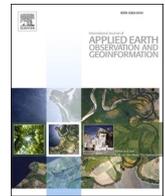




Contents lists available at ScienceDirect

# International Journal of Applied Earth Observations and Geoinformation

journal homepage: [www.elsevier.com/locate/jag](http://www.elsevier.com/locate/jag)

## Tree species classification using Sentinel-2 imagery and Bayesian inference

Arvid Axelsson<sup>a,\*</sup>, Eva Lindberg<sup>a</sup>, Heather Reese<sup>b</sup>, Håkan Olsson<sup>a</sup><sup>a</sup> Department of Forest Resource Management, Swedish University of Agricultural Sciences, 901 83 Umeå, Sweden<sup>b</sup> Department of Earth Sciences, University of Gothenburg, 405 30 Gothenburg, Sweden

### ARTICLE INFO

#### Keywords:

Multi-temporal  
Satellite  
Bayesian  
Land cover  
Classification  
Maximum likelihood  
Sequential

### ABSTRACT

The increased temporal frequency of optical satellite data acquisitions provides a data stream that has the potential to improve land cover mapping, including mapping of tree species. However, for large area operational mapping, partial cloud cover and different image extents can pose challenges. Therefore, methods are needed to assimilate new images in a straightforward way without requiring a total spatial coverage for each new image. This study shows that Bayesian inference applied sequentially has the potential to solve this problem. To test Bayesian inference for tree species classification in the boreo-nemoral zone of southern Sweden, field data from the study area of Remningstorp (58°27'18.35" N, 13°39'8.03" E) were used. By updating class likelihood with an increasing number of combined Sentinel-2 images, a higher and more stable cross-validated overall accuracy was achieved. Based on a Mahalanobis distance, 23 images were automatically chosen from the period of 2016 to 2018 (from 142 images total). An overall accuracy of 87% (a Cohen's kappa of 78.5%) was obtained for four tree species classes: *Betula spp.*, *Picea abies*, *Pinus sylvestris*, and *Quercus robur*. This application of Bayesian inference in a boreo-nemoral forest suggests that it is a practical way to provide a high and stable classification accuracy. The method could be applied where data are not always complete for all areas. Furthermore, the method requires less reference data than if all images were used for classification simultaneously.

### 1. Introduction

Tree species information is currently one of the key parameters of interest to ecologists and forest managers alike. The increased stream of freely available optical satellite data provides more opportunities to use a multitude of images, making it possible to map more thematically detailed classes with higher accuracy. As an example, the Sentinel-2 satellites from the European Space Agency's (ESA) Copernicus program provides images with a temporal frequency of five days at the equator, and higher frequency at higher latitudes (e.g., every one to three days over Sweden). The use of multi-temporal data for land cover classification has been shown to improve accuracy of tree species classification (Amani et al., 2017; Immitzer et al., 2019). Yet, processing a large number of images poses a new challenge in the process of land cover mapping: how can the stream of partially overlapping data best be used? Hence, there is a need for methods that provide highly accurate and thematically detailed (e.g., tree species) maps over large areas that can be easily updated as new images are acquired.

Earlier work on tree species classification from optical data benefited from using a few multi-temporal images where phenological differences

were used to separate the classes (Fassnacht et al., 2016; Hagner and Reese, 2007; Hill et al., 2010; Reese et al., 2002; Wolter et al., 1995). Much of the recent work using Sentinel-2 data have used a selection of scenes rather than the continuous data flow. Immitzer et al. (2019) classified tree species using 18 Sentinel-2 images (from 2015 to 2017) and a random forest classification. They found that six to seven dates of imagery provided as accurate a model as using all 18 images (an out-of-bag overall accuracy of 85.7% for 12 tree species), and that images from April to August contributed most to a higher accuracy. Persson et al. (2018) used four images, from April, May, July, and October of the same year in a random forest classification of five tree species. They achieved the highest overall accuracy of 88.2% when using all images and confirmed that the spring and autumn images, which had different phenological states, were the most useful. Puletti et al. (2018) used three Sentinel-2 images from spring, summer and autumn in a random forest classification to discriminate between coniferous, broad-leaved and mixed forest, achieving a maximum overall accuracy of 86.2% using a separate validation data set.

Research on determining vegetation phenology (including tree species) using a continuous flow of remote sensing data has been conducted

\* Corresponding author.

E-mail addresses: [arvid.axelsson@slu.se](mailto:arvid.axelsson@slu.se) (A. Axelsson), [eva.lindberg@slu.se](mailto:eva.lindberg@slu.se) (E. Lindberg), [heather.reese@gu.se](mailto:heather.reese@gu.se) (H. Reese), [hakan.olsson@slu.se](mailto:hakan.olsson@slu.se) (H. Olsson).

<https://doi.org/10.1016/j.jag.2021.102318>

Received 10 April 2020; Received in revised form 31 January 2021; Accepted 10 February 2021

Available online 2 April 2021

0303-2434/© 2021 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

at a site in Sweden by Jönsson et al. (2018). Their study evaluated a new method that would better accommodate data gaps in time series of Sentinel-2, Landsat and MODIS data, and compared the method to that used in TIMESAT (Jönsson and Eklundh, 2004), among others. They applied seasonal shape priors combined with box constrained separable least squares fit to logistic model functions, which was a robust approach when processing time series with sparse or missing data (e.g., due to cloud occlusions).

When making land cover classification over large areas with multi-temporal images, there are other issues apart from handling the large number of images. One such issue is the problem of occluded areas in the images, leading to missing data. One of the primary causes of this is clouds and cloud shadows. Another issue may be differences in images taken under different atmospheric conditions or with different illumination conditions (e.g., sun angle). Yet another issue is that each area may have been imaged a different number of times, resulting in an unequal number of observations at different times between sites. Therefore, one motivation behind the current study is to investigate a method that could utilise a data stream of images that are only partly overlapping.

As the amount of available data have increased, new approaches to efficiently and accurately classifying a continuous stream of optical data have been developed, including convolutional neural networks (Yuan et al., 2020), distribution-based thresholding (Nink et al., 2019), and support vector machines (Shelestov et al., 2017). In past decades, Maximum likelihood classification was the most widely used classification method for making thematic land cover maps from remote sensing data. One advantage of Maximum likelihood classification is that it can be used in a Bayesian framework where probabilities can be updated by using new observations. Strahler (1980) proposed that the method may make “use of time-sequential information in making the outcome of a later classification contingent on an earlier classification.” How this is done depends on the decision rule used (Swain, 1978). With the high temporal resolution of current earth observation missions, combined with the increase in computational power since 1980, this Bayesian method can be worth revisiting.

Similar to Strahler’s proposal (Strahler (1980)), Bayesian updating of land cover was used by Cardille and Fortin (2016) to solve the problem of partially overlapping data when classifying an area of central Quebec into water, forest and burnt area from a time series consisting of eleven Landsat 8 images. They used the method to ignore noise caused by clouds and smoke, and to avoid problems due to images covering only parts of the area. By using Bayesian updating, they achieved an overall accuracy of more than 90% at each time step, compared to a median overall accuracy of 78% when images were used separately. Crowley et al. (2019) used the same method with a combination of six Landsat 8 images, ten Sentinel-2 images, and six MODIS images to produce maps of burned and unburned areas in British Columbia.

The objective of this study was to investigate the utility of applying a Bayesian inference method to a continuous flow of Sentinel-2 data for tree species classification in the boreo-nemoral forest landscape of southern Sweden. The aim was to test the method of Bayesian inference by applying Maximum Likelihood classification to the posterior probabilities produced by the method, similar to that proposed by Strahler (1980), and using data with high temporal resolution. The focus was not on finding the optimal spectral bands or combination image dates for classification. This method could provide a relatively simple approach to achieve continuously updated and highly accurate large area tree species maps, using a continuous flow of data with different image date combinations and only partial overlap. In the case of Sweden, such tree species maps are of interest to the forest agency and the environmental protection agency, and at European level for example for keeping the Corine land cover data base up to date. By applying Bayesian inference and updated posterior probabilities to multiple images in a sequence, the expected outcome was to obtain a higher and more stable overall accuracy than when classifying single or few images.

## 2. Materials and methods

### 2.1. Study site and field data

Field data were collected during the summer of 2016 and the spring of 2017 in the study area of Remningstorp in the boreo-nemoral zone of southern Sweden (58°27′18.35″ N, 13°39′8.03″ E). In total, 335 circular plots with a radius of 10 meters were available (Fig. 1). Of these plots, 265 were placed in the north-west and central parts of the study area, in managed forest stands mainly dominated by conifers. These plot coordinates had been determined by an earlier inventory campaign using systematic sampling in a grid with random start. A new grid with random start was placed in the south-east part of the study area, that was dominated by broad-leaved forest, mainly in a nature reserve that was historically used for grazing.

During the field inventory of the south-eastern area, the grid points were used as a reference, but the plot centres were allowed to be shifted by up to 20 m to the nearest area containing a dominant representation of a single tree species. The tree species composition was assessed with an angle gauge at the grid point as well as positions 20 m north, east, south and west of the grid point. If one species constituted more than 70% of the basal area at one of the positions, that position was selected as the new centre coordinate for the field plot. In total 70 new field plots were established.

All plot centres were positioned using a real-time kinematic GPS with an accuracy of one metre. Stem diameter at 1.3 m above ground ( $d_{bh}$ ) and species was recorded for all trees where  $d_{bh}$  was at least 4 cm. Only plots where mean  $d_{bh}$  was at least 10 cm were included. Since angle gauge measurements of basal area might differ from the basal area within a circular plot (Gregoire and Valentine, 2008, chapter 8), only those plots where one species constituted 70% or more of the basal area calculated from  $d_{bh}$  were used. To assure enough training data for each species, only those that dominated more than 10 plots were kept. In total, 169 plots with four species classes remained (Table 1). These four tree species represent over 90% of the forest biomass in Sweden (Nilsson et al., 2020).

### 2.2. Remote sensing data

Satellite imagery from both Sentinel-2A and Sentinel-2B were used for classification. The mission provides data around every three days over the study area. Each image consists of thirteen spectral bands with varying ground sampling distance: 10 metres, 20 metres and 60 metres (European Space Agency, 2015). Images from the period of 2016-07-31 to 2018-08-30 from granule T33VVE were downloaded from Copernicus Open Access Hub. Only Level-1C images were available for all dates, therefore, we performed atmospheric correction to Level-2A using the Sen2Cor program, version 2.8 (European Space Agency, 2019), provided by ESA. Both processing levels were evaluated and due to the way in which the proposed classification method works, combined with a fairly small study area, no significant differences in the outcomes when using Level-1C or Level-2A were expected.

Due to the low sun angles and presence of snow in winter images, only images from the months of April to October of each year were used. In total, 142 images were obtained from this time period, and 22 images, or 15%, were free of clouds and cloud shadows in the study area.

Among the spectral bands available from different sensors, the most common ones used for vegetation mapping are green, red and near-infrared. Since the aim of this study was to evaluate the method of Bayesian inference on a data stream (i.e., rather than to determine optimal band combinations), it was deemed most important to have constant parameters for every image through the time series; therefore, the green, red, and near-infrared wavelength bands were used (bands 3, 4, and 8), which all have the same pixel size of 10 metres. The visible blue band was excluded, however, since it is very sensitive to variations in atmospheric effects, and therefore less useful for large-area

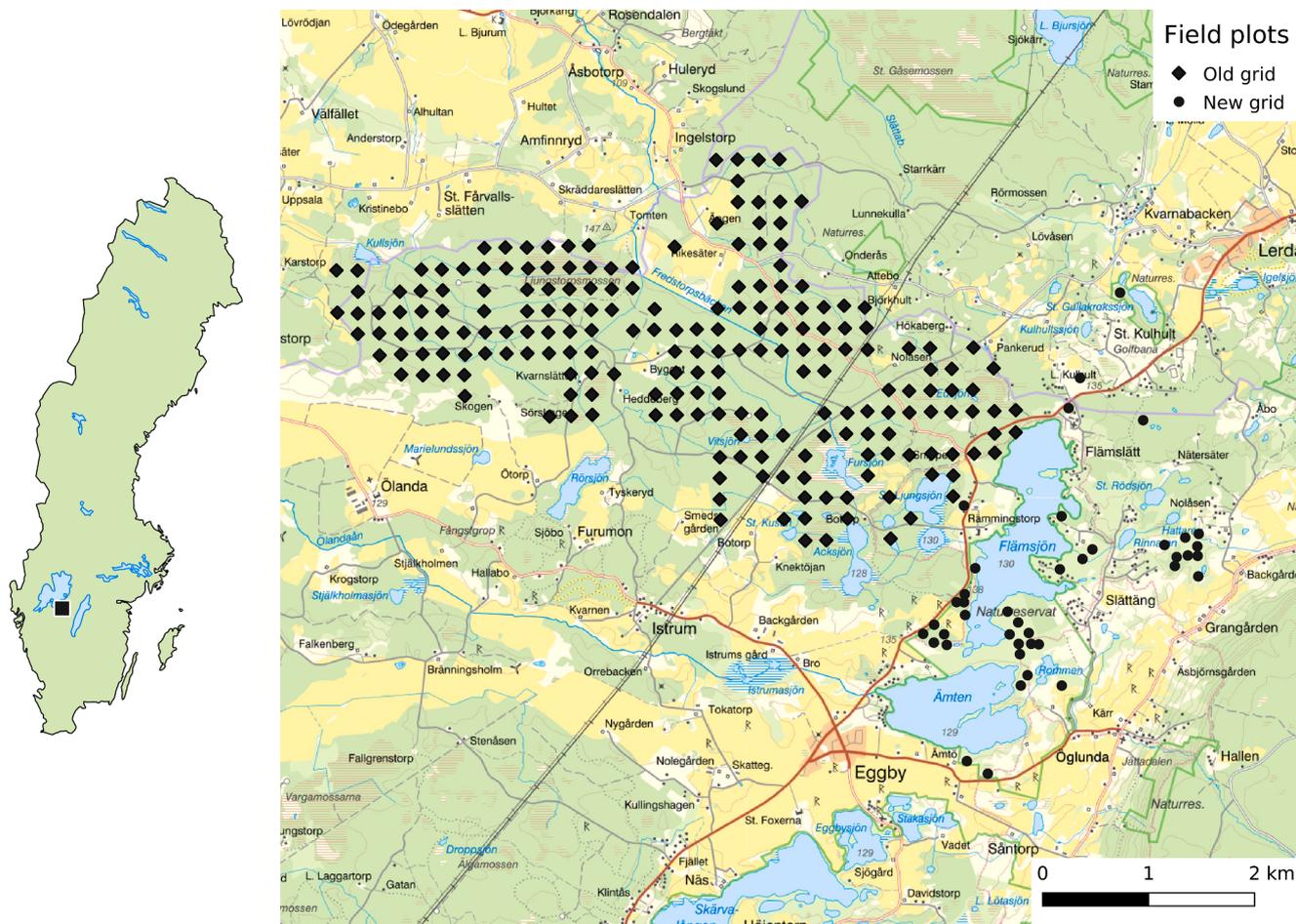


Fig. 1. Map of the Remningstorp study area and field plots therein. Background map is GSD - Terrain Map, © Lantmäteriet.

Table 1

Number of plots ( $n$ ) per species and their properties in field data. Diameter at 1.3 meter above ground is denoted  $d_{bh}$ , basal area weighted height is denoted  $h$ , and basal area is denoted  $q$ .

Species	$n$	$\bar{d}_{bh}$ (cm)	$SD(d_{bh})$ (cm)	$\bar{h}$ (m)	$SD(h)$ (m)	$\bar{q}$ (m <sup>2</sup> )	$SD(q)$ (m <sup>2</sup> )
<i>Betula spp.</i>	25	15	7	19	6	17	13
<i>P. abies</i>	99	20	9	19	6	25	13
<i>P. sylvestris</i>	26	23	8	21	5	26	7
<i>Q. robur</i>	19	21	11	21	3	24	10

operational mapping of tree species.

When extracting pixel values, field plot coordinates were transformed from SWEREF 99TM (EPSG 3006) to WGS84/UTM zone 33 N (EPSG 32633), which are almost identical in their placement for this study area. Spectral data for the field plot locations were extracted with bilinear interpolation.

### 2.3. Classification method

Maximum likelihood (ML) classification is made by using a decision rule based on an expression describing conditional probabilities of different classes given an observation (Swain, 1978, pp. 152–158). A decision rule for making an ML classification using multiple images was formulated, with the assumption that observations were conditionally independent given a class. The probability of interest was  $Pr(\omega_k | X_1 \cap X_2 \cap \dots \cap X_n)$ , where  $\omega_k$  is the event that a pixel has class  $k$ , and  $X_t$  is the event of observing the vector  $\vec{x}_t$  of band values in the pixel

of an image taken at time  $t$ . The selection rule, equivalent to the probability of interest, which was used for classification was: *Select  $k$  to maximize*

$$\ln(Pr(\omega_k)) + \sum_{t=1}^n \ln(p(X_t, \vec{\theta}_{k,t})), \tag{1}$$

where  $p$  is the probability density function,  $\vec{\theta}_{k,t}$  is the parameter vector of that function for species  $k$  and image  $t$ .

Bayesian classification is often made using a normal distribution, but can be used with any likelihood function (Gorte and Stein, 1998). Two alternative distributions were tested: normal distribution and t-distribution. The t-distribution has wider tails, possibly allowing for a more robust classification. Kernel methods for obtaining empirical distributions were not an option due to the small sample size for some of the tree species in this study.

### 2.4. Image selection

Images were assigned a grade according to the level of class separation within them. For each pair of classes, a Mahalanobis distance (Mahalanobis, 1936), hereafter denoted as  $Z$  or  $Z$ -value, of the difference between the class population means was calculated as

$$Z = (\widehat{\mu}_a - \widehat{\mu}_b)^T \left( \frac{\widehat{\Sigma}_a}{m_a} + \frac{\widehat{\Sigma}_b}{m_b} \right)^{-1} (\widehat{\mu}_a - \widehat{\mu}_b), \tag{2}$$

where  $\widehat{\mu}$  is sample mean vector,  $\widehat{\Sigma}$  is the sample variance–covariance matrix,  $m$  is sample size, and subscripts  $a$  and  $b$  denote the different

classes. To select images, they were put in decreasing order by Z-value and for each pairwise combination of classes, the highest ranking images were selected. The number of images to be selected was set as close as possible to fifteen percent of the total number of images ( $N = 142$ ), corresponding to the proportion of cloud-free images. There were a total of four thematic classes, giving six different class combinations. The value closest to 15% of 142 that is evenly divisible by six is 24. The total number of images selected per class combination by their Z-value was increased until the number of images selected for classification was no more than 24.

### 2.5. Accuracy evaluation

Classification model accuracy was evaluated using leave-one-out cross-validation. The main measure of accuracy was overall accuracy, that is defined as the number of correctly classed elements divided by the total number of elements.

To assess the impact on accuracy due to adding more images, combinations of a different number of images were evaluated. This was done both for all available images ( $N = 142$ ) as well as the subset selected by their Z-values ( $N \leq 24$ ). For each integer number  $i$  from one to the number of available images  $N$  minus one,  $N$  image combinations were sampled from a set of image combinations,  $S_i$ . The set  $S_i$  consisted of all possible combinations of  $i$  images. The overall accuracies when using each of those sampled combinations for classification were recorded and plotted to view the effect of using multiple images. When only one image was used for classification, it was the same as making an ordinary ML classification.

### 3. Results

When using images selected by their Z-value, no difference between the use of a normal distribution and a t-distribution was found. Only the results from calculations made using normal distributions are presented here. In addition, there was no difference when using Level-1C and Level-2A data for classification, and therefore Level-1C images were used for the study.

The images selected by their Z-values ranged from April to October, but most tended to be from May, June, and July, which is the beginning and middle of the summer in Sweden. To separate between *P. abies* and *P. sylvestris*, only images from the middle and end of summer were selected, and to separate *Q. robur* from other classes, mostly images from early summer were selected. *P. abies* and *P. sylvestris* were the two most spectrally similar classes, as shown by low Z-values for images selected for separation of that combination. The highest Z-values were obtained for class-pairs that included the *Q. robur* class (Table 2).

The overall accuracy for the classification was 87% when using the 23 images selected by their Z-values (Table 3). When all 142 images—both those with high and low Z-values—were used for classification, an overall accuracy of 85% was achieved. When using only a single image, the highest overall classification accuracy obtained was 83%, the average was 63%, and the lowest was 33%. When using all 142 images as well as when using the images selected by their Z-value, it could be seen that a maximum overall accuracy occurred when using only a subset of the images (Fig. 2). When sampling image combinations, both from the total of 142 images and the 23 images selected by Z-values, the average overall accuracy of the samples increased with increasing sample size, as expected. At the same time, variation in classification accuracy decreased (Fig. 2) (see Fig. 3).

The confusion matrix (Table 3) shows that it was primarily the two coniferous species that were conflated. *Q. robur* was only conflated with *Betula spp.* while *Betula spp.* itself was conflated at least one time with each of the other species.

**Table 2**

Images selected by their Z-value (a Mahalanobis distance, calculated using Eq. (2)). B stands for *Betula spp.* (birch), S stands for *Picea abies* (spruce), P stands for *Pinus sylvestris* (pine), and O stands for *Quercus robur* (oak). Bold face means that the image in question was among the best ones for discriminating between those two classes.

Cloud free	Date	B vs. S	B vs. P	B vs. O	S vs. P	S vs. O	P vs. O
Yes	2017-04-07	219	<b>114</b>	36	16	<b>508</b>	242
Yes	2018-04-10	122	87	<b>48</b>	44	<b>518</b>	<b>384</b>
Yes	2018-04-12	147	111	32	46	<b>441</b>	<b>323</b>
Yes	2018-05-07	119	74	33	28	404	<b>323</b>
Yes	2018-05-15	<b>255</b>	93	<b>135</b>	61	<b>443</b>	253
No	2018-05-20	<b>228</b>	105	<b>154</b>	54	338	<b>341</b>
Yes	2018-05-22	181	110	<b>115</b>	25	<b>429</b>	<b>376</b>
Yes	2017-05-27	<b>238</b>	92	<b>128</b>	61	291	200
Yes	2018-05-30	<b>238</b>	<b>133</b>	23	46	293	312
Yes	2018-06-01	162	92	22	41	<b>448</b>	<b>356</b>
Yes	2018-06-04	225	<b>112</b>	18	42	<b>459</b>	273
No	2017-06-06	189	<b>122</b>	<b>134</b>	13	410	<b>345</b>
Yes	2018-06-24	221	97	7	<b>107</b>	335	186
Yes	2018-06-26	<b>225</b>	<b>118</b>	8	59	263	188
No	2018-06-29	<b>253</b>	104	7	<b>91</b>	241	158
Yes	2018-07-04	185	71	9	<b>120</b>	383	190
No	2017-07-09	<b>241</b>	<b>123</b>	27	34	422	269
Yes	2018-07-14	205	70	18	<b>109</b>	403	177
Yes	2018-07-16	190	88	16	<b>98</b>	192	165
No	2018-07-21	147	59	11	<b>102</b>	264	130
No	2018-07-24	167	62	15	<b>120</b>	328	148
No	2017-08-30	184	<b>122</b>	13	6	214	141
Yes	2016-09-12	102	44	<b>79</b>	74	209	202

**Table 3**

Confusion matrix when using 23 images chosen by Z-value in Table 2. UA is Users Accuracy, PA is Producers Accuracy, and the entry in the lower right corner is Overall Accuracy. Cohen's kappa for this confusion matrix is 78.5%. Columns are field-inventoried classes and rows are the classification result.

	B. spp	P. abies	P. sylvestris	Q. robur	UA (%)
B. spp	20	1	2	3	76.9
P. abies	1	91	4	0	94.8
P. sylvestris	1	7	20	0	71.4
Q. robur	3	0	0	16	84.2
PA (%)	80.0	91.9	76.9	84.2	<b>87.0</b>

### 4. Discussion

The use of Bayesian inference applied sequentially on multiple Sentinel-2 images resulted in a higher overall classification accuracy of

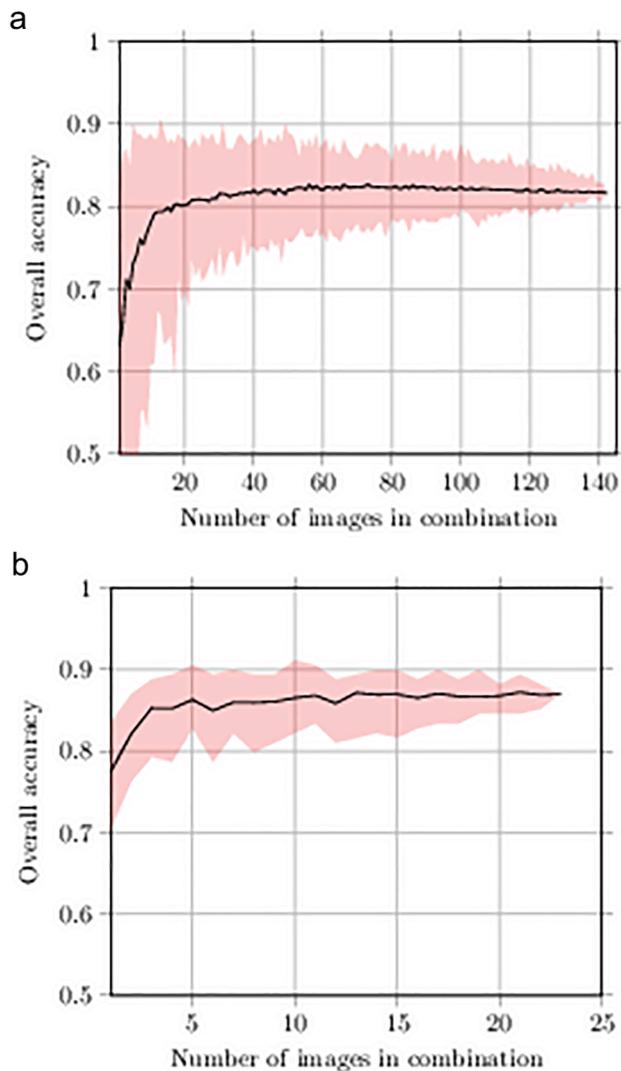


Fig. 2. Overall accuracy over number of images used for classification. The upper and lower limits of the red areas are the maximum and minimum overall accuracy in the respective samples. The black lines show the average overall accuracy.

tree species in a boreo-nemoral forest. In addition, the variation of overall accuracy between samples was smaller as compared to the result when Bayesian classification was applied to a single date image. Even though the method presented in this article does not utilize the temporal relationship between images, it does improve the classification accuracy by using information from multiple observations. Similar to other studies (e.g., Hill et al., 2010; Persson et al., 2018), a higher accuracy was obtained when using multiple image dates as opposed to a single date image. A big advantage of the proposed method is that it does not require that all areas are covered with the same sequence of image data, which allows, for example, cloudy areas to be handled in an efficient way.

There are many classification algorithms available for use with remote sensing data, however, there are certain advantages to Bayesian inference. For example, in large area mapping, each cell in a fixed grid over the area could be assigned the probability density values of different tree species, and updated with new probabilities and class membership as each new image became available. In this study, images were selected based on a quality measure derived for the entire image (i. e., Z-values). Sentinel-2 data products include pixel-wise quality measurements which are routinely used to flag pixels with poor quality. This

could be utilized to select the pixels to include or exclude when making a Bayesian inference.

In addition, when using Bayesian inference on individual images in a sequential manner, the curse of dimensionality can be avoided; if bands from every image in the time series were added as a new dimension of the data set, this would cause observations to be too sparse and overfitting would be a concern. Otherwise, to utilize a stream of images simultaneously, reduction in dimensionality would be needed each time an image was added to the model. With Bayesian inference where probabilities can be updated sequentially, this is not a problem as long as the assumption of conditional independence holds. When processing single images sequentially as opposed to combining many images, less training data would be required for this Bayesian approach.

Selecting images by their Z-value improved the overall classification accuracy, but more importantly, the overall accuracy initially increased by a relatively large percentage when adding more images that had relatively high Z-values. This supports the practice of selecting images that are cloud-free and of good radiometric quality (Kempeneers and Soille, 2017). When using Z-values to select images, mainly cloud-free images were identified, even though the emphasis of the method lies in class separation. The main advantage of using Z-values is that it is easily implemented and provides an automated approach to find a subset of optimal images from the data stream for the classification process. When all images were used, as shown in Fig. 2(a), the increase in accuracy started to flatten out when using between 50 and 100 images. When this result is compared to the overall accuracy obtained using images chosen by their Z-value, it can be seen in Fig. 2(b) that the variation in overall accuracy quickly decreased with increased sample size. Most importantly, the average classification accuracy was higher when using images with higher Z-values, as compared to using all images.

There were high Z-values for different tree species at different times of the year, supporting the use of multi-date images for catching phenological differences. *Q. robur* and *Betula spp.* green-up at different times in the spring, which is shown by the high separability between *Q. robur* and *Betula spp.* when using spring images, which for this study was primarily mid to late May images (Table 2). *P. abies* and *P. sylvestris* had the highest separation during late June and all through July. The reason for this might be that *P. sylvestris* trees have a sparser crown structure in comparison to *P. abies* trees, which leads to a sparser canopy in *P. sylvestris* stands as compared to *P. abies* stands. For *P. sylvestris* stands, this allows more light to penetrate the canopy. In the study area, *P. sylvestris* stands often have an undergrowth of *Corylus avellana* (hazel), resulting in some pixels with a mixed spectral response of *P. sylvestris* and *C. avellana*; this can result in difficulties in obtaining a pure *P. sylvestris* signature and lead to confusion in the classification with other spectrally overlapping classes. For the results from this paper, it should be kept in mind that a drought occurred in Sweden during the summer of 2018, potentially affecting the forest and other vegetation.

As for all parametric methods, some assumptions of data distributions need to be made. Two assumptions were tested: (1) that data were normally distributed, and (2) that all species followed a t-distribution with as many degrees of freedom as there were plots in the training data. The assumption of a t-distribution was selected since it allowed for a higher probability density for values far from the mean, especially in classes with few field plots. Both assumptions produced the same overall accuracy, which means that the variation in data is greater than the difference between the two distributions. An example where both of these assumptions would not be met would be during the leafing out of deciduous trees; a part of the population could be leafed-out while another could still be leafless. This would result in a bimodal distribution with the two subgroups appearing as local maxima. If more field data were available, it might be feasible to construct empirical distributions using kernel methods.

Sentinel-2 has some issues with the geometric quality of images. The geometric correction of Sentinel-2 imagery has proven to be of varying

## Classification

## False color orthophoto

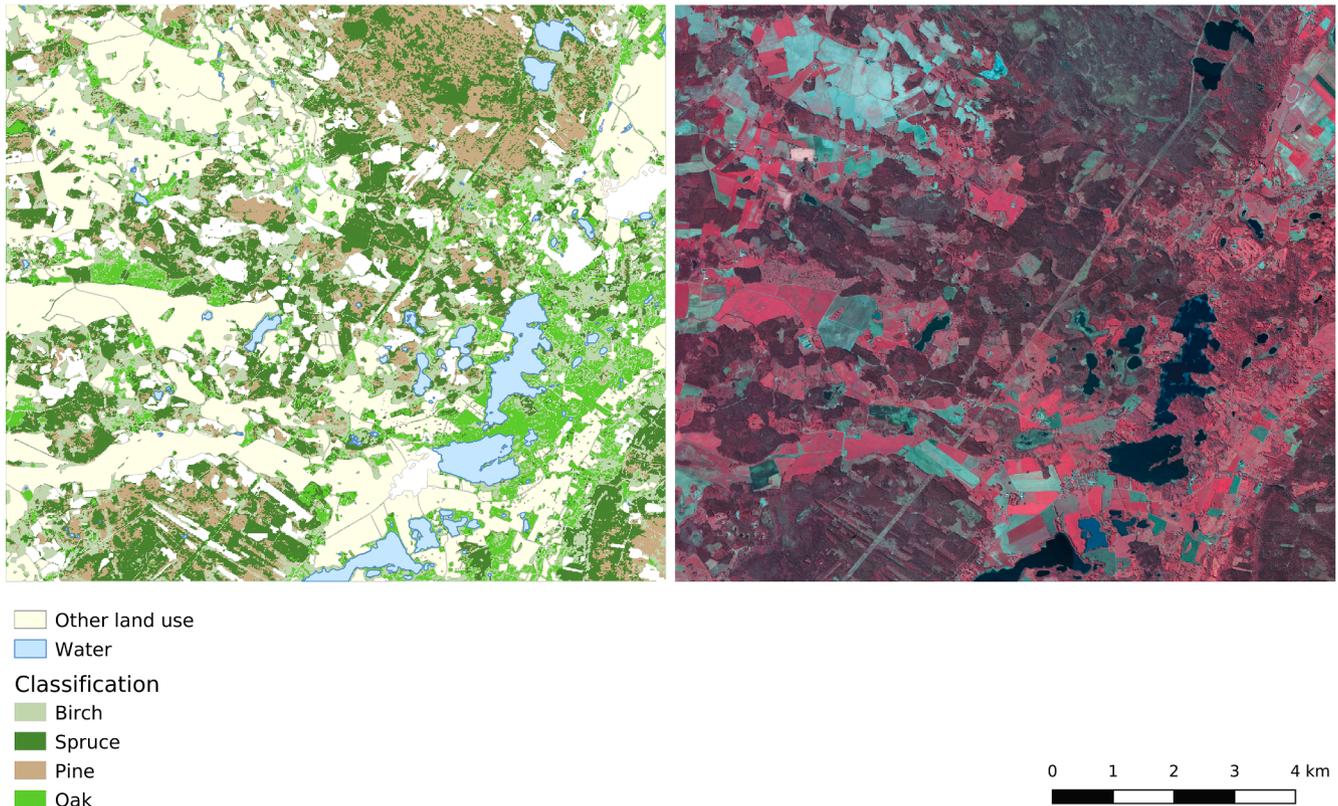


Fig. 3. Classification results when using the 23 images chosen by Z-value. An existing land cover map was used here for non-forest land cover. Right subfigure is GSD-Orthophoto, 0.5m, color-IR © Lantmäteriet.

quality, with geo-location inaccuracies in Level-1C data of up to around 10 m on unrefined images and 8 m on refined images (Gascon et al., 2017). Such shifts in the pixel geometry, and therefore occasional poor geographic co-location between images, was also seen in this study. The classification results could be expected to improve with better co-location of pixels, but the general results from this study should be valid even without a perfect geometry.

An image-level linear calibration from digital numbers to radiance or reflectance will not influence the classification result since each image is classified with the aid of ground reference plots. However, for large-area applications, any variation in atmospheric optical thickness within an image might be an error source. For this reason, and to obtain a good signal to noise ratio, it is important to select images that are as cloud and haze free as possible and only use the part of the images with a good image quality, which is possible with the presented method. In the case of large-area applications, the use of data at processing level 2A (with Bottom of Atmosphere correction) might provide a better basis for tree species classification, since the effects of non-linear atmospheric calibration could be greater.

Additional modeling could be used with Bayesian inference. A Hidden Markov Model (HMM) could provide a way to incorporate a model for phenological change to improve classification accuracy (Siachalou et al., 2015). To do this, a model of probabilities for phenological change as a function of time would be needed. To do this in the best way, field-based information on vegetation phenology over the study area should be used to formulate the transition matrices needed for HMM; however, this type of field data is not currently collected in a comprehensive and available manner for all of Sweden.

All images were treated as equally important for classification

purposes, regardless of how close they were in time to the field inventory. One idea for future development would be to smooth out the distributions with a weighting according to the difference in time from the inventory data. This would mean that an observation taken at a date close to the image acquisition date would be trusted more than one further away in time.

To further examine the usefulness of Bayesian inference as a classification method of tree species, a study on a greater landscape level would be of interest. If data from a national forest inventory (NFI) were to be used, it might be advisable to use whole swaths of satellite data, as opposed to individual  $100 \times 100$  km granules, to get better estimations of distributions. Since NFI data are often acquired on a regular basis, weighting observations by temporal difference might be useful.

The method of Bayesian inference, where pixel-wise probabilities can be calculated and then updated as more observations are made, provides a way to easily automate classification of a data stream consisting of multi-temporal images. The results in this study might provide a foundation for a practical method for continuous large area forest mapping where new images are screened for image quality using a Z-value, and high quality images are used for improving the previous classification. The method looks promising and should be tested in more forest ecosystems.

## 5. Conclusions

This study demonstrated the utility of classical Bayesian inference applied sequentially, where accumulated class likelihoods were used as prior probabilities from an image stream. The study used an initial stream of 142 Sentinel-2 images to map four tree species over a study

area in southern Sweden. Using Z-scores, a best subset of 23 images was identified. The overall classification accuracy of the 23 images was 87%, while for the 142 images it was 85% and for the single best image it was 83%. In addition, the variation of overall accuracy between samples was smaller when Bayesian inference applied sequentially was used. The method is a straight forward way to handle cases where images have different geographical extents or contain occlusions (e.g., due to cloud cover). The method thus has the potential to be an efficient way for utilizing a stream of multi-temporal images for large area tree species or land cover classification.

### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgements

This research was supported by SNSA, the Swedish National Space Agency, through the project "Iterative land cover classification using assimilation of multiple Sentinel-2 images with Hidden Markov Models" with the SNSA DNR number 136/15, by Formas, the Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning, through the project "Tree species classification from remote sensing - next generation forest maps for ecology and management" with the Formas DNR number 2018-01161, by the Swedish Forest Society Foundation through the project "New forest maps with tree species information from time series of satellite data" with the Id-nummmmer: 2019-660-Stege 2 2018, by Mistra, the Swedish Foundation for Strategic Environmental Research, through the research program "Mistra Digital Forest", by the Hildur and Sven Wingquist Foundation for Forest Science Research, and by the Swedish University of Agricultural Sciences

### References

Amani, M., Salehi, B., Mahdavi, S., Granger, J., Brisco, B., 2017. Evaluation of multi-temporal landsat 8 data for wetland classification in newfoundland, Canada. In: 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 62292153–62317003. <https://doi.org/10.1109/IGARSS.2017.8128432>. ISSN: 2153-7003.

Cardille, J.A., Fortin, J.A., 2016. Bayesian updating of land-cover estimates in a data-rich environment. *Remote Sens. Environ.* 186, 234–249. <https://doi.org/10.1016/j.rse.2016.08.021>.

Crowley, M.A., Cardille, J.A., White, J.C., Wulder, M.A., 2019. Multi-sensor, multi-scale, Bayesian data synthesis for mapping within-year wildfire progression. *Remote Sens. Lett.* 10, 302–311. <https://doi.org/10.1080/2150704X.2018.1536300>.

European Space Agency, 2015. Sentinel-2 User Handbook. Technical Report 1. [https://sentinel.esa.int/documents/247904/685211/Sentinel-2\\_User\\_Handbook](https://sentinel.esa.int/documents/247904/685211/Sentinel-2_User_Handbook).

European Space Agency, 2019. S2 MPC – Sen2Cor Configuration and User Manual. Technical Report 2. <http://step.esa.int/thirdparties/sen2cor/2.8.0/docs/S2-PDGS-MPC-L2A-SUM-V2.8.pdf>.

Fassnacht, F.E., Latifi, H., Sterenczak, K., Modzelewska, A., Lefsky, M., Waser, L.T., Straub, C., Ghosh, A., 2016. Review of studies on tree species classification from remotely sensed data. *Remote Sens. Environ.* 186, 64–87. <https://doi.org/10.1016/j.rse.2016.08.013>.

Gascon, F., Bouzinac, C., Thépaut, O., Jung, M., Francesconi, B., Louis, J., Lonjou, V., Lafrance, B., Massera, S., Gaudel-Vacaresse, A., Languille, F., Alhammoud, B., Viallefont, F., Pflug, B., Bieniarz, J., Clerc, S., Pessiot, L., Trémas, T., Cadau, E., De Bonis, R., Isola, C., Martimort, P., Fernandez, V., 2017. Copernicus Sentinel-2a Calibration and Products Validation Status. *Remote Sens.* 9, 584. <https://doi.org/10.3390/rs9060584>.

Gorte, B., Stein, A., 1998. Bayesian classification and class area estimation of satellite images using stratification. *IEEE Transactions on Geoscience and Remote Sensing* 36, 803–812. <https://doi.org/10.1109/36.673673>.

Gregoire, T.G., Valentine, H.T., 2008. Sampling strategies for natural resources and the environment. *Applied environmental statistics*. Chapman & Hall/CRC, Boca Raton.

Hagner, O., Reese, H., 2007. A method for calibrated maximum likelihood classification of forest types. *Remote Sens. Environ.* 110, 438–444. <https://doi.org/10.1016/j.rse.2006.08.017>.

Hill, R.A., Wilson, A.K., George, M., Hinsley, S.A., 2010. Mapping tree species in temperate deciduous woodland using time-series multi-spectral data. *Appl. Veg. Sci.* 13, 86–99. <https://doi.org/10.1111/j.1654-109X.2009.01053.x>.

Immitzer, M., Neuwirth, M., Böck, S., Brenner, H., Vuolo, F., Atzberger, C., 2019. Optimal Input Features for Tree Species Classification in Central Europe Based on Multi-Temporal Sentinel-2 Data. *Remote Sens.* 11, 2599. <https://doi.org/10.3390/rs11222599>.

Jönsson, P., Cai, Z., Melaas, E., Friedl, M.A., Eklundh, L., 2018. A method for robust estimation of vegetation seasonality from landsat and sentinel-2 time series data. *Remote Sens.* 10, 635.

Jönsson, P., Eklundh, L., 2004. Timesat—a program for analyzing time-series of satellite sensor data. *Comput. Geosci.* 30, 833–845.

Kempeneers, P., Soille, P., 2017. Optimizing Sentinel-2 image selection in a Big Data context. *Big Earth Data* 1, 145–158. <https://doi.org/10.1080/20964471.2017.1407489>.

Mahalanobis, P.C., 1936. On the generalized distance in statistics. In: Proceedings of the National Institute of Sciences of India, National Institute of Science of India. pp. 49–55.

Nilsson, P., Roberge, C., Fridman, J., 2020. Skogsdata 2020. Technical Report. Swedish University of Agricultural Sciences.

Nink, S., Hill, J., Stoffels, J., Buddenbaum, H., Frantz, D., Langshausen, J., 2019. Using Landsat and Sentinel-2 Data for the Generation of Continuously Updated Forest Type Information Layers in a Cross-Border Region. *Remote Sens.* 11, 2337. <https://doi.org/10.3390/rs11202337>.

Persson, M., Lindberg, E., Reese, H., 2018. Tree Species Classification with Multi-Temporal Sentinel-2 Data. *Remote Sens.* 10, 1794. <https://doi.org/10.3390/rs10111794>.

Puletti, N., Chianucci, F., Castaldi, C., 2018. Use of Sentinel-2 for forest classification in Mediterranean environments. *Annals Silvicult. Res.* 42 <https://doi.org/10.12899/asr-1463>.

Reese, H.M., Lillesand, T.M., Nagel, D.E., Stewart, J.S., Goldmann, R.A., Simmons, T.E., Chipman, J.W., Tessar, P.A., 2002. Statewide land cover derived from multiseasonal Landsat TM data: A retrospective of the WISCLAND project. *Remote Sens. Environ.* 82, 224–237. [https://doi.org/10.1016/S0034-4257\(02\)00039-1](https://doi.org/10.1016/S0034-4257(02)00039-1).

Shelestov, A., Lavreniuk, M., Kussul, N., Novikov, A., Skakun, S., 2017. Exploring Google Earth Engine Platform for Big Data Processing: Classification of Multi-Temporal Satellite Imagery for Crop Mapping. In: *Frontiers in Earth Science* 5. <https://doi.org/10.3389/feart.2017.00017>.

Siachalou, S., Mallinis, G., Tsakiri-Strati, M., 2015. A Hidden Markov Models Approach for Crop Classification: Linking Crop Phenology to Time Series of Multi-Sensor Remote Sensing Data. *Remote Sens.* 7, 3633–3650. <https://doi.org/10.3390/rs70403633>.

Strahler, A.H., 1980. The use of prior probabilities in maximum likelihood classification of remotely sensed data. *Remote Sens. Environ.* 10, 135–163. [https://doi.org/10.1016/0034-4257\(80\)90011-5](https://doi.org/10.1016/0034-4257(80)90011-5).

Swain, P.H., 1978. Fundamentals of Pattern Recognition in Remote Sensing. In: *Remote Sensing; The Quantitative approach*. McGraw-Hill International Book Company, pp. 136–187.

Wolter, P.T., Mladenoff, D.J., Host, G.E., Crow, T.R., 1995. Using Multi-Temporal Landsat Imagery. *Photogram. Eng. Remote Sens.* 61, 1129–1143.

Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J., Gao, J., Zhang, L., 2020. Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sens. Environ.* 241, 111716. <https://doi.org/10.1016/j.rse.2020.111716>.