Forestry, 3100 Jefferson Way, Corvallis OR 97333, USA; galapits@oregonstate.edu; jon.souder@oregonstate.edu; temesgen.hailemariam@oregonstate.edu; bogdan.strimbu@oregonstate.edu (*corresponding author)*
[2]*Clean Water Services, 2550 SW Hillsboro Hwy, Hillsboro OR 97123, USA; GoetzJ@CleanWaterServices.org; protzmanb@cleanwaterservices.org; shepardb@cleanwaterservices.org*
[3]*Polyethnic University Timisoara, Faculty of Civil Engineering, Traian Lalescu No. 2A, 300223 Timisoara, Romania; sorin.herban@upt.ro*
[4]*Universidad de Valladolid, Department of Plant Production, Campus Duques de Soria sn, 42005, Soria, Spain; francisco.mauro@uva.es*
[5]*National Institute of Research and Development for Biological Sciences, 296 Independenței Bd. District 6, 060031 – Bucharest, Romania*

Abstract

Arguably the most popular remote-sensing products are classified images. However, there are no definitive procedures to assess classification accuracy that simultaneously consider resources available and field efforts. The explosive usage of unmanned aerial systems (UAS) in land surveys adds new challenges to classification assessment, as orthorectified images usually contain significant artifacts. This study aims to identify the optimal ratio between training and validation sample size within a supervised classification approach applied to UAS orthophotos. As a case study, we used a wetland area west of Portland, OR, USA, treated with various glyphosate formulations to control *Phalaris arundinacea*, commonly known as reed canary grass. A completely randomized design with five replications and six glyphosate formulations was used to assess *P. arundinacea* vigor following repeated herbicide applications. The change in *P. arundinacea* vitality was monitored with high-resolution four-band imagery acquired with a SlantRange 3PX camera installed on a DJI Matrice 210. The orthophotos created from images were produced with Pix4D, which was subsequently preprocessed with ERDAS Imagine 2020 to reduce the noise, shadows, and artifacts. All images were classified with the maximum likelihood classification algorithm. Simple random and stratified random sampling methods were applied to collect training and validation samples, evaluating eight ratios of training to validation samples to assess their classification accuracy. We found that increasing the training-to-validation sample size ratio enhances accuracy, with the 3:1 ratio being the most reliable in classifying *P. arundinacea* vigor. Our study provides evidence that image preprocessing and enhancement are essential for UAS-based imagery.

Keywords: image classification; image enhancement; maximum likelihood classification; reed canary grass; *Phalaris arundinacea*; simple random sampling

Received: 28 May 2023. Received in revised form: 28 Jun 2023. Accepted: 30 Jun 2023. Published online: 05 Sep 2023.
From Volume 49, Issue 1, 2021, Notulae Botanicae Horti Agrobotanici Cluj-Napoca journal uses article numbers in place of the traditional method of continuous pagination through the volume. The journal will continue to appear quarterly, as before, with four annual numbers.

Introduction

Phalaris arundinacea (reed canary grass) is a cold-seasonal, perennial rhizomatous plant belonging to the Poaceae grass family (Heath and Barnes, 1985; Paveglio and Kilbride, 2000). *P. arundinacea* is a significant threat to wetland ecosystems in the Pacific Northwest region of the United States and is considered an invasive species due to its aggressive nature and rapid growth. *P. arundinacea* was introduced to Oregon in 1918 for its agronomical potential (i.e., pasture forage) and for erosion control. Subsequently, *P. arundinacea* became abundant in the Pacific Northwest, commonly in low-elevation wetlands, flood plains, and ditches. Its rapid growth and spread restrict vegetation diversity and impede ecological restoration efforts.

Controlling *P. arundinacea* is difficult due to its high vegetative and sexual reproduction capabilities, as well as its persistent rhizome system. Several methods are available to control *P. arundinacea,* including mowing, tilling, grazing, prescribing fire, introducing competitive species, and applying herbicides. However, most *P. arundinacea* control methods have proven ineffective except for herbicide treatments. Among the herbicides available to control *P. arundinacea*, glyphosate was identified by Adams and Galatowitsch (2006) to be one of the most effective. Inundation and dense vegetation complicate herbicide treatments (Sanna, 2020); and public resistance to their use led to a desire to identify the least glyphosate concentration that would effectively control *P. arundinacea*. We used remote sensing to evaluate the efficacy of experimental treatments using different herbicide formulations as it has shown application in forestry and agriculture to map species distribution and monitor invasive species (Kaskie *et al*., 2019).

Compared to conventional field survey methods, remote sensing techniques, particularly unmanned aerial systems (UAS), are more efficient and comprehensive in detecting and monitoring invasive species. One of the main features of using UAS in vegetation detection is the production of high-resolution orthoimages, which allows the delineation of small-sized plant species (Wang *et al*., 2021). The vast majority of the UAS is equipped with devices that have limited positioning abilities, namely GPS (i.e., location on Earth's Surface) and with respect to the ground (i.e., pitch, roll, and yaw). The main procedures to produce orthoimages from photos captured by sensors installed on UAS are based on structure from motion algorithms (SfM) (Schonberger and Frahm, 2016; Jiang *et al.*, 2020). However, the UAS-based orthoimages contain many artifacts that directly result from the combination of the lack of accuracy and precision of the equipment with the algorithm itself.

Remotely sensed products widely used in scientific investigations and decision-making are evaluated based on their intended use (Stehman and Czaplewski, 1998). Therefore, the assessment of a classified image became essential (Millard and Richardson, 2015). Various approaches may be employed for accuracy assessment, ranging from expert knowledge to contingency matrices based on sampling strategies (Lu and Weng, 2007). Many factors affect the assessment of the classified image (Foody, 2009; Millard and Richardson, 2015), including the size of the training sample (Congalton, 1991; Foody and Mathur, 2004), the number of classes in the classification (Pal and Mather, 2003), the ability of the training data to adequately characterize the classes being mapped (Foody and Mathur, 2004), and dimensionality of the data (Pal and Mather, 2003). Moreover, ground reference data play a significant role in supervised image classification (e.g., training), and they are generally assumed to be error-free, even though it is not necessarily true (Foody *et al*., 2016). Training samples are generally acquired from fieldwork or fine spatial-resolution images. The effects of reference data error on training a supervised classifier are less studied, and growing literature highlights various issues and concerns associated with reference data on image classification (e.g., Radoux *et al*., 2014; Powell *et al*., 2004).

Reference data is separated in two: one set for training and one set for validation. Reference data can be obtained using different collection strategies (i.e., a single pixel, seeds, or polygon) and sampling designs, which pose challenges to the classification and would influence classification results (Chen and Stow, 2002). For example, the variability induced by different sampling designs could cause subsequent issues, the most famous being unbalanced samples (Foody *et al*., 2016). Small reference data errors can lead to large errors when

assessing the accuracy of maps or producing estimates of class size. Foody *et al.* (2016) found that the estimated value of ecosystem services for the conterminous USA determined using the National Land Cover Database (2006) changed by almost 50% (i.e., from $1118 billion/year to $600 billion/year) after adjustment for known errors.

The core concept of collecting training data is to generate a set of statistics that describe the spectral response pattern for each landcover type to be classified (Lillesand *et al.*, 2015). UAS-based orthophotos are commonly analyzed using the same methods as aerial or space-borne images. We argue that the similarity in accuracy assessment starts after the UAS images are preprocessed, which ensures a comparable quality to these other types of images with respect to their spectral and spatial information. The main objective of the present study is to identify an appropriate sampling strategy, i.e., ratio between training and validation data, for classifying orthophotos created using SfM algorithms.

Materials and Methods

Our study area is the Jackson Bottom Wetlands Reserve, west Portland, Oregon, USA. This area is a former agricultural field currently dominated by *P. arundinacea*, *Raphanus raphanistrum* (wild radish), and other herbaceous species. Along the central drainage ditch, a set of 5 plots covering 1,300 m^2 were randomly located (Figure 1). To account for the spatial variability within the riparian area, we have replicated the treatments on five sites with concentrations of 2% glyphosate as the active ingredient (Rodeo), an adjuvant, modified vegetable oil (Competitor) with the addition of ammonium sulfate as a water conditioner (Kicker).

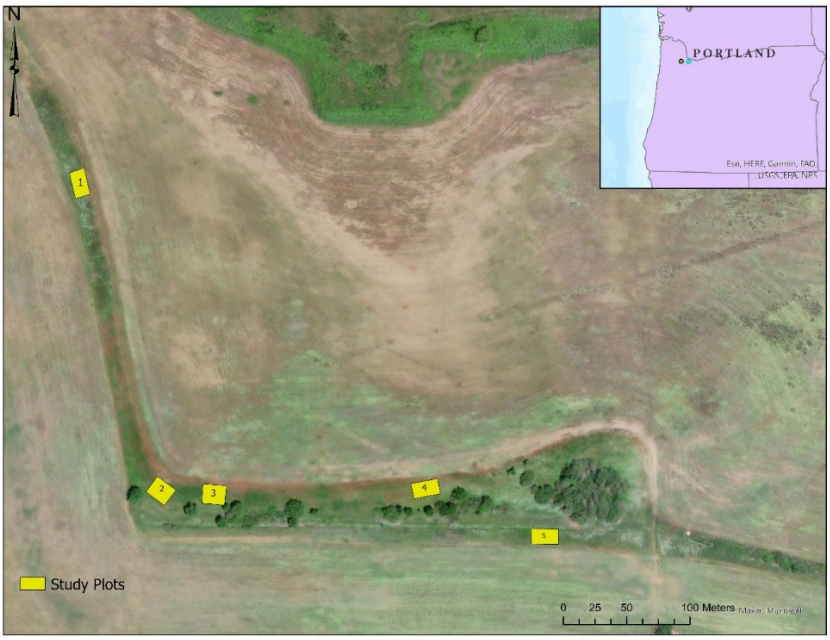

Figure 1. Study area and the treatment plots

UAS image collection and image processing

The UAS images were collected using a multi-rotor DJI Matrice 210 platform equipped with the multispectral sensor SlantRange 3PX. The sensor records spectral reflectance in four bands (i.e., green, red, red edge, and near-infrared) with a 2.5 cm spatial resolution. For ground validation, we also acquired true-color images (i.e., red, green, and blue bands) using a Zenmuse X4S camera, which has a spatial resolution of 1.8 cm.

To delineate the *P. arundinacea,* we carried out a two-stage procedure (Jensen, 1996; Lillesand *et al*., 2015), the first focused on preparing the image for analysis (i.e., preprocessing stage), and the second concentrated on extracting information from the image (i.e., the processing stage). Because the images captured with the UAS were not orthorectified, before starting the analysis, we combined them with the SfM algorithm implemented in the Pix4D software (Pix4D 2017). The orthophotos were preprocessed to eliminate artifacts and enhance the spectral and spatial signatures and finally classified (Figure 2).

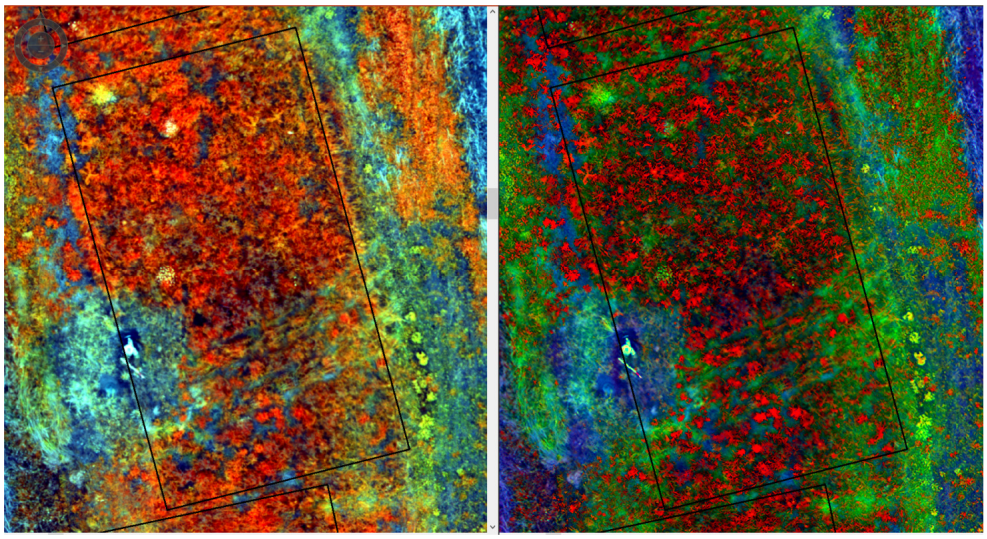

Figure 2. Raw images and unprocessed orthophotos were created from the images acquired with the SlantRange 3PX sensor

We preprocessed and enhanced the UAS-based orthophotos from the SfM algorithms to remove artifacts and errors by using routines within ERDAS Imagine 2020 (ERDAS IMAGINE Help, 2019). The best method(s) for image enhancement depend on the user's needs; no predetermined or standard method fits all situations. We used an 8-step preprocessing approach to create an image deemed suitable for classification. Four main image enhancement techniques, including Fourier transformation, radiometric, spatial, and spectral enhancements were applied to increase the amount of information extraction from the raw (unrefined) data (Figure 3). To improve classification accuracy, we added five spectral indices (Jensen, 1996) to the four-band images resulting from the preprocessing. These techniques amplify the visual distinctions or segregation among the features in the scene and improve visual interpretability (Lillesand *et al*., 2015).

The initial step for image enhancement is the Fourier Transform (FT) of the raw image, which decomposes the digital image into its sine and cosine components (Haque and Uddin, 2011). This tool provides options to reduce striping, noise, and other periodic anomalies associated with raster imagery. Editing the resulting Fourier image is an interactive process, and there is no specific rule(s) for determining what works best for the given image. Thus, the image analyst has to manipulate the FT analysis with the tools available in the Fourier editor to select the best techniques and parameters that perform well (ERDAS IMAGINE Help, 2019).

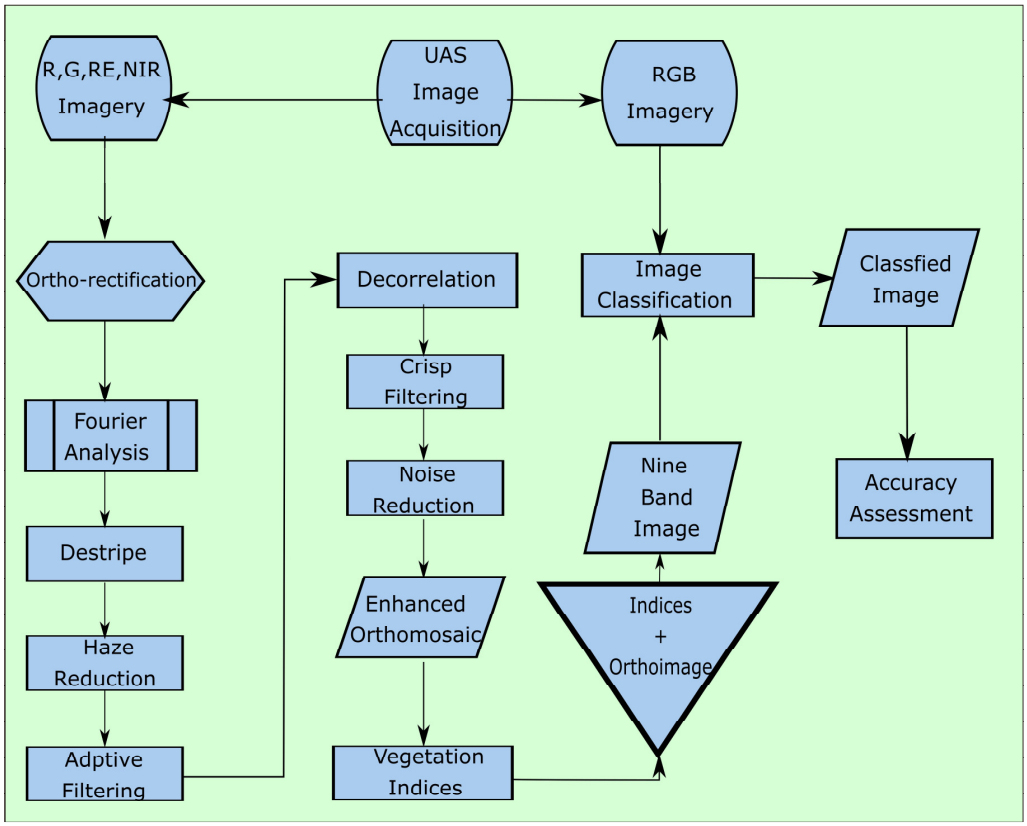

Figure 3. Flow diagram deployed for image per-processing and enhancement. Flow chart symbols are defined according to Hebb (2012)

We sharpened and edge-enhanced the features in the image using a high-pass filter to enervate the low-frequency components (i.e., bright areas) and pass the high-frequency components (i.e., dark areas); and then a wedge filter that extends from the center to the outer edge of the FT image to help remove the horizontal and vertical low-frequency components. Even after performing the FT analysis, the resulting orthomosaic images contained systematic horizontal banding patterns possibly introduced during the image mosaicing stage; these were removed by a de-strip algorithm.

To aid in identifying the different vegetation types, a Wallis adaptive filter was used to increase contrast and improve luminance using a Principle Components transformation (ERDAS IMAGINE Help, 2019; Noviello *et al*., 2013). This filter uses an analyst-specified moving window to scan the raster, stretching the contrast to alter the distribution of the image digital number values within the 0 - 255 range of the display device and utilize the full range of values linearly). The result of this procedure is that the clarity between each landcover type increased, allowing the analyst to demarcate the vegetation types more confidently.

The Crisp filter was used to additionally sharpen the images through convolution with an inverted point spread function (PSF) kernel. The PSF measures the image's blurring due to the sensor system's characteristics (ERDAS IMAGINE Help, 2019). This unique filter increases the fine details of the image, such as *P. arundinacea* clump edges, without affecting broader-scale features. This option also sharpens the overall luminance of the image.

Interference of atmospheric effects on image acquisition can limit the dynamic range of imagery and appear to have haze or reduced contrast. These effects can be reduced by applying a noise-removal filter that identifies noisy pixels by comparing each pixel with its surrounding pixels. This filter removes noise along edges in flat areas while preserving the subtle details in an image, such as thin lines (ERDAS IMAGINE Help, 2019).

Additional spectral enhancement techniques were applied to the multi-band images to generate new bands which are interpretable to the human eye (i.e., creating indices) through mathematical transformations and algorithms (ERDAS IMAGINE Help, 2019). These indices enhance or differentiate the landcover features that cannot be detected in the original color bands while minimizing shadow effects. For this study, we generated five indices: the Difference Vegetation Index (DVI), Normalized Difference Vegetation Index (NDVI), Green Normalized Difference Vegetation Index (GNDVI), Improved Modified Chlorophyll Absorption Ratio Index (MCARI), and Modified Red Edge Simple Ratio Index (MRESR). A nine band-enhanced ortho-mosaic resulted from the application of these multiple filters and algorithms.

Image classification and accuracy assessment

We identified five land cover classes for image classification to assess the efficacy of the glyphosate formulations to control *P. arundinacea*: healthy *P. arundinacea* (H), unhealthy *P. arundinacea* (U), ground (G), other vegetation (V), and shadow (S). Among the rich ecosystem of image classification procedures, we evaluated several that were parsimonious in terms of training and accuracy sample sizes. Procedures that require large samples, like most machine learning techniques, inherently preclude a formal assessment of the effort that should be placed into selecting these samples. We, therefore, focused on a classification approach that best provides accurate results as a function of training and validation data size. Several trial-and-error tests suggested that the maximum likelihood algorithm, a supervised classification approach, provided the desired traits. Our target population was the pixels of the preprocessed orthophoto that represent the five land cover classes.

To evaluate the accuracy of the image classification, Stehman & Czaplewski (1998) described three essential components: (i) sample design, (ii) response design, and (iii) estimation and analysis protocol. We considered a set of homogeneous pixels on each land cover class, a polygon, as population elements in the training stage, with the classified individual pixels considered population elements in the validation stage. We used in-situ information extracted through photo-interpretation of the high-resolution true-color images to identify training sample polygons. Given the small number of land cover classes (five) and the homogeneity of the land within each plot, we randomly distributed each class's training sites throughout each plot to acquire a set of spatially homogeneous data (Khorram *et al*., 2012). To ensure statistical validity (Strimbu, 2014), we collected 10 polygons for each category to adequately capture the variability in each land cover category without incurring significant field sample collection effort. This resulted in 50 training polygons (i.e., 10 polygons for each of the five classes) for each herbicide treatment plot, for a total of 250 polygons.

For validation, which is based on the information acquired during the response design, we considered individual pixels belonging to each class as the population elements. Previous studies showed that choosing the sample unit (for validation data) depends on the project objectives, landscape characteristics, and practical constraints (Seltman, 2015). We selected individual pixels, rather than sets of pixels, as the sample units to cope with the large-scale landscape features (e.g., *P. arundinacea*-dominated vegetation) and take advantage of the high spatial resolution of the orthophoto. To ensure consistency across the training and validation stages of the classification, we used simple random sampling to select a constant number of pixels from each class. This method allows stratifying a user-defined number of random points to the distribution of thematic layer classes. We collected 10 pixels for each land cover class. Because the response design involves collecting information relevant to the reference land cover, we used high-resolution true-color images and in-*situ* (field) collected photographs to validate the selected pixels. We chose this approach for validation because of the reduced size of each individual to be classified (e.g., leaves or flowers) and because previous studies by Janssen and Vanderwel (1994) and Franklin *et al*. (1991) showed that individual pixels are more appropriate for a pixel-based classification approach.

We assessed the quality of the classification by constructing a confusion matrix from which we computed the user (Eq. 1) and producer (Eq. 2) accuracy for each class, as well as the overall accuracy (Eq. 3), and Cohen's Kappa coefficient (Eq. 4). Producer accuracy is the probability that a value in a given class was classified correctly, whereas user accuracy is the probability that a value predicted to be in a particular class really

is that class. The Kappa coefficient measures the agreement between classification and truth values. For example, a Kappa coefficient of 1 represents a perfect agreement, while a value of 0 represents no agreement (Foody, 2002).

$$User\ Accuracy\ = \frac{Number\ of\ correctly\ classified\ pixels\ in\ each\ category}{Total\ number\ of\ pixes\ in\ that\ catagory} \quad 1$$

$$Producer\ Accuracy\ = \frac{Number\ of\ correctly\ classified\ pixels\ in\ each\ category}{Total\ number\ of\ training\ set\ pixels\ of\ that\ catagory} \quad 2$$

$$Overall\ Accuracy\ = \frac{Total\ number\ of\ correclty\ classifed\ pixels}{Total\ number\ of\ reference\ pixels} \quad 3$$

$$Kappa\ (k) = \frac{\mathrm{N}\sum_{i=1}^{n} m_{i,i} - \sum_{i=1}^{n}(G_i\ C_i\ )}{N^2 - \sum_{i=1}^{n}(G_i\ C_i\ )} \quad 4$$

where: i is the class number, N is the total number of classified values compared to truth values, mi, i is the number of values belonging to the truth class i that have also been classified as class i (i.e., values found along the diagonal of the confusion matrix), C_i is the total number of predicted values belonging to class i, and G_i is the total number of truth values belonging to class i.

Experimental design for assessing the sampling strategy

We used a partial factorial experiment of eight trials that covered a large range of combinations between training and validation plots to evaluate the performance of the supervised image classification (Table 1). Treatment plots 1 and 2 for used for training, and 3, 4, and 5 for validation, except when three plots were used for training, in which case plot 3 was used for training. For simplification, we convey the findings as the ratio between the number of plots used for training and the number of plots used for validation (Table 1).

Table 1. The experimental design was used to assess the sampling strategy

Trial	Training plot(s)	Validation plot(s)	Training samples	Validation samples	Ratio
1	1	3	5x10	5x10	1:1
2	1	3+4	5x10	5x20	1:2
3	1	3+4+5	5x10	5x30	1:3
4	1+2	3	5x10 +5x10	5x10	2:1
5	1+2	3+4	5x10 + 5x10	5x20	2:2
6	1+2	3+4+5	5x10 + 5x10	5x30	2:3
7	1+2+3	4	5x10 +5x10 +5x10	5x10	3:1
8	1+2+3	4+5	5x10 + 5x10 +5x10	5x20	3:2

We used ANOVA to identify the most suitable ratio of sizes between training and validation *(Winer et al*., 1991); and evaluated the normality assumptions using the Shapiro-Wilk test, the homoskedasticity assumption with the Levene test, and the independence assumption with the Durbin-Watson test. For the ANOVA, the dependent variable is one of the six assessment metrics, and the predictor variable is the ratio between the size of the training data and validation data (Eq. 5). Rather than using the ratios as actual values, we have grouped them in three classes: one which has the training dataset larger than validation dataset (labeled Larger), one that has the training and validation equal in size (labeled Even), and the last one that represents the case when training data is smaller than the validation data (labeled Smaller).

$$Assessment\ metric = Ratio\ training\ to\ validation\ data \quad (5)$$

Where the assessment metric is one of six variables: overall accuracy, the Kappa value, and the producer or the user accuracy associated with the healthy or unhealthy *P. arundinacea*; the number of training plots takes the value 1, 2, and 3, and the number of validation plots also takes the value 3, 4, or 5.

Results

The primary goal of the study is to use orthophotos to estimate the efficacy of various herbicide formulations to control *P. arundinacea*. Obtaining accurate classifications is vital to obtain valid estimates of *P. arundinacea* vigor among the experimental herbicide formulations. However, collecting training data requires methodological choices that are in essence tradeoffs between quality (i.e., high accuracy) and quantity (i.e., the effort to collect training and validation data) (Millard and Richardson, 2015). Determining the correct proportion of samples for the training and validation data is challenging because of the size of the individual objects being classified (i.e., leaves and flowers) as well as the land cover variance.

Sampling strategies using only one training dataset showed a gradual decrease in overall accuracy and Kappa-value when the validation sample size increased. The decrease is dramatic, as for a 1:1 ratio training – validation, the overall accuracy and Kappa values were 68% and 60%, respectively, whereas, for a 1:3 ratio, the same metrics were 37% and 20%, respectively. The same trend was observed for the user and producer accuracies of each class, which decreased when the training-to-validation sample ratio changed from 1:1 to 1:3 (Table 2).

Table 2. Assessment metrics for the classifications executed using maximum likelihood. The symbols in the table are H for healthy *P. arundinacea*, U for unhealthy *P. arundinacea*, G for ground, V for other vegetation, and S for shadow

Trial	Ratio	H	U	G	V	S	OA	Kappa	Accuracy type
1	1:1	81.8	61.5	54.5	66.7	100.0	68.0	60.2	Producer
		90.0	80.0	66.5	80.0	27.3			User
2	1:2	87.5	67.0	76.2	70.8	27.3	70.0	62.0	Producer
		88.0	66.7	72.7	68.0	33.4			User
3	1:3	60.0	37.0	32.0	33.4	13.5	37.0	20.0	Producer
		60.0	26.3	26.7	40.5	21.4			User
4	2:1	100.0	90.0	100.0	60.0	70.0	84.0	80.0	Producer
		83.3	81.8	90.9	66.7	100.0			User
5	2:2	81.8	77.8	60.0	46.2	50.0	63.0	54.0	Producer
		68.2	50.0	71.0	66.7	63.7			User
6	2:3	76.5	80.0	67.9	48.3	52.0	65.0	56.0	Producer
		72.0	53.0	76.0	51.0	88.2			User
7	3:1	90.0	100.0	80.0	100.0	30.0	80.0	75.0	Producer
		82.0	66.7	89.0	83.0	100.0			User
8	3:2	64.3	85.7	87.5	95.0	40.0	75.0	68.5	Producer
		90.0	60.0	88.7	82.6	66.8			User

When two training datasets were used, we observed a similar trend of decreasing overall accuracies and Kappa values, similar to one training dataset. A reduction in Kappa value and overall accuracy was noticed when the validation sample size increased (Table 2) from 80% to 56% and 84% to 65%, respectively. The decreasing trend in assessment statistics with the increase in the number of validation datasets was found when

three training datasets were used, as the overall accuracy reduced from 80% to 75% and the Kappa value from 75% to 68.5% (Table 2).

Similar to the overall accuracy and Kapaa coefficient, we observed enhancements of both producer and user accuracies at the class level with the increase in the number of training samples relative to the number of validation samples. Notably, an improvement in the healthy and unhealthy *P. arundinacea* classes was observed, which is helpful in achieving the study's primary goals (Table 2).

The ratio between training and validation data revealed no significant differences among the trials (p-value=0.17), which was surprising given the importance of training data. A close examination of the results pointed to two reasons for these results: the first is related to the small sample size (only eight cases), and the second is due to the large variability induced by some unrealistic scenarios, such as when one plot was used for training and three for validation (Table 1). We conducted a separate ANOVA, which considered only realistic cases (i.e., commonly encountered in applications); namely, the validation is not triple the size of the training data (Figure 4). In this situation, we found significant differences among the approaches (p-value = 0.027). The contrast occurs between the larger and even sampling plots (p-value = 0.03). Interestingly, the larger and smaller are not significantly different (p-value > 0.05), which is the direct result of the large variability still existing within each sampling strategy.

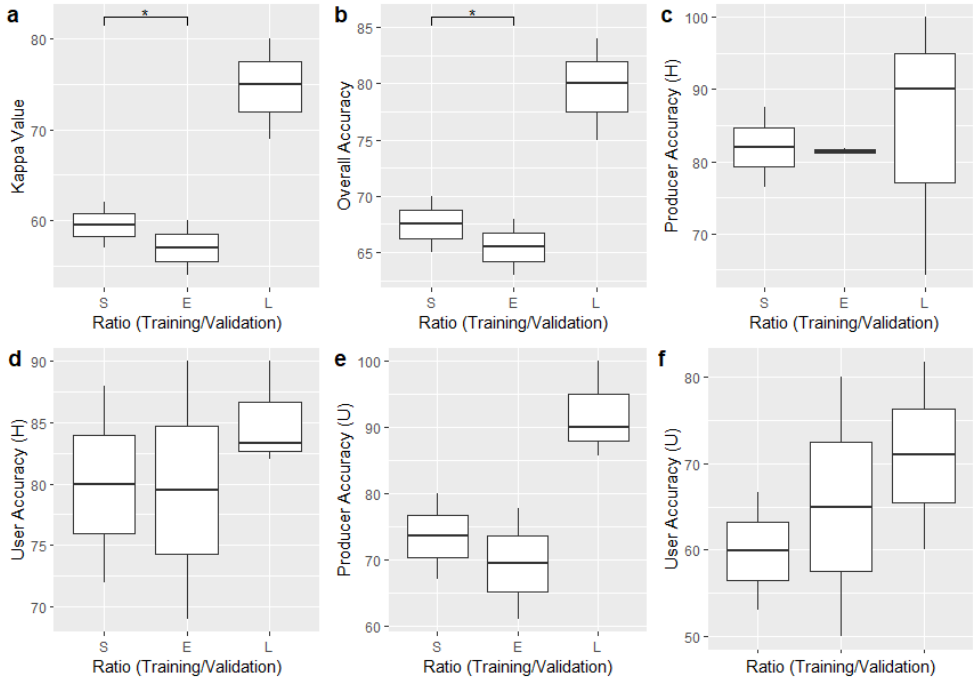

Figure 4. Box plot of the six assessment statistics across the ratio of the training and validation datasets sizes. The symbols are S for < 1 ratios, E for even ratios, and L for > 1 ratios. The accolades represent non-significant differences.

Discussion

Increasing the number of training samples while the number of validation samples is kept constant (i.e., increase the ratio training-validation) refines the land-cover classification, particularly for classes with closer characteristics (e.g., spectral reflectance and structural appearance). Furthermore, the acquisition of a larger

number of training samples can increase the robustness of the classifier in the allocation of the pixels in the appropriate class.

We have selected the image classifier following the recommendation of Sisodia *et al.* (2014), who argued that maximum likelihood is a robust approach because there are fewer chances of misclassification if the data is normally distributed (Lillesand *et al.,* 2015). However, the normal distribution of training samples is rarely achieved, so the outcomes would be affected by violating the Gaussian assumption. Furthermore, more training data leads to a narrower probability density function while reducing the overlaps between the central tendency describing each class. Narrower probability functions increase the probability of assigning a pixel to the appropriate class with confidence but reduce the chance of assigning all pixels to predefined classes (i.e., there will be many unclassified pixels).

When the ratio of training to validation samples is larger than 1, particularly the 3:1 and 3:2 ratios, we observed a lower producer accuracy for the shadow class compared to when the ratio is less than or equal to one. This result is due to the improvement in maximum likelihood performance when the training sample size is larger than the validation sample size. Previously classified pixels as a shadow in one or less training-validation ratios were allocated to other classes (i.e., healthy and unhealthy *P. arundinacea* or other vegetation), leading to a lower producer accuracy for class S (shadow).

Due to the terrain, sun angle, or vegetation type, airborne sensors capture lower reflectance values when there is a greater vegetative cover that produces shade. Consequently, image prepossessing and enhancing steps are essential before performing image classification, particularly for images acquired from UAS. Generally, we classify the low-radiance areas as shadows; therefore, some vegetation will be included as shadows. Using pre-processed images and a high sampling ratio would potentially reduce the effects of low radiance and eventually enhance image classification accuracy by distinguishing the fine spectral characteristics captured by the UAS sensors.

For equivalent sizes of training and validation data, we observed a high overall accuracy Kappa value and producer accuracy for the vigorous *P. arundinacea*. Decreasing the training-to-validation sample ratio lowered the overall performance and Kappa value (Figure 4). The ANOVA pointed to a significant decrease in the performances of all the assessment metrics when the ratio of training to validation increased above one. This supports the findings of other studies that maximum likelihood supervised classification performs better with a higher number of training samples.

Conclusions

Remote-sensing image classification is a complex process that requires consideration of many factors, among which the classification algorithm and selection of training and validation samples are critical choices (Lu and Weng, 2007). We argue that image preprocessing and enhancing are essential for the classification of orthophotos created using SfM algorithms, which is the case for almost all images acquired with UASs. Besides removing many artifacts created by the SfM, image preprocessing and enhancement allow the usage of spectral information even when less energy is available. Acquisition of a sufficient number of training samples and their representativeness is critical for image classifications (Landgrebe, 2003; Pal and Mather, 2003), especially when the landscape is complex and heterogeneous. The present study confirms that increasing the ratio between training and validation sample size enhances classification accuracy in complex vegetation. Supraunitary ratios (i.e., 2:1 or 3:1) are more reliable than unitary or subunitary ratios (e.g., 2:2 or 1:3) because the maximum likelihood classifier performs robustly with a higher number of training samples. The increase in the number of training samples narrows the probability density function describing each land class, which is essential in the separation of complex vegetation types and reduces the shadow effect.

Conventional supervised image classification approaches aim to fully describe all land cover classes based on their spectral characteristics (Sampat *et al*., 2005). Successful supervised classification combines a method

suitable to the data along with sufficient training samples (Lu and Weng, 2007). The ability to determine or quantify the functional and optimum number of training samples increases the efficiency of image classification. Throughout past decades scientists and practitioners have made progress in developing various classification techniques for improving classification accuracy (Stuckens *et al*., 2000; Pal and Mather, 2003; Gallego, 2004). However, classifying remotely sensed data into a thematic map remains a challenge because of many factors, such as the complexity of the landscape in a study area, selected remotely sensed data, and image-processing and classification approaches (Lu and Weng, 2007). In this study, we suggest that image preprocessing and establishing an appropriate training data collection procedure that is convenient for the study will lead to an accurate thematic map.

Authors' Contributions

Conceptualization: SW, BMS, FM, and HT; Data curation: SW; Formal analysis SW and BMS; Methodology: SW; Project administration JS and JG; Resources: JG, BP, and BS; Software: SH; Supervision: BS, JS, and JG; Validation: SW and BMS; Writing - original draft: SW; Review and editing: BMS, JS, JG, BP, BS, SH, FM, and TH. All authors read and approved the final manuscript.

Ethical approval (for researches involving animals or humans)

Not applicable.

Acknowledgements

The study was funded by a grant from Clean Water Services, Oregon, the U.S. Department of Agriculture, grant number 2019-67019-29462, and the Nucleus Program of the National Plan for Research, Development, and Innovation 2022-2027, funded by MCID, project 23020101, contract 7N from 03.01.2023.

Conflict of Interests

The authors declare that there are no conflicts of interest related to this article.

References

Adams CR, Galatowitsch SM (2006). Increasing the effectiveness of reed canary grass (*Phalaris arundinacea* L.) control in wet meadow restorations. Restoration Ecology 14(3):441-451. *https://doi.org/10.1111/j.1526-100X.2006.00152.x*

Chen D, Stow D (2002). The effect of training strategies on supervised classification at different spatial resolutions. Photogrammetric Engineering and Remote Sensing 68 (11):1155 -1162.

Congalton RG (1991). A review of assessing the accuracy of classifications of remotely sensed data. Remote Sensing of Environment 37(1):35-46. *https://doi.org/10.1016/0034-4257(91)90048-B*

ERDAS IMAGINE Help (2019). Retriveved 2022 December 15 from: *http://localhost:8080/imaginehelp/html/index.htm#ii_haze_hazediag.htm*

Foody GM (2002). Status of land cover classification accuracy assessment. Remote Sensing of Environment 80:185-201 *https://doi.org/10.1016/S0034-4257(01)00295-4*

Foody GM (2009). Sample size determination for image classification accuracy assessment and comparison. International Journal of Remote Sensing 30(20):5273-5291. *https://doi.org/10.1080/01431160903130937*

Foody GM, Mathur A (2004). Toward intelligent training of supervised image classifications: Directing training data acquisition for SVM classification. Remote Sensing of Environment 93(1):107-117. *https://doi.org/10.1016/j.rse.2004.06.017*

Foody GM, Pal M, Rocchini D, Garzon-Lopez CX, Bastin L (2016). The sensitivity of mapping methods to reference data quality: training supervised image classifications with imperfect reference data. ISPRS International Journal of Geo-Information 5(11):11. *https://doi.org/10.3390/ijgi5110199*

Franklin SE, Peddle DR, Wilson BA, Blodgett CF (1991). Pixel sampling of remotely sensed digital imagery. Computers & Geosciences 17(6):759-775. *https://doi.org/10.1016/0098-3004(91)90059-M*

Gallego FJ (2004) Remote sensing and land cover area estimation. International Journal of Remote Sensing 25:3019-3047. *https://doi.org/10.1080/01431160310001619607*

Haque MN, Uddin MS (2011). Accelerating Fast fourier transformation for image processing using graphics processing unit. Journal of Emerging Trends in Computing and Information Sciences 2(8):317-336.

Hebb N (2012). Flowchart symbols defined. BreezeTree Software.

Heath ME, Barnes RF (1985). Forages: The science of grassland agriculture. (No. 633.2 H42 1985).

Janssen LLF, Vanderwel FJM (1994). Accuracy assessment of satellite derived land-cover data: A review. Photogrammetric Engineering and Remote Sensing; (United States) 60:4. *https://www.osti.gov/biblio/6448244*

Jensen JR (1996). Introductory digital image processing: a remote sensing perspective (No. Ed. 2). Prentice-Hall Inc.

Jiang S, Jiang C, Jiang W (2020). Efficient structure from motion for large-scale UAV images: A review and a comparison of SfM tools. ISPRS Journal of Photogrammetry and Remote Sensing 167:230-251. *https://doi.org/10.1016/j.isprsjprs.2020.04.016*

Kaskie KD, Wimberly MC, Bauman PJ (2019). Rapid assessment of juniper distribution in prairie landscapes of the northern Great Plains. International Journal of Applied Earth Observation and Geoinformation 83:101946. *https://doi.org/10.1016/j.jag.2019.101946*

Khorram S, Koch FH, Wiele CF van der, Nelson SAC (2012). Remote Sensing. Springer Science & Business Media.

Landgrebe DA (2003). Signal Theory Methods in Multispectral Remote Sensing. John Wiley & Sons

Lillesand T, Kiefer RW, Chipman J (2015). Remote Sensing and Image Interpretation. John Wiley & Sons

Lu D, Weng Q (2007) A survey of image classification methods and techniques for improving classification performance. International Journal of Remote Sensing 28:823-870. *https://doi.org/10.1080/01431160600746456*

Millard K, Richardson M (2015). On the importance of training data sample selection in random forest image classification: a case study in peatland ecosystem mapping. Remote Sensing 7(7):7. *https://doi.org/10.3390/rs70708489*

Noviello M, Ciminale M, De Pasquale V (2013). Combined application of pansharpening and enhancement methods to improve archaeological cropmark visibility and identification in QuickBird imagery: two case studies from Apulia, Southern Italy. Journal of Archaeological Science 40:3604-3613. *https://doi.org/10.1016/j.jas.2013.04.013*

Pal M, Mather PM (2003). An assessment of the effectiveness of decision tree methods for land cover classification. Remote Sensing of Environment 86(4):554-565. *https://doi.org/10.1016/S0034-4257(03)00132-9*

Paveglio FL, Kilbride KM (2000). Response of vegetation to control of reed canary grass in seasonally managed wetlands of southwestern Washington. Wildlife Society Bulletin 28(3):730-740.

Pix4D (2017) Manual – Support. Retrieved 2023 May 2 from: *https://support.pix4d.com/hc/en-us/sections/360003718992-Manual*

Powell RL, Matzke N, de Souza C, Clark M, Numata I, Hess LL, Roberts DA (2004). Sources of error in accuracy assessment of thematic land-cover maps in the Brazilian Amazon. Remote Sensing of Environment 90(2):221-234. *https://doi.org/10.1016/j.rse.2003.12.007*

Radoux J, Lamarche C, Van Bogaert E, Bontemps S, Brockmann C, Defourny P (2014). Automated training sample extraction for global land cover mapping. Remote Sensing 6(5):5. *https://doi.org/10.3390/rs6053965*

Sampat MP, Bovik AC, Aggarwal JK, Castleman KR (2005) Supervised parametric and non-parametric classification of chromosome images. Pattern Recognition 38:1209-1223. *https://doi.org/10.1016/j.patcog.2004.09.010*

Sanna A (2020). Monitoring wetland invasive vegetation with drones: pilot study on reed canary grass. *https://expo.uw.edu/expo/apply/577/proceedings/result?student_name=astrid&commit=Search*

Särndal C-E, Swensson B, Wretman J (2003). Model Assisted Survey Sampling. Springer Science & Business Media.

Schonberger JL, Frahm JM (2016). Structure-from-motion revisited. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4104-4113.

Seltman HJ (2018). Experimental Design and Analysis. Carnegie Mellon University pp 428. Retrieved 2023 April 14 from: *https://www.stat.cmu.edu/~hseltman/309/Book/Book.pdf*

Sisodia PS, Tiwari V, Kumar A (2014). Analysis of supervised maximum likelihood classification for remote sensing image. International Conference on Recent Advances and Innovations in Engineering (ICRAIE-2014), pp 1-4. *https://doi.org/10.1109/ICRAIE.2014.6909319*

Stehman SV, Czaplewski RL (1998). Design and analysis for thematic map accuracy assessment: fundamental principles. Remote Sensing of Environment 64(3):331-344. *https://doi.org/10.1016/S0034-4257(98)00010-8*

Strimbu BM (2014). Comparing the efficiency of intensity-based forest inventories with sampling-error-based forest inventories. Forestry 87:249-255. *https://doi.org/10.1093/forestry/cpt061*

Stuckens J, Coppin PR, Bauer ME (2000). Integrating contextual information with per-pixel classification for improved land cover classification. Remote Sensing of Environment 71:282-296. *https://doi.org/10.1016/S0034-4257(99)00083-8*

Sun J, Yang J, Zhang C, Yun W, Qun J (2013). Automatic remotely sensed image classification in a grid environment based on the maximum likelihood method. Mathematical and Computer Modelling 58(3):573-581. *https://doi.org/10.1016/j.mcm.2011.10.063*

Wang L, Zhou Y, Hu Q, Tang Z, Ge Y, Smith A, Awada T, Shi Y (2021). Early detection of encroaching woody juniperus virginiana and its classification in multi-species forest using UAS imagery and semantic segmentation algorithms. Remote Sensing 13(10):10. *https://doi.org/10.3390/rs13101975*

Winer BJ, Brown DR, Michels KM (1991). Statistical principles in experimental design. 3rd edn. McGraw-Hill, Boston MA.