Tree species classification in Norway from airborne hyperspectral and airborne laser scanning data

Øivind Due Trier*, Aart-Borre Salberg*, Martin Kermit*, Øystein Rudjord*, Terje Gobakkenb, Erik Næssetb and Dagrun Aarstenc

*Section for Earth Observation, Norwegian Computing Center, Oslo, Norway; bFaculty of Environmental Science and Natural Resource Management, Norwegian University of Life Sciences, Ås, Norway; cTerraTec AS, Oslo, Norway

ABSTRACT
This article compares four new automatic methods to discriminate between spruce, pine and birch, which are the dominating tree species in Norwegian forests. Airborne laser scanning and hyperspectral data were used. The laser scanning data was used to mask pixels with low or no vegetation in the hyperspectral data. A green–blue ratio was used to remove shadow areas from tree canopies, and the normalized difference vegetation index to remove dead vegetation and non-vegetation. The best method was hyperspectral pixel classification with 160 spectral channels in the visible and near-infrared spectrum, using a deep neural network. This method achieved 87% correct classification rate. Partial least squares regression for hyperspectral pixel classification achieved 78%. Deep neural network image classification using canopy height blended with three hyperspectral channels achieved 74%. A simple pixel classification method based on two spectral indices resulted in 67% correct classification. A possible future improvement is to find a better way to combine hyperspectral data with canopy height data in a deep neural network.

Introduction

The current situation in the forest industry in Norway is under pressure due to fluctuating timber prices and high labour cost. The primary focus of current forest mapping is timber for production of building material. Current forest mapping methods combine airborne laser scanning (ALS) data with species classification from manual interpretation of multispectral stereo imagery (broad visual and near-infrared channels). The manual photo interpretation is expensive and inaccurate.

In this perspective, more automated methods for forest mapping are needed to reduce costs and improve the accuracy. Specifically, we will focus on methods which combine data from ALS and imaging spectrometer. The ALS data already provide information on vegetation height. The hyperspectral data may provide information on biophysical and biochemical parameters, and species composition. Combined, the two types of data have the potential of more accurate forest mapping. Correct species classification is important, since the wood from spruce, pine, birch and other deciduous trees are priced differently in the market.

Goetz (2009) reviews the technological development of hyper-spectral remote sensing from the 1970s. The airborne visible/infrared imaging spectrometer was a major technological advancement in 1987 and is still an important provider of hyperspectral data. Other essential developments include field spectrometers, e.g. by Analytical Spectral Devices Inc., and software for hyperspectral data analysis, e.g. ENVI. Gao, Montes, Davis, and Goetz (2009) review atmospheric correction algorithms for hyper-spectral data. Atmospheric correction is needed to convert at-sensor radiances to surface reflectance, so that data collected at different times and/or locations may be compared. However, Korpela, Mehtätalo, Seppänen, and Kangas (2014) found that for a small study with hyperspectral data collection within 2 h, classification performance did not improve with atmospheric correction. Plaza et al. (2009) review hyper-spectral image processing techniques. One important aspect is integration of spatial and spectral information, as this is complementary data. Also, the high dimensionality of the data requires efficient processing methods.

Schaepman et al. (2009) observe that most studies on forests have focused on young to mature forests, rather than more structurally complex old-growth forests. This may be due to smaller degrees of uncertainty. Hence, young to mature forests are more suited to model parameterization and validation. Another observation is that there are two causes for
mixed pixels. One is that the pixel size is too large, so that several disjoint materials are captured within a pixel. The other is that two materials may be combined into a homogeneous, intimate mixture, which cannot be resolved by increasing the resolution.

Ustin et al. (2009) review spectral wavelengths and combinations of wavelengths for remote sensing of plant pigments, and to detect plant stress. Kokaly, Asner, Ollinger, Martin, and Wessman (2009) review hyper-spectral remote sensing of non-pigment substances in vegetation, including water, nitrogen, cellulose and lignin. One application is to quantify net primary production over large geographical areas.

Fassnacht, Latifí, and Koch (2012) compared angular vegetation indexes with traditional vegetation indexes for forest damage detection in the Bavarian National Park from 7 m resolution hyper-spectral data. The best overall accuracy was 91%, using six classes and a combination of angular vegetation indexes. The best combination of traditional vegetation indexes gave 84% overall accuracy.

Clark, Roberts, and Clark (2005) used hyper-spectral data for species classification in a tropical rain forest with seven different species. For laboratory spectrometer measurements of individual leaves, 100% correct classification was obtained. For airborne hyper-spectral images with 1.6 m pixel size, the best results (92%) were obtained using average pixel spectra within manually delineated crowns, clearly better than pixel-level classification (81%), and also better than majority voting of classified pixels within each tree crown (84%). By only using sunlit pixels, species classification of individual pixels improved (87%), and also improved using majority voting (86%), but species classification of average spectra within crowns did not improve (91%) by keeping only the sunlit pixels.

Dalponte, Ørka, Gobakken, Gianelle, and Næsset (2013) studied species classification in a Boreal forest from two different hyper-spectral sensors: one visible and near-infrared (VNIR) sensor with 0.4 m pixel size at 1500 m flying height, and one short-wave infrared (SWIR) sensor with 1.5 m pixel size at the same flying height. The study area, located in Aurskog-Høland municipality, Akershus County in South-East Norway, is dominated by Spruce (35%) and Pine (50%), but also has some Birch (11%) and other broadleaved species (4%). The best result (93%) was obtained with the VNIR sensor on manually delineated tree crowns, and by using a support vector machine (SVM) classifier, a slightly better result than pixel-level classification on 0.4 m pixels with the same sensor and classifier (90%). Using the two sensors together had no positive effect. The authors see a huge potential in combining the hyper-spectral VNIR sensor with simultaneous acquisition of ALS. Manual delineation of tree crowns needs to be replaced by an automatic method in a practical application. Dalponte, Ørka, Ene, Gobakken, and Næsset (2014) has recently suggested a method for doing so, but it was only able to detect 48% of the tree crowns.

Dalponte, Bruzzone, and Gianelle (2008) combined hyper-spectral and ALS data for forest classification of a nature reserve in Italy, containing 19 different tree species. The ALS data provided better separability between tree species with similar spectral signatures but different average tree height. A SVM classifier performed better (89%) than Gaussian maximum likelihood and k-nearest neighbour.

Koetz, Morsdorf, van den Linden, Curt, and Allgöwer (2008) combined hyper-spectral and ALS data for land cover classification for forest fire management. They observed that for a few land cover classes, ALS data alone was equally good as the combination of hyper-spectral and ALS data for classification, whereas for several other land cover classes, hyper-spectral alone was equally good as the combination. For some land cover classes, classification using only ALS data failed completely, whereas for hyper-spectral data alone, this did not happen. Overall, the best classification rates (75%) were obtained with the combined data.

Jones, Coops, and Sharma (2010) studied the combination of hyper-spectral and ALS data for tree species classification in a national park in south-western Canada. 11 species were mapped. By combining hyper-spectral and ALS data, the overall classification accuracy (73%) was only slightly better compared to using hyper-spectral data alone (72%), but there was a substantial increase for some species. First, the hyperspectral data was atmospherically corrected using FLAASH MODTRAN4. Then, water absorption regions, from 1350 to 1416 nm and from 1796 to 1970 nm, were removed. Also, regions from 395 to 429 nm and from 2401 to 2500 nm were removed due to high levels of noise. At this stage, 453 channels were still remaining. Then, a laboratory experiment with leaf samples was conducted to determine the 40 bands that minimized the within-species variance and at the same time maximized the between-species variance.

Clark, Roberts, Ewel, and Clark (2011) combined ALS and hyper-spectral data for above-ground biomass estimation in a tropical rain forest. The two data sets were used to compute features at the plot level. The hyper-spectral data was used for spectral un-mixing and for narrowband indices that were used to estimate photosynthetic vegetation, structure, senescence, health, water content and lignin content. These features were less successful in predicting above-ground biomass than vegetation height features derived from the ALS data. The height features included mean, 95th percentile, median, maximum, standard deviation, kurtosis, skewness and the ratio of median to maximum. Also, gap percentage was used.
Swatantran, Dubayah, Roberts, Hofton, and Blair (2011) also combined ALS and hyper-spectral data for biomass estimation. They found that ALS data was better than hyper-spectral for average biomass estimation, but that hyper-spectral data could provide species classification and thus could be used to adjust the biomass estimates. Also, hyper-spectral data provides canopy state information, such as stress.

By comparing the above studies (Table 1), we observe that the best correct classification rate in each study varies from 73% to 93%. Direct comparison of classification rates may not be fair, since the studies have different data and different number of species. The two studies on species classification that investigate the effect of including ALS features find that this effect is small with respect to the overall classification accuracy (Dalponte et al., 2008; Jones et al., 2010). However, the number of ALS features is small. The inclusion of ALS features results in higher increase in accuracy for land cover classification (Koetz et al., 2008). On the other hand, ALS features are important in biomass estimation (Clark et al., 2011; Swatantran et al., 2011). Two studies find that using average pixel spectra within individual tree crowns improves classification accuracy compared with pixel classification (Clark et al., 2005; Dalponte et al., 2013). The latter study also finds that majority voting of pixel classification within individual tree crowns give a smaller improvement in classification accuracy.

Fassnacht et al. (2016) have reviewed a large number of studies on tree species classification from remotely sensed data. They observe that early studies tended to use supervised maximum likelihood classifiers or unsupervised clustering methods. After 1995, decision-tree classifiers and neural network classifiers emerged. Some recent studies have preferred random forest or support vector machine classifiers.

Deep neural networks have recently emerged as the preferred methodology in image classification (Krizhevsky, Sutskever, & Hinton, 2012; LeCun, Bengio, & Hinton, 2015), and may also be used for hyperspectral pixel classification (Hu, Huang, Wei, Zhang, & Li, 2015). Mørchhale, Pauca, Plemmons, and Torgersen (2016) extended the deep neural network of Hu et al. (2015) to include elevation data from ALS for pixel-based land cover classification. The main strength of deep neural networks is that they may be able to extract combinations of the input data that are difficult to describe explicitly by humans. Species classification from remote sensing data is challenging, as the traits that are used by humans during fieldwork, e.g. leaf shapes and bark texture, are not available in remote sensing data. Instead, hyperspectral signatures and/or tree canopy shapes are being used. These may be quite similar for some species.

In this article, we compare two deep neural network methods with two alternative methods for tree species classification of airborne hyperspectral and laser scanning data of a boreal forest in Norway.

**Data**

ALS data (Table 2) for Gran, Lunner and Jevnaker municipalities were acquired on several dates in the period 2 May–20 June 2015. The ALS data have been processed so that each three-dimensional point is labelled as either “ground” or “other”, the latter including vegetation. A number of other flags are also attached to each point, including whether it is a first return or not.

Hyperspectral data (Table 3) were acquired on 6 September 2015 for a few selected locations within the ALS coverage. Two sensors were used simultaneously, one for the VNIR wavelengths and one for the SWIR wavelengths of reflected solar illumination. The SWIR data were not used in the experiments. The data was orthorectified and geocoded using PARGE (http://www.rese.ch/products/parge/) and appears as raw radiances in the range 0.0–0.13. No atmospheric correction was used.

The forest area has a mixed species composition that is typical of many Norwegian forests, with approximately 50% Norway spruce, 40% Scots pine, 7% birch and 3% other deciduous species. The less productive areas are dominated by pine, whereas the more productive areas are dominated by spruce.

The spruce forest is mainly planted, and within each forest stand of spruce, most trees are of the same age. The forest management includes one or two iterations of thinning and a final clear cutting. On clear cuts, some isolated pine and birch trees had been left. Also within pine stands, most of the trees were of the same age, due to the forest management. Single birch trees occurred within some spruce stands. Some forest stands had a mixed species composition.

Manual photointerpretation experts have experienced that the spectral variability within spruce and pine has an age component. Young spruce trees tend to be brighter than mature spruce trees, indicating that young spruce trees may be spectrally quite close to birch. Mature pine trees tend to have less saturated colours than young pine trees. Young pine trees may be spectrally quite close to mature spruce trees.

Field work was conducted on 2 June, 5 September and 14 September 2016 to identify areas with a single tree species. That is, for each polygon, the border was drawn so that only trees from the same species were included. To prepare for the field work, paper maps were printed in scale 1:1000 (Figure 1), containing
| Authors             | Data | # HS channels | Wavelengths | Resolution | # classes | Prediction task | Unit | # HS Features | # ALS features | Method                                           | Accuracy |
|---------------------|------|---------------|-------------|------------|-----------|----------------|------|---------------|---------------|-------------------------------------------------|----------|
| Fassnacht et al. (2012) | HS   | 125           | 450–2480 nm | 7 m        | 6         | Damage classes | Pixel | 8             | N/A           | Angular vegetation indexes + support vector machines | 91%      |
| Clark et al. (2005)  | HS   | 210           | 400–2500 nm | 1.6 m      | 7         | Species classes | Tree crown | Sunlit tree crown | Pixel          | Vegetation indexes + support vector machines | 84%      |
| Dalponte et al. (2013) | HS   | 160           | 400–1000 nm | 0.4 m      | 4         | Species classes | Tree crown | Pixel          | N/A           | Band selection + support vector machines | 93%      |
| Dalponte et al. (2008) | HS+ALS | 126           | 400–990 nm  | 1 m        | 19        | Species classes | Tree crown | Pixel          | 2             | Band selection + support vector machines | 89%      |
| Jones et al. (2010)  | HS+ALS | 492           | 395–2305 nm | 2 m        | 11        | Species classes | Pixel | 40            | 0             | Band selection by discriminant analysis of laboratory spectra. Support vector machines | 73%      |
| Koetz et al. (2008)  | HS+ALS | 102           | 454–923 nm  | 1 m        | 9         | Land cover classes | Pixel | 97            | 7             | Support vector machines | 75%      |
| Clark et al. (2011)  | HS+ALS | 210           | 400–2500 nm | 1.6 m      | N/A       | Biomass estimate | Plot | 9             | 9             | Narrowband spectral indexes. ALS vertical profile statistics | N/A      |
| Swatantran et al. (2011) | HS+ALS | 224           | 350–2500 nm | 3.3 m      | N/A       | Biomass estimate | Plot | 38            | 20            | 19 hyperspectral indices: mean and standard deviation. Quartile heights of ALS energy return + canopy cover: minimum, maximum, mean, standard deviation | N/A      |
The field data was split in two geographical areas, one for training of the methods and one for validation. After removal of shadows, non-vegetation and low vegetation (see later), the training data consisted of 36,470 pixels at 0.5 m pixel spacing, and the validation data 30,390 pixels (Table 4).

**Methods**

**Preprocessing**

In order to extract the tree canopy pixels for classification, a number of masks were applied to remove unwanted pixels: shadows, non-vegetation and low vegetation.

**Shadow mask**

Image pixels with no direct sunlight are discarded for two reasons. Firstly, the small amount of indirect light available for these pixels results in a low signal-to-noise ratio. Secondly, the reflected light from any of these pixels may be dominated by a spectral mixture of the object at the pixel itself, reflections from nearby objects and atmospheric scattering. On the other hand, for pixels with direct sunlight, the direct reflection dominates. By comparing the spectral profiles from well-illuminated tree canopy pixels with tree canopy pixels in shadow (Figure 2), it appears that a possible shadow criterion is that the intensity in the blue wavelengths is higher than in the green wavelengths. We selected the wavelength bands at 446 and 544 nm for blue and green, respectively, denoted as $\text{blue}_{446}$ and $\text{green}_{544}$. The green and blue intensity may be compared visually by computing a green-to-blue ratio. Shadows in the hyperspectral
image (Figure 3(a)) correspond well with dark areas in the green-to-blue ratio (Figure 3(b)).

Pixels with blue > green are classified as shadow (black in Figure 3(f)).

By visual inspection at several locations in the hyperspectral image, it appears that this criterion only excludes parts of tree canopies that are in full shadow. Partial shadows are not excluded.

Non-vegetation mask
The normalized difference vegetation index (NDVI; see, e.g., Aase & Siddoway, 1981; Campbell, 2006; Townshend, Goff, & Tucker, 1985) may be used to estimate the amount of plant chlorophyll, and thus to mask pixels with NDVI (Figure 3(c)) below a threshold as non-vegetation (black in Figure 3(g)).

\[
NDVI = \frac{NIR - red}{NIR + red}
\]

The NDVI threshold was selected by visual inspection to exclude buildings and dead trees, while keeping all live trees.

Pixels with NDVI < 0.55 are classified as non-vegetation.

For red, the average of the spectral channels with central wavelengths in the range 640–670 nm was used, and for near-infrared, 850–880 nm. These are the same wavelength ranges as for the Landsat 8 OLI satellite sensor.

Spectral normalization
The average spectra (Figure 4) suggest that birch is, on average, brighter than pine, which in turn is slightly brighter than spruce. However, there are large within-class variances in brightness, due to varying illumination conditions, for example, a well-illuminated spruce pixel at a tree top may be brighter than the average of birch pixels.

By scaling each pixel to have the same total intensity over all spectral channels, minor differences in the spectral profiles become apparent (Figure 5). This is fine for manual analysis of the spectral profiles to locate suitable spectral channels for species discrimination. However, for automatic analysis, a more robust way to adjust for varying illumination is that the spectrum for each pixel is normalized to standard illumination (Figure 6) by subtracting the mean intensity across all wavelengths for that pixel, and dividing by the standard deviation. This is called standard normal variate transformation (Barnes, Dhanoa, & Lister, 1989). The effect is that tree canopies have a more homogeneous intensity (Figure 3(e)). Note that intensity values are now...
negative in some parts of the spectrum (Figure 6), and may also be zero, meaning that spectral indices like NDVI cannot be used after spectral normalization.

**Classification methods**

Four new classification methods were compared in this study: three pixel classification methods and one image classification method.

**Indices method**

Rudjord and Trier (2016) proposed to compare average spectral profiles of spruce, pine and birch to locate a few wavelengths that may be used in ratios to separate the tree species. As noted earlier, the scaled radiance spectra (Figure 5) are better suited than the original spectra (Figure 4) for this analysis. In the lower blue wavelengths (415–413 nm), pine and spruce are equal and brighter than birch. However, we suspect that this may be due to different internal shadow patterns in birch canopies compared to spruce and pine canopies, so that well-illuminated pixels may not have the same trend. Instead, we suggest using the following wavelength bands. The near-infrared 751 nm channel is brighter for birch than for pine and spruce. The red-edge 715 nm channel is brighter for pine than for spruce. The green 534 nm channel is
brighter for pine and spruce than for birch, but may not be ideal since pine is slightly brighter than spruce.

Rudjord and Trier (2016) suggested using two indices: a spruce index (SI) to separate spruce from pine/birch, and a conifer index (CI) to separate birch from pine/spruce. Based on the above observations of spectral differences, we propose to use:

\[
CI = \frac{NIR_{751} - \text{green}_{534}}{NIR_{751} + \text{green}_{534}}
\]

\[
SI = \frac{NIR_{751} - \text{RE}_{715}}{NIR_{751} + \text{RE}_{715}}
\]

For both indices, histograms were prepared. The histogram counts were scaled so that each species had the same total count. For the conifer index (Figure 7), a threshold at 0.41 may be used to separate birch from pine, and for the spruce index (Figure 8), a threshold at 0.3725 may be used to separate spruce from birch and pine. However, there are large overlaps between the histograms, indicating that the misclassification rates may be high.

Figure 6. Normalized spectra for spruce, pine and birch.

Figure 7. Conifer index histograms.

Figure 8. Spruce index histograms.
A simple decision tree was used. First, the spruce index was used to classify tree canopy pixels with SI > 0.3725 as spruce. Then the conifer index was used to classify the remaining tree canopy pixels as pine if CI > 0.41, otherwise birch. Note that these thresholds have been established using at-sensor radiances. In order to make them applicable to other hyperspectral datasets, they need to be re-established using atmospherically corrected hyperspectral data, i.e. surface reflectance values.

**Partial least squares method**

Instead of selecting wavelengths by manually comparing spectral profiles, as in the indices method described earlier, one may use an automatic method for this. Also, we need to investigate if the use of more wavelengths may improve classification performance.

We used partial least squares regression (Höskuldsson, 1988; Martens & Martens, 1986; Westad & Martens, 2000) to identify a set of 14 spectral channels for each tree species (Figure 9, Table 5). Thus, for each tree species, there is one classifier, which, ideally, returns +1 for a hyperspectral pixel of the same class, and −1 otherwise. In practice, a floating point value is returned, indicating how well the pixel resembles that class. The three single-species pixel classifiers are combined, by returning the class name of the classifier that gives the highest output value for a particular pixel.

Partial least squares regression methods are related to other multivariate modelling methods like principal component regression and ordinary least squares. As high-resolution spectra exhibit a large degree of correlation between spectral responses which are close in wavelength, leading to co-linearity in the data, ordinary least squares methods tend to become numerical unstable due to the matrix inversion involved in the solution.

Partial least squares regression exploits the co-linearity between input spectra and the predictive response by dimensionality reduction through maximization of variance between the spectral input and the variables to predict. The regression coefficients effectively indicate interesting spectral regions contributing to the modelling result as well as regions or specific wavelengths which may be ignored without affecting the modelling result. Thus, partial least squares regression is a method of choice when dealing with analysis of spectral data.

The best selection of wavelengths may be systematically searched for during cross-validation. In this approach, a portion (here, 10%) of the input spectra and responses are systematically kept out during the training of the model, and instead used for testing the predictive performance of the model. When all the

---

**Figure 9.** Partial least squares regression coefficients per selected wavelength, for each tree species. Top: birch, middle: spruce, bottom: pine.

**Table 5.** Selected wavelengths (nm) for the partial least squares regression models for birch, spruce and pine.

|        | Birch | Spruce | Pine |
|--------|-------|--------|------|
|        | 522.7 | 526.3  | 526.3|
|        | 526.3 | 573.5  | 653.3|
|        | 693.3 | 696.9  | 682.4|
|        | 700.5 | 704.3  | 686.0|
|        | 718.7 | 722.3  | 729.5|
|        | 725.9 | 725.9  | 733.2|
|        | 736.8 | 736.8  | 736.8|
|        | 740.4 | 740.4  | 740.4|
|        | 769.5 | 769.5  | 769.5|
|        | 758.6 | 758.6  | 758.6|
|        | 762.2 | 762.2  | 762.2|
|        | 754.9 | 754.9  | 754.9|
|        | 729.5 | 729.5  | 729.5|
|        | 733.2 | 733.2  | 733.2|
|        | 736.8 | 736.8  | 736.8|
|        | 740.4 | 740.4  | 740.4|
|        | 769.5 | 769.5  | 769.5|
|        | 820.3 | 820.3  | 823.9|
data has been used both for training and for testing, the variability of regression coefficients during cross-validation is analysed and compared for each model built during the cross-validation. High-variability coefficients are too dependent on single spectra in the training data, while more stable coefficients tend to indicate the underlying structure that is needed for adequate model performance.

By repeating the cross-validation procedure through several iterations, removing less important coefficients while still upholding the predictive performance of the model at every step, a final set of coefficients representing the important wavelengths describing the relationship between spectra and predictive responses can be found.

Partial least squares has been used in other forest-related studies utilizing ALS to estimate properties such as wood volume, tree heights and mean stem diameters simultaneously (e.g. Næsset, Bollandsås, & Gobakken, 2005).

**Deep learning image classification using canopy shape and three hyperspectral bands**

Since 2012, deep learning has been the preferred methodology in computer vision, including automatic object detection and classification (e.g. see Krizhevsky et al., 2012; LeCun et al., 2015). These methods have been used on standard colour photographs of natural scenes. In order to transform our ALS and hyperspectral data to something resembling three-channel colour images, suitable for such methods, we have combined ALS vegetation height with three selected hyperspectral channels.

The Caffe library (Jia et al., 2014) is a python implementation of neural network components and may be used to construct and train a deep neural network. A network with the following structure was defined as follows:

1. Input layer, accepting images of size 32 × 32 pixels × 3 bands
2. Convolution layer with 20 convolution kernels of size 5 × 5 pixels, stride 1
3. Rectified linear units (ReLU) layer (no parameters)
4. Max pooling layer with kernel size 2 × 2 pixels and stride 2
5. Convolution layer with 50 convolution kernels of size 5 × 5 pixels
6. ReLU layer
7. Max pooling layer with kernel size 3 × 3 pixels and stride 2
8. Fully connected layer with 100 output nodes
9. ReLU layer
10. Fully connected layer with 3 output nodes, i.e. one output node per class

The purpose of the convolution layers (2 and 5) is to look for interesting subimage structures, e.g. edges, bright or dark spots, texture, etc., that may contribute to the objects being searched for. The ReLU layers (3, 6 and 9) introduce non-linearity into the decision boundary. They suppress negative feedback and copy positive feedback by using the function f(x) = max(x, 0). The max pooling layers (4 and 7) provide some degree of spatial invariance as to where the image primitives are located within the image. The fully connected layers (8 and 10) provide a high degree of flexibility in how the neural network learns from the training data.

Training and validation data (Table 6) for the network was extracted as follows. For each tree within the labelled training data, a 32 × 32 subimage (6.4 m × 6.4 m) centred on the tree was extracted. Three bands from the hyperspectral data were extracted and used to blend the single channel vegetation height image into a three channel image. For each pixel in the blended image, the three channels were scaled so that their sum equalled the vegetation height in that pixel.

In this study, we did not use a proper tree canopy segmentation method, although several exist (e.g. Dalponte et al., 2014; Ene, Næsset, & Gobakken, 2012). For the purpose of evaluating a classifier, a simplified tree detection method was designed as follows. On vegetation height images with 20 cm pixel spacing, a sliding window of size 32 × 32 pixels (6.4 m × 6.4 m) with step size 10 pixels (2.0 m) was used. At each window location, the class label of the centre pixel was inspected. If the class label was one of “shadow”, “no vegetation”, “low vegetation”, “no data” or “unknown”, then the location was skipped. However, if the class label was one of the tree species (“spruce”, “pine” or “birch”), then the highest pixel within the window is regarded as a tree top and centre of the tree canopy. If the class label of the tree top is the same as of the window centre, then the tree top is accepted and added to the list of identified tree tops.

In order to create eight times as many training images, a copy of each training image was flipped, and then rotated versions were made with 90°, 180° and 270° rotation.

Training of the deep neural network with cross-validation on the training data reached 88% overall correct classification (Figure 10) after 1800 iterations of training the net, with class-specific classification rates at 93%, 89% and 68% for spruce, pine and birch, respectively.

**Table 6.** The number of trees used for training and validation for the deep learning image classification method.

|         | Spruce | Pine | Birch | Total |
|---------|--------|------|-------|-------|
| Training data | 170    | 336  | 63    | 569   |
| Validation data | 226    | 196  | 49    | 471   |
Deep learning using hyperspectral pixels

Although designed for image data, the Caffe library may be used to construct a deep neural network for hyperspectral image pixels. The network structure suggested by Hu et al. (2015) was defined as follows:

1. Input layer, accepting “images” of size 160 × 1 pixels × 1 band
2. Convolution layer with 20 convolution kernels of size 17 × 1 pixels
3. ReLU layer
4. Max pooling layer with kernel size 4 × 1 and stride 1
5. Fully connected layer with 100 output nodes
6. ReLU layer
7. Fully connected layer with 3 output nodes, i.e. one output node per class

Training of the deep neural network with cross-validation on the training data reached 90% overall correct classification (Figure 11) after 8000 iterations, with class-specific rates at 92%, 90% and 88% for spruce, pine and birch, respectively.

Results

The four different methods were applied on the validation set. Visually, the classification results had some similarities and some differences (Figure 12). The partial least squares method (Figure 12(c)) resulted in more predicted pine than for the other methods, while both deep learning methods (Figure 12(d,e)) resulted in more predicted birch than for the two other methods (Figure 12(b,c)). For the labelled validation data (Figure 12(f)), the deep learning hyperspectral pixel method was best (Table 7), with 87% correct classification, which was close to the estimated 90% correct classification from cross validation on the training data.

Both the index method (Table 8) and the partial least squares method (Table 9) overestimated the amount of pine in the validation data. For the deep learning image method (Table 10), both pine and spruce were underestimated, with a large overestimation of birch.

The deep learning hyperspectral pixel method (Table 11) underestimated spruce by 10%, while birch was underestimated by 30%. However, by looking at the parts of the validation image without any ground truth, this method had the largest number of predicted birch pixels (Figure 12(d)). This indicates that another field trip to study the differences in classification results would be useful. For pine, the number of true pine pixels that were misclassified was balanced by the number of spruce and birch pixels that were misclassified as pine (Table 11).

The classification results were also compared visually in some selected areas. In an open area with scattered trees (Figure 13), the index method split all the trees into parts of different species (Figure 13(d)). The partial least squares method correctly classified most of the spruce and pine trees, but misclassified the birch trees (Figure 13(g)). The deep learning hyperspectral pixel method correctly classified the spruce and pine trees, and one of the birch trees, while the other birch tree was classified as a mix of all three species (Figure 13(h)). The deep learning image method correctly classified one pine and one birch but misclassified the other birch tree (Figure 13(f)). The spruce tree was missed (Figure 13(f)), possibly due to its small size.
In a pine forest with 50–60% canopy cover (Figure 14), all methods worked fairly well, with relatively few misclassifications. The partial least squares method worked best (Figure 14(g)), most likely because it was overestimating the amount of pine (Table 9). The deep learning image method made one misclassification of pine as spruce (Figure 14(f)).

In a dense spruce forest with 80–100% canopy cover (Figure 15), more than half of the pixels were excluded from classification due to large amounts of shadow (Figure 15(b)). All the methods worked well. The index method misclassified one spruce tree as pine (Figure 15(d)). The partial least squares method misclassified parts of a few trees (Figure 15(g)). The deep learning hyperspectral pixel method misclassified one spruce tree as

![Classification maps for the validation data, with pink = birch, green = pine, blue = spruce. (a) Three channels of the hyperspectral data: red, near infrared, green. (b) Pixel classification with indices method. (c) Pixel classification with partial least squares method. (d) Pixel classification with deep learning. (e) Image classification with deep learning. (f) Ground truth from field work for validation data.](image-url)
Table 7. Classification rates for the different methods on the validation set.

| True class | Indices | Partial least squares regression | Deep learning hyperspectral pixels | Deep learning images |
|------------|---------|----------------------------------|-----------------------------------|---------------------|
| Spruce     | 66%     | 95%                             | 95%                               | 77%                 |
| Pine       | 80%     | 88%                             | 88%                               | 67%                 |
| Birch      | 26%     | 61%                             | 61%                               | 86%                 |
| Total, test data | 67% | 78%                             | 87%                               | 74%                 |
| Cross-validation on training data | | | | 90% 88% |

Table 8. Confusion matrix for the index method on the validation set.

| True class | Spruce | Pine | Birch | Sum | Correct classification rate |
|------------|--------|------|-------|-----|-----------------------------|
| Spruce     | 7990   | 3422 | 631   | 12,043 | 66% |
| Pine       | 1819   | 11,329 | 980   | 14,128 | 80% |
| Birch      | 1285   | 1836 | 1098   | 4219   | 26% |
| Sum        | 11,094 | 16,587 | 2709   | 30,390 | 67% |
| Over/underestimation | −8% | 17% | −36% | |

Table 9. Confusion matrix for the partial least squares method on the validation set.

| True class | Spruce | Pine | Birch | Sum | Correct classification rate |
|------------|--------|------|-------|-----|-----------------------------|
| Spruce     | 9562   | 2464 | 17    | 12,043 | 79% |
| Pine       | 275    | 13,846 | 7     | 14,128 | 98% |
| Birch      | 428    | 3578 | 213   | 4219   | 5% |
| Sum        | 10,265 | 19,888 | 237   | 30,390 | 78% |
| Over/underestimation | −15% | 41% | −94% | |

Table 10. Confusion matrix for the validation set, for classification of $32 \times 32$ pixels images with three channels, using deep learning.

| True class | Spruce | Pine | Birch | Sum | Correct classification rate |
|------------|--------|------|-------|-----|-----------------------------|
| Spruce     | 173    | 34   | 19    | 226 | 77% |
| Pine       | 32     | 132  | 32    | 196 | 67% |
| Birch      | 1      | 6    | 42    | 49  | 86% |
| Sum        | 206    | 172  | 93    | 471 | 74% |
| Over/underestimation | −9% | −12% | 90% | |

Table 11. Confusion matrix and classification accuracy for the validation set, for pixel classification with 160 hyperspectral visible and near infrared channels using deep learning.

| True class | Spruce | Pine | Birch | Sum | Correct classification rate |
|------------|--------|------|-------|-----|-----------------------------|
| Spruce     | 11,497 | 407  | 139   | 12,043 | 95% |
| Pine       | 1482   | 12,381 | 265  | 14,128 | 88% |
| Birch      | 279    | 1384 | 2556  | 4219   | 61% |
| Sum        | 13,258 | 14,172 | 2960  | 30,390 | 87% |
| Over/underestimation | 10% | 0% | −30% | |

All the pixel based methods misclassified a few edge pixels on other trees (Figure 15(d,g and h)). The deep learning image method misclassified one spruce tree as birch (Figure 15(f)).

**Discussion and conclusions**

The result of using a pixel-based classification appeared less noisy than expected. Rather, the classified pixels often appeared in homogeneous clusters that represent single trees or small groups of trees. However, there were also examples of single pixels or a few pixels that appeared as misclassified. If we had a delineation of the individual tree crowns, these noisy pixels could be removed by majority voting.

One potential source of misclassification could be that the shadow mask had a too weak shadow criterion. For example, by increasing the green-to-blue ratio limit from 1.0 to 1.1, the deep learning hyperspectral method reduced the number of misclassified pine pixels by 231. However, 1084 correctly classified pine pixels were also removed. Thus, a stricter shadow mask does not appear to be a good solution.

One major difference between the four proposed methods is the number of features. The indices method used only two features, extracted from three channels. One may suspect that this method could perform better than 67% classification accuracy by adding more features.
The partial least squares method uses 14 features per class, from a total of 27 channels (Table 5), but performs better (78% accuracy) than the deep learning image classification method (74% accuracy). The latter method uses $32 \times 32$ pixels $\times 3$ bands = 3072 input features. However, there is some hidden feature extraction inside a deep neural network. Still, the lower accuracy of the deep learning image classification method may be due to an insufficient number of training examples to learn all the internal parameters in the neural network.

The deep learning hyperspectral pixel method used 160 input features. It achieved slightly worse (87%) accuracy than the pixel method (90%) of Dalponte et al. (2013), using the same sensor, Hyspex VNIR 1600. However, the resolution in our data was slightly coarser (0.5 m) than in their data (0.4 m).

The next planned action is to make a field revisit, with the classification maps, to identify misclassified trees, in order to include these in an extended training set and then to re-calibrate and re-run the classification method. Further, the results presented here are for one acquisition time only, with training data...
from the same acquisition. We may expect to see somewhat different species-specific spectra at other dates. Spruce is known to be more bright green in June, when fresh needles appear at the tips of the branches. It would be interesting to repeat the experiments with hyperspectral data from another time of the year to assess the amount of training needed for a new acquisition.

There is also potential for improvement of the best method, deep learning with hyperspectral pixels, by also including the tree height and/or canopy shape in the classification method, as this information is currently not used, except for masking. A simple extension of the deep neural network to include canopy height, in a similar fashion as proposed by Morchhale et al. (2016), was attempted but did not work. One could design a deep neural network that combines hyperspectral pixel data and vegetation height image data in a better way than in the current deep learning image method, which only uses three hyperspectral bands. One challenge is to collect sufficiently large training and validation data sets for an image-based method that uses all the hyperspectral channels. A possible solution could be to use two classifiers in parallel: a deep neural network for classification of single-channel vegetation height images and the current deep neural network for hyperspectral pixel classification.

The tree height is a quite good proxy for tree age. In photointerpretation of broadband multispectral images, it is a well-known problem that young spruce may be confused spectrally with birch, and old spruce with young pine. It would be interesting to investigate if this is also true for narrow bands in hyperspectral data. In more general, how does the tree age (or height) influence the entire spectra for spruce, pine and birch?

In conclusion, we obtain about 87% correct classification rate in the three-class problem of discriminating between spruce, pine and birch, by using deep learning for hyperspectral pixel classification with 160 spectral channels in the visible and near-infrared spectrum. We have identified a number of possible improvements that we plan to pursue in the near future.

Geolocation information: 10°43′34.50″E–10°44′17.70″E, 60°10′43.20″N–60°13′45.14″N

Acknowledgments

This research was financed in part by the BIONÆR program of the Research Council of Norway, project no. 244,599. This article presents a research collaboration to develop more automated methods for forest mapping in Norway and Scandinavia. TerraTec AS is a commercial provider of airborne remote sensing data, including airborne laser scanning (ALS), multispectral imagery and hyperspectral imagery. The Norwegian Computing Center is specializing in developing (semi-) automatic methods for the processing of remote sensing data.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was financed in part by the BIONÆR program of the Research Council of Norway, project no. 244,599.

References

Aase, J.K., & Siddoway, F.H. (1981). Assessing winter wheat dry matter production via spectral reflectance measurements. Remote Sensing of Environment, 11, 267–277. doi:10.1016/0034-4257(81)90025-0
Barnes, R.J., Dhanoa, M.S., & Lister, S.J. (1989). Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. Applied Spectroscopy, 43, 772–777. doi:10.1366/0003702894202201
Campbell, J.B. (2006). Introduction to remote sensing (4th ed.). Oxon, UK: Taylor and Francis.
Clark, M.L., Roberts, D.A., & Clark, D.B. (2005). Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. Remote Sensing of Environment, 90(3–4), 375–398. doi:10.1016/j.rse.2005.03.009

Clark, M.L., Roberts, D.A., Ewel, J.J., & Clark, D.B. (2011). Estimation of tropical rain forest aboveground biomass with small-footprint lidar and hyperspectral sensors. Remote Sensing of Environment, 115(11), 2931–2942. doi:10.1016/j.rse.2010.08.029

Dalponte, M., Bruzzone, L., & Gianelle, D. (2008). Fusion of hyperspectral and lidar remote sensing data for classification of complex forest areas. IEEE Transactions on Geoscience and Remote Sensing, 46(5), 1416–1427. doi:10.1109/TGRS.2008.916480

Dalponte, M., Ökra, H.O., Gobakken, T., Gianelle, D., & Naesset, E. (2013). Tree species classification in boreal forests with hyperspectral data. IEEE Transactions on Geoscience and Remote Sensing, 51(5), 2632–2645. doi:10.1109/TGRS.2012.2216272

Dalponte, M., Ökra, H.O., Ene, L.T., Gobakken, T., & Naesset, E. (2014). Tree crown delineation and tree species classification in boreal forests using hyperspectral and ALS data. Remote Sensing of Environment, 140, 306–317. doi:10.1016/j.rse.2013.09.006

Ene, L., Naesset, E., & Gobakken, T. (2012). Single tree detection in heterogeneous boreal forests using airborne laser scanning and area-based stem number estimates. International Journal of Remote Sensing, 33(16), 5171–5193. doi:10.1080/01431161.2012.657363

Fassnacht, F.E., Latifi, H., & Koch, B. (2012). An angular vegetation index for imaging spectroscopy data—Preliminary results on forest damage detection in the Bavarian National Park, Germany. International Journal of Applied Earth Observation and Geoinformation, 19, 308–321. doi:10.1016/j.jag.2012.05.018

Fassnacht, F.E., Latifi, H., Stereiczak, K., Modelezewskaja, A., Lefsky, M., Waser, I.T., … Ghosh, A. (2016). Review of studies on tree species classification from remotely sensed data. Remote Sensing of Environment, 186, 64–87. doi:10.1016/j.rse.2016.08.013

Gao, B.-C., Montes, M.J., Davis, C.O., & Goetz, A.F.H. (2009). Atmospheric correction algorithms for hyperspectral remote sensing data of land and ocean. Remote Sensing of Environment, 113, S17–S24. doi:10.1016/j.rse.2007.12.015

Goetz, A.F.H. (2009). Three decades of hyperspectral remote sensing of the Earth: A personal view. Remote Sensing of Environment, 113, S5–S16. doi:10.1016/j.rse.2010.12.014

Höskuldsson, A. (1988). PLS regression methods. Journal of Chemometrics, 2(3), 211–228. doi:10.1002/cem.1180020306

Hu, E., Huang, Y., Wei, L., Zhang, F., & Li, H. (2015). Deep convolutional neural networks for hyperspectral image classification. Journal of Sensors, Article ID 258619, 12. doi:10.1155/2015/258619

Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., … Darrell, T. (2014). Caffe: Convolutional architecture for fast feature embedding. Proceedings of the 22nd ACM International Conference on Multimedia (MM ’14), 3–7 November 2014, Orlando, Florida, USA (pp 675–678).

Jones, T.G., Coops, N.C., & Sharma, T. (2010). Assessing the utility of airborne hyperspectral and LiDAR data for species distribution mapping in the coastal Pacific Northwest, Canada. Remote Sensing of Environment, 114(12), 2841–2852. doi:10.1016/j.rse.2010.07.002

Koetz, B., Morsdorf, F., van den Linden, S., Curt, T., & Allgöwer, B. (2008). Multi-source land cover classification for forest fire management based on imaging spectrometry and LiDAR data. Forest Ecology and Management, 256(3), 263–271. doi:10.1016/j.foreco.2008.04.025

Kokaly, R.F., Asner, G.P., Ollinger, S.V., Martin, M.E., & Wessman, C.A. (2009). Characterizing canopy biochemistry from imaging spectroscopy and its application to ecosystem studies. Remote Sensing of Environment, 113, S78–S91. doi:10.1016/j.rse.2008.10.018

Korpeš, I., Mehtätalo, L., Seppänen, A., & Kangas, A. (2014). Tree species identification in aerial image data using directional reflectance signatures. Silva Fennica, 48(3), article id 1087. doi:10.14214/sf.1087

Krizhevsky, A., Sutskever, I., & Hinton, G.E. (2012). December 3–8). ImageNet classification with deep convolutional neural networks. Proceedings of the Neural Information Processing Systems Conference (NIPS 2012) (pp. 1106–1114), Lake Tahoe, Nevada.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521, 436–444. doi:10.1038/nature14539

Martens, M., & Martens, H. (1986). Partial least squares regression. In J.R. Piggot (Ed.), Statistical procedures in food research (pp. 293–359). London, UK: Elsevier.

Morchhalé, S., Pauca, V.P., Plennmons, R.J., & Torgersen, T.C. (2016, August 21–24). Classification of pixel-level fused hyperspectral and lidar data using deep convolutional neural networks. 8th Workshop on Hyperspectral Image and Signal Processing (WHISPERS): Evolution in Remote Sensing (5 pp.). Los Angeles, USA. doi:10.1109/WHISPERS.2016.8071715

Naesset, E., Bollandsås, O.M., & Gobakken, T. (2005). Comparing regression methods in estimation of biophysical properties of forest stands from two different inventories using laser scanner data. Remote Sensing of Environment, 94(4), 541–553. doi:10.1016/j.rse.2004.11.010

Plaza, A., Benediktsson, J.A., Boardman, J.W., Brazile, J., Bruzzone, L., Camps-Valls, G., … Trianni, G. (2009). Recent advances in techniques for hyperspectral image processing. Remote Sensing of Environment, 113, S110–S122. doi:10.1016/j.rse.2007.07.028

Rudjord, Ø., & Trier, O.D. (2016, August 21–24). Tree species classification with hyperspectral imaging and lidar. 8th Workshop on Hyperspectral Image and Signal Processing (WHISPERS): Evolution in Remote Sensing (4 pp.). Los Angeles, USA. doi:10.1109/WHISPERS.2016.8071665

Schaepe, M.E., Ustin, S.L., Plaza, A.J., Painter, T.H., Verrelst, J., & Liang, S. (2009). Earth system science related imaging spectroscopy—An assessment. Remote Sensing of Environment, 113, S123–S137. doi:10.1016/j.rse.2009.03.001

Svantran, A., Dubayah, R., Roberts, D., Hofton, M., & Blair, J.B. (2011). Mapping biomass and stress in the Sierra Nevada using lidar and hyperspectral data fusion. Remote Sensing of Environment, 115(11), 2917–2930. doi:10.1016/j.rse.2010.08.027

Townshend, J.R.G., Goiff, T.E., & Tucker, C.J. (1985). Multitemporal dimensionality of images of normalized difference vegetation index at continental scales. IEEE Transactions on Geoscience and Remote Sensing, 23(6), 888–895. doi:10.1109/TGRS.1985.289474

Ustin, S.L., Gitelson, A.A., Jacquemoud, S., Schaepe, M., Asner, G.P., Gamon, J.A., & Zarco-Tejada, P. (2009). Retrieval of foliar information about plant pigment systems from high resolution spectroscopy. Remote Sensing of Environment, 113, S67–S77. doi:10.1016/j.rse.2008.10.019

Westad, F., & Martens, H. (2000). Variable selection in near infrared spectroscopy based on significance testing in partial least squares regression. Journal of Near Infrared Spectroscopy, 8(2), 117–124. doi:10.1255/jnirs.271