Transferability and the effect of colour calibration during multi-image classification of Arctic vegetation change

Samira Kolyaie[1], Urs Albert Treier[1,2,3], Gary Richard Watmough[1,4], Bjarke Madsen[1], Peder Klith Bøcher[1], Achilleas Psomas[2], Ruedi Bösch[2], Signe Normand*[1,3,5]

[1] Section for Ecoinformatics & Biodiversity, Center for Biodiversity Dynamics in a Changing World, Department of Bioscience, Aarhus University, Ny Munkegade 116, DK-8000 Aarhus C, Denmark

[2]Swiss Federal Research Institute WSL, Remote Sensing Group, Zürcherstrasse 111, CH-8903 Birmensdorf, Switzerland

[3] Arctic Research Centre, Department of Bioscience, Aarhus University, Ny Munkegade 116, DK-8000 Aarhus C, Denmark

[4] School of Geosciences, University of Edinburgh, EH8 9XP Edinburgh, United Kingdom

[5] Swiss Federal Research Institute WSL, Landscape Dynamics Group, Zürcherstrasse 111, CH-8903 Birmensdorf, Switzerland

*Corresponding author: signe.normand@bios.au.dk, +45 23718009, https://orcid.org/0000-0002-8782-4154

Acknowledgements

The research was initiated with a WSL internal innovative research grant and supported by a Villum Young Investigator grant (VKR023456), an Aarhus University (AU), Research Foundation grant (AUFF-E-2015-FLS-8-73) and an AU Science and Technology Synergy grant. Fieldwork was possible due to funding from the AU Arctic Research Centre. We are grateful to Rok Kreslin and Peter Peer for providing their code in MATLAB for the DT method; to Constantinos Tsirogiannis for his valuable help in code optimization and to Ditte Grube Barild for geo-referencing the images.

This document is the accepted manuscript version of the following article:
Kolyaie, S., Treier, U. A., Watmough, G. R., Madsen, B., Bøcher, P. K., Psomas, A., … Normand, S. (2019). Transferability and the effect of colour calibration during multi-image classification of Arctic vegetation change. Polar Biology, 42(7), 1227-1239.
https://doi.org/10.1007/s00300-019-02491-7

Abstract

Mapping changes in vegetation cover is essential for understanding the consequences of climate change on Arctic ecosystems. Classification of ultra-high spatial-resolution (UHR, <1cm) imagery can provide estimates of vegetation cover across space and time. The challenge of this approach is to assure comparability of classification across many images taken at different illumination conditions and locations. With warming, vegetation at higher elevation is expected to resemble current vegetation at lower elevation. To investigate the value of classification of UHR imagery for monitoring vegetation change, we collected visible and near infrared images from 108 plots with handheld cameras along an altitudinal gradient in Greenland and examined the classification accuracy of shrub cover on independent images (i.e. classification transferability). We implemented several models to examine if colour calibration improves transferability based on an in-image calibration target. The classifier was trained on different number of images to find the minimum training subset size. With a training set of ~20% of the images the overall accuracy levelled off at about 81% and 68% on the non-calibrated training and validation images, respectively. Colour calibration improved the accuracy on training images (1-4%) while it only improved the classifier transferability significantly for training sets <20%. Linear calibration only based on the target's grey series improved transferability most. Reasonable transferability of Arctic shrub cover classification can be obtained based only on spectral data and about 20% of all images. This is promising for vegetation monitoring through multi-image classification of UHR imagery acquired with hand-held cameras or Unmanned Aerial Systems.

KEY WORDS: Arctic tundra, climate change, colour calibration, standardization, spectral data, classification transferability

Introduction

The Arctic is warming faster than the rest of the world (Masson-Delmotte et al. 2013). Several studies document recent vegetation changes in response to the increasing temperatures (Tape et al. 2006; Myers-Smith et al. 2011; Elmendorf et al. 2012; Myers-Smith et al. 2015; Guay et al. 2015; Nielsen et al. 2017) and model projections highlight the potential for large future changes (Pearson et al. 2013; Normand et al. 2013). Arctic shrub species have been found to increase growth, cover, and height in response to warming, but to varying degrees depending on local environmental conditions (Tape et al. 2006; Elmendorf et al. 2012; Myers-Smith et al. 2015; Nielsen et al. 2017). Increased height and dominance of Arctic shrubs are expected to negatively affect the cover of bryophytes and lichens (Elmendorf et al. 2012), change composition of arthropod communities (Hansen et al. 2016), speed up climate change (Myers-Smith et al. 2015), and lead to profound changes in Arctic ecosystems (Post et al. 2009). Mapping and monitoring changes of Arctic shrub cover is crucial for understanding the spatial magnitude of the potential biodiversity and ecosystem consequences of climate change in the Arctic.

Studying changes in vegetation at different spatial and temporal scales is a central challenge in ecology. Fine resolution data are required for studying local changes in vegetation cover (Elmendorf et al. 2012) and for upscaling locally observed patterns across larger areas (e.g., Liu and Treitz 2016). Point framing or visual cover estimation in the field are commonly used methods for providing fine resolution data (Luscier et al. 2006; Liu and Treitz 2016). However, providing these data is either expensive (time/cost; point-frame method) or has reduced reproducibility due to the observer's bias, with an unknown error distribution, which limits the inference of vegetation changes (Neeser et al. 2000; Tichy 2016; Kercher et al. 2003). Vegetation cover estimation using ultra-high spatial-resolution (UHR) images taken by handheld-cameras or Unmanned Aerial Systems is a promising and pragmatic approach that can speed up field data collection (Booth and Cox 2008; Bold et al. 2010; Bricher 2012). Moreover, vegetation cover can be measured by image classification with known accuracy (Lengyel et al. 2008) and the source data can be archived for objectivity and reproducibility of measurements. This can improve our ability for fine-scale vegetation mapping and monitoring (Lengyel et al. 2008; Zlinszky et al. 2015; He et al. 2015) as well as for detailed investigations of vegetation characteristics (Neumann et al. 2015). Here, we seek for an effective and standard processing method to improve field-based observations by UHR images. The goal is to increase comparability of vegetation cover estimates across space and time.

Several researchers have used UHR images taken by handheld cameras from vegetation plots and provided measurements that are more reliable compared to field-based observations. They mainly applied object-based image classification to measure ground or vegetation cover (Luscier et al. 2006; Chen et al. 2010; Liu and Treitz 2016). The key aspect for operational use of UHR imagery in ecological field-based studies is the ability to semi-automatically classifying large numbers of images based on reference data collected from the smallest possible number of images. Therefore, we need novel methodologies for the analysis of UHR imagery to obtain vegetation cover and other ecologically relevant parameters efficiently. A pragmatic approach is to train the classifier based on a limited number of images and using that classifier to classify other images (i.e. transferring the classifier). Monitoring vegetation change with this approach can be further challenging because the images are from different locations and different times (hours, days, years). Therefore, vegetation composition, health and life stage, as well as illumination conditions, are likely to vary among images. Vegetation characteristics (species, age, and health condition) and illumination conditions both influence the chromatic outcomes of the vegetation in the images (Jackowski et al. 1997; Villafuerte and Negro 1998; Ritchie et al. 2008; Menesatti et al. 2012; Wang et al. 2013). This leads to high intra-class variation for the classification and make obtaining a representative reference dataset of the images challenging (Gehler and Nowozin 2009).

Colour calibration is an important approach to mitigate intra-class variation of reflectance due to the illumination differences among images (Finlayson and Trezzi 2004; Gehler et al. 2008; Wang et al. 2013). Colour calibration has successfully improved image interpretation and analysis for applications in ecology (Villafuerte and Negro 1998), environmental monitoring (Hyman 2010), food science (Quevedo et al. 2010), medicine (Wang et al. 2013), as well as art and museum documentation (Berns et al. 2005). Using colour calibration, the spectral values of the images are converted to standard values using a mathematical model, e.g., polynomial regression models (Wang and Zhang 2010), exponential models (Fischer et al. 2012), or transformation using Delaunay Triangulation (DT) (Kreslin et al. 2014). Defining the models' parameters depends on the relationship between standard values of a calibration target (e.g., Macbeth colour checker, McCamy et al. 1976), placed in the image swath at the image acquisition time, and the values measured from the calibration target on the acquired images. Jackowski et al. (1997) calibrated 20 images based on a Gaussian basis function and the Macbeth colour checker and achieved calibrated images with values closer to the standard values on the calibration target. Polynomial regression models are widely used for colour calibration purposes (Wang and Zhang 2010). Wang and Zhang (2010) calibrated over 300 images for disease diagnosis and showed that a polynomial-based regression provided the best calibration, compared with calibrated images with ridge, support vector, and neural network regressions. Kreslin et al. (2014) tested different colour calibration models on 568 images (each containing a Macbeth colour checker) acquired under different indoor and outdoor illumination conditions. They found that DT-based transformation outperformed other calibration models in producing closer values to the Macbeth colour checker's standard values. The above studies applied colour calibration on imagery including the three visible bands (R, red; G, green; B, blue, hereafter RGB). Near-infrared data are often valuable for vegetation classification and monitoring (Fischer et al. 2012). Using an exponential equation for radiometric calibration of RGB and near-infrared (NIR) images of biological soil crust, Fischer et al. (2012) documented a high linear correlation (r^2 = 0.91) between estimates of the normalized difference vegetation index (NDVI) from the calibrated images and data obtained from a field spectrometer. Hence, polynomial and DT-based colour calibration models show promise for colour calibration of large RGB image datasets and exponential equations show promise for calibration of NIR images. Nonetheless, while it is documented that colour calibration provides a good standardization of reflectance values across images, the importance of colour calibration for reducing intra-class variation and improving classifier transferability during classification of multiple images remains unknown.

Patterns of vegetation composition considerably change along elevational gradients (Engler et al. 2011; Morueta-Holme et al. 2015). With warming, vegetation at higher elevations potentially will become more similar to the current vegetation at lower elevations (Engler et al. 2011; Morueta-Holme et al. 2015). We acquired RGB and NIR images with two handheld cameras from 108 plots distributed across an altitudinal gradient in western Greenland to assess the spatial and temporal classification transferability of UHR imagery for shrub cover quantification. Our overall goal was to examine the effect of colour calibration on the transferability of the classifier and to optimize a multi-image classification framework to automatize monitoring of Arctic vegetation change. Specifically, we addressed the following questions on spectral data: (i) How accurate can we classify images in a multi-image classification framework, (ii) does colour calibration increase classification transferability, and (iii) what is the minimum reference data set for optimising classification transferability.

Materials and methods

Study area and sampling design

Digital images of 108 permanent plots (80×80 cm) were sampled from the 21st to 24th of July 2013 in a valley in the inner Nuuk fiord (Latitude: 64.2093; Longitude: -50.2920) (Fig. 1). The plots were distributed stratified random across altitudinal isoclines (at 20, 100, 200, 300, 400, and 500 m a.s.l.). Three groups of six plots were approximately 500 m apart along each isocline and plots were placed 10 m apart within each plot group (for more details on the sampling design see Nabe-Nielsen et al. 2017). Vegetation in the area is composed of a mosaic of several dwarf, low and tall shrub species (*Betula nana, Cassiope tetragona, Dryas integrifolia, Empetrum nigrum, Ledum groenlandicum, Ledum palustre, Phyllodoce coerulea, Salix glauca, Salix arctophila, Vaccinium vitis-idaea, Vaccinium uliginosum*), graminoids (*Juncaceae, Cyperaceae, Poaceae*), other herbs, bryophytes, lichens, pteridophytes, and bare ground. In this study, we define shrubs as encompassing dwarf, low, and tall shrubs (cf. Myers-Smith et al. 2015).

Image data

We used two handheld single-lens reflex cameras (Canon EOS 550D) to collect the image data. We acquired the visible (VIS) light spectrum with one of the cameras and modified the other one to acquire the near-infrared (NIR) light spectrum (Fig. 1 & 3) by replacing the low-pass filter to restrict the cameras sensitivity to wavelengths above 800nm (http://www.optic-makario.de/transmissionskurven/: IR LP2-830nm). Raw images were converted to 8-bit TIFF images by applying the appropriate lens correction model with standard parameters (Adobe Photoshop Camera Raw 6.7). Due to the build in Bayer filter both cameras provided images with 3 bands, hereafter defined for the unmodified (VIS) camera R (red), G (green) and B (blue) bands and for the modified (NIR) camera NIR-R, NIR-G, and NIR-B bands. Four sticks marked each of the corners of the field-plot and allowed for geometric correction. A Macbeth colour checker was placed next to the plot, within the image swath at the image acquisition time. All the images were recorded at about two meters height above the plots from a central nadir position to minimize distortions as much as possible.

Image processing and analyses

Our methodology to assess the effect of colour calibration on classifier transferability for Arctic vegetation change studies across the sampled altitudinal gradient had four main parts (Fig. 2): (i) data preparation including geometric correction, (ii) colour calibration, (iii) defining and extracting reference data, and (iv) image classification and accuracy assessment in a multi-image classification framework.

Data preparation and geometric correction

All 108 VIS and 108 NIR images were geo-referenced two times (Fig. 1): (1) relative to the four plot corners (80×80cm) and extracting the plot area image with 2500×2500 pixels, and (2) relative to the cross marks in the corners of the Macbeth colour checker and extracting the *colour checker image* with 570×860 pixels . Both extractions resulted in a ~0.3 mm pixel resolution on the ground. We did the georeferencing in *ArcGIS 10.3.1* (ESRI Redlands, California, USA).

Colour calibration

We implemented colour calibration based on the Macbeth colour checker and 11 different calibration models (Fig. 2). The Macbeth colour checker has been used in several studies using close range photography (McCamy et al. 1976; Jackowski et al. 1997; Kreslin et al. 2014). Reference reflectance values of the 24 colours on the colour checker were based on Ritchie et al. (2008). We extracted the 24 colour values (DN: digital numbers) of each plot area image in six spectral bands (R, G, B, NIR-R, NIR-G, NIR-B), using *Python 2.7.8* (Python Software Foundation, Beaverton, USA). To assess the effect of the Bayer filter on NIR images, we compared the standard deviations of the DN for the NIR-R, NIR-G, and NIR-B bands within sampling grids (i.e., each of the 24 colours). NIR-R had the lowest standard deviation (Online Resource 1) and thus provided the most consistent NIR spectral information. Therefore, only NIR-R was calibrated and used in the classification.

First, we did colour calibration on RGB images by implementing 11 different calibration models based on first (1^{st}) and second (2^{nd}) order polynomial regression models, an exponential model, logistic regressions, and the DT-based transformation (Table 1). All calibration models were implemented in *R 3.3.5* (R Development Core Team). Based on DN from the colour checker image, represented as a vector $V: (R_i, G_i, B_i)$ $(i = 1, 2, \dots, 24)$, and the corresponding reference reflectance values as given in Ritchie et al. (2008), represented as *sRGB*, with $S : (sR_i, sG_i, sB_i)$ $(i = 1, 2, \dots, 24)$, the parameters (a) of the calibration algorithms were defined. For example, a simple linear transformation (i.e., 1^{st} order polynomial transformation x: [$R, G, B,$ 1]), was formulated as follows:

$$sR_i = a_{11}R_i + a_{12}G_i + a_{13}B_i + a_{14}$$

$$sG_i = a_{21}R_i + a_{22}G_i + a_{23}B_i + a_{24}$$

$$sB_i = a_{31}R_i + a_{32}G_i + a_{33}B_i + a_{34} \quad (1)$$

Where:	
sR, sG, sB	Reflectance values from Ritchie et al. (2008)
R, G, B	Digital numbers extracted from the in-image colour checker
(i = 1, 2…, N)	Fields on the Macbeth colour checker (N = 24)

We implemented three 1st order polynomial transformations (M1 – M3, Table 1). Four 2nd order polynomial transforms (M4 – M7) were used to increase the transformation accuracy. For M2 and M4 only corresponding bands were used for the calibration (Table 1). M5 was highly parametrized and resulted in overfitting and false colours occurred. Furthermore, we implemented two logistic regression models with sigmoid curves with exponential growth (M8 and M9) and an exponential regression model (M10) (Table 1). Finally, DT was implemented in R by converting the MATLAB code of Kreslin et al. (2014).

We first applied all the colour calibration algorithms on RGB images to compare the values of the colour checker from the calibrated images with the reference values and with the values from the non-calibrated images. Four calibration methods (M2, M8, M9 and DT) resulted in colour values closer to the reference values compared to the non-calibrated images and did not change the natural colour space (Online Resource 2). We therefore selected these four calibration models and calibrated the NIR band based on the $NIR_{800\text{-}900}$ reflectance values from Ritchie et al. (2008) (Table 1). Since the DT model cannot be applied to only one band, we here used the sNIR values obtained with M8.

Images calibrated with these four models were taken forward for the classification. Moreover, since the spectral reflectance is almost constant across wavelengths for the grey colours compared to other colours (see Fig.4 in Berns et al. 2005), we also calibrated all images with only the grey colours of the colour checker with M2, M8 and M9 (hereafter M2_g, M8_g and M9_g). Hence, in total eight image data sets were used for classification, including the non-calibrated images.

Reference data preparation

For each plot, we created reference polygons for the following four cover classes: shrub, other vegetation cover (i.e., graminoids, pteridophytes, lichens, bryophytes, and herbs), other cover classes (i.e., markings), and ground (including bare ground and stones). All reference polygons were drawn by the same person to reduce observer bias. We extracted spectral values per pixel of each image within the defined polygons and used them as the reference dataset. On average, 105,328 pixels of 6.25MP = 1.7% (min = 0.5%, max = 9%) were selected per image, and on average 41,095 (39%) of these pixels were the shrub class. The process was automated in R 3.3.5, and was done for each of the 108 plots in each of the eight calibrated image sets. The large reference data set provides a unique opportunity for assessing classification transferability for Arctic vegetation change studies using images stratified randomly across almost 500 altitudinal meters.

Multi-image classification

The four cover classes were classified using random forest classification with four (R, G, B and NIR) parameters. We implemented pixel-based classification, as the aim was to investigate the effect of the colour calibration on spectral classification transferability and not to obtain the most accurate classification of each image. We used the random forest classifier because of its robustness (Rodriguez-Galiano et al. 2012).

The following steps were taken to investigate the classifier's transferability, i.e., to what extent the classifier can be applied to other images beyond a training subset, and to find the optimum size of reference data regarding classification transferability. We trained the random forest classifiers on randomly selected portions of images (5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60 and 70%, hereafter training subset) and subsequently applied the classifier on the remaining images (remaining subset, Fig. 2). Reference data of the images in training subsets were merged and used to train the classifier. The remaining subsets included all the images in each image set except the images of training subsets. Moreover, for training portions ≤50%, we selected a random subset with the equal number of images as in the specific training subset. This procedure was implemented to examine if the overall trend of accuracy could be captured in smaller datasets. We repeated the classification 10 times for each portion, to assess the classification transferability regardless of the specifications of training subsets (Fig. 2). We repeated the whole procedure for the non-calibrated images as well as the seven selected calibration methods (M2, M2_g, M8, M8_g, M9, M9_g, DT) to assess the effect of each calibration method on the transferability of the classifier.

To assess the degree of transferability, we calculated overall accuracy (OA), the Kappa coefficient, and the user's (UA) and the producer's (PA) accuracies per class for each classified image (Foody 2002). Classification accuracies of the remaining subsets compared to the classification accuracies of training subsets were used to assess the classifier's transferability. As a benchmark for classification accuracy, we also implemented single image classification on each non-calibrated image. Accuracy of the single image classification was preformed based on training and testing data obtained from the same image to which the classifier was applied. We assessed if calibration methods significantly improved transferability of the classifier by performing Dunnett's test using the *'DescTools'* package with the non-calibrated image set as control group (Signorell et al. 2017).

Results

The average (±standard deviation) OA and kappa for single image classification of the four ground cover classes across all non-calibrated images was 93% (±3%) and 88% (±6%), respectively. In the multi-image classification framework, OA on the training subsets decreased to 78% (±8%) with increasing size of the training subset (Fig. 4). OA on the remaining subsets increased with increasing subset size and reached average OA of 68% (±12%) and 72% (±11%) when respectively 20% and 70% of the non-calibrated images were used for training (Fig. 4, 5). A similar trend was observed on the testing subsets. In general, classification OAs are levelling off with a training set of ca. 20%, with a similar trend for all the image sets (calibrated and non-calibrated) (Fig. 4, 5). OA variations in the remaining subsets decreased with increasing sizes of the training subsets; with the highest variation for portion=5%. For shrub cover average UA and PA of 62% and 68% where reached when 20% of the non-calibrated images were used for training . For the non-shrub vegetation class these values were 61% and 60%, respectively, while they were only 28% and 39% for the ground cover class.

Training subset OA based on all the colour calibrated image-sets were higher than OA on the non-calibrated image-set (1-4%, Fig. 4). However, colour calibration only slightly increased transferability of the classifier for small subset sizes (≤20%, Fig. 4) and only M2_g and M8_g had significant positive effect on classification transferability (Fig. 5). M9 and M9_g had a significant negative effect on classification transferability (Fig. 5). These trends were also captured on testing subsets. Similarly, in relation to shrub cover classification, M2_g and M8_g significantly increased shrub cover class UA for remaining subsets when assessed across all portions (Fig. 6). However, all other calibration models, except M2_g and DT, decreased shrub cover class producer's accuracies significantly (Fig. 6). When 20% of the images were used for calibration with M2_g, average UA for the shrub class improved with 1.9% (relative to non-calibrated data), while only PA improved (2.6%) for the non-shrub class. Both UA (2.1%) and PA (7.9%) increased for the ground class. Similar results were obtained with M8_g and a subset size of 20%; here PA for the ground class increased by 9%.

Discussion

In ecological field-based studies, researchers estimate vegetation cover visually or with point framing, for analysis and monitoring of vegetation change at fine scale (Luscier et al. 2006; Liu and Treitz 2016). This method, however, is time consuming, might be biased, and provides only limited data for upscaling (Neeser et al. 2000; Kercher et al. 2003; Rose et al. 2015; Tichy 2016). Mitigating these challenges was the main motivation to use UHR imagery for vegetation cover estimation. However, usage of UHR imagery requires classification of a large number of images taken at different locations and times (Gehler and Nowozin 2009; Cimpoi et al. 2014). The main challenge is increased intra-class variability (e.g., due to varying illumination condition, various vegetation characteristics) which might reduce classification accuracy and makes selection of a representative reference dataset difficult (Gehler and Nowozin 2009). We assessed the value of UHR imagery for vegetation cover estimation by testing the classification transferability during a multi-image classification of images taken stratified-random across almost 500 altitudinal meters in Western Greenland. Our findings show reasonable transferability of Arctic shrub cover classification, with average overall accuracy of 68% ±12% on independent images when 20% of the images were used to parametrize the classifier. This relatively good transferability based only on spectral data is promising. It illustrates, that monitoring of vegetation cover with UHR imagery is achievable, not only for images taken under varying field conditions, but also for images covering the range of vegetation class variation, which is expected under future climate change.

The training subset size affected classification transferability. The aim was to find the smallest possible random subset of images, assuring reliable training of the classifier and minimizing the time spend on creating reference data. Reference data were created by delimitation of polygons for each of the targeted vegetation classes. As expected, increasing the training subset size improved the transferability of the classifier (i.e., classification accuracy on remaining subsets) (Fig. 4). By using more images for training, classification transferability increases as the classifier recognizes more variation of each class due to different species, shadows, age, and health as well as (mitigated) illumination effects. However, due to increased intra-class variation the overall accuracies of classification on training subsets decreased with the increasing number of training images (Fig. 4). Classification accuracy tends to level off when about 20% of all images are included in the training set and this trend is similar for all image sets (calibrated and non-calibrated). Therefore, we concluded that about one fifth of the images would possibly be the optimum size for a training subset to provide an image classifier that is transferrable to all the images.

All calibration models improved classification accuracies on the training subsets (Fig. 4). Even though colour calibration slightly increased transferability of the classifier for small subset sizes (≤20%, Fig. 4), only M2_g and M8_g significantly improved transferability of the classifier for small subsets (portions ≤15%, Fig. 5). However, for bigger sizes of training subsets (>20%), overall accuracies of the calibrated image-sets and non-calibrated image-set were similar. This shows that the classifier possibly captured most of the variation of illumination effects when a random subset of at least 20% of the images was used to train the models. Transferability of the classification for the shrub, non-shrub, and ground class increased more with calibration models that only included the grey scales of the calibration target. Transferability of each of the three classes increased 2-9% with these colour calibrations when only 20% of the images were used for training. This underlines that colour calibration is important for maximizing transferability when small portions of the data are used for training, but also that its importance depends on the cover class of interest.

Images calibrated with M9_g had the highest classification accuracy on training subsets, compared to the other image-sets (Fig. 4). However, M9 and M9_g had the lowest classification accuracy on the remaining subsets (Fig. 4, 5). This behaviour might be explained by M9 models having a lower dynamic range compared to other models, due to the model specification. In addition, although the DT calibration model enhanced the images best for visualization purposes (Kreslin et al. 2014), it did not improve the transferability of the pixel-based random forest classifier. These results show that different calibration methods could be useful for different applications. Importantly the increase in accuracy (1-4%) on our training data documents that colour calibration is important when classification is performed on one or few images were reference data is available for all images.

Colour calibration is one approach to mitigate intra-class variation in reflectance due to illumination differences among images (Finlayson and Trezzi 2004; Gehler et al. 2008; Wang et al. 2013). Another approach is using spatial signatures (like texture and shape); these measures are less sensitive to illumination variation (Gehler and Nowozin 2009; Johansen et al. 2014). In recent studies using high-resolution imagery, object-based classification methods provided more accurate results than pixel-based classification methods (Whiteside et al. 2011). Because we aimed at a fully objective classification approach, which minimized user decisions and optimized time efficiency, we here applied a pixel-based classification method. However, integrating texture measures in a pixel-based classification is likely to improve the classification accuracy.

Mapping and monitoring changes in Arctic shrub cover is crucial for understanding the spatial magnitude of the potential biodiversity and ecosystem consequences of climate change in the Arctic. Efficiently obtaining fine-scale ground truthing information of vegetation cover is especially important in the Arctic due to the short field season and the logistical challenges related to cover large areas during one field campaign. Classification of UHR images show promise for providing comparable estimates of vegetation cover across space and time. Moreover, such remotely sensed data can improve, add and speed up the traditional field-based data collection (Neeser et al. 2000; Luscier et al. 2006; Lengyel et al. 2008; Fischer et al. 2012; Tichy 2016) and provide fine-scale ground truth data which in combination with satellite-based remote sensing will enable upscaling of fine scale observations across larger areas (Liu and Treitz 2016).

Conclusion

The goal of this study was to investigate the effect of training data size and colour calibration on transferability of a pixel-based classification. Here, for shrub cover estimation, a simple linear model (M2) based on the grey series of the calibration target worked better than the other models. A random selection of 20% of all images was the optimal size for the training subset. The transferability of the classifier with an overall classification accuracy of about 70% is promising for the use of UHR imagery to assist field-based ecological studies. These results are useful for automating Arctic vegetation monitoring. Further improvement of classification accuracy might be reached by including spatial signatures in the classification.

Compliance with ethical standards

Conflict of interest

The authors declare that they have no conflict of interest.

Figure legends

Figure 1 Study area, sampling design and example of geometric correction. (a) Location of the study area in the inner Nuuk Fiord, western Greenland. Vegetation classification based on Karami et al. (2018). (b) Distribution of the 108 vegetation plots across altitudinal isoclines within the study area. The vegetation plots are distributed in groups of six plots (inlet). The distance between plot groups was 500 m and distance between each of the six plots 10 m. (c) From each image the plot area (80×80 cm) and the colour checker area were extracted as separate images and (e) geometrically corrected (see text for details).

Figure 2 Framework of the applied data processing steps. VIS: visible light spectrum, NIR: near-infrared light spectrum, and R: Red, G: Green, B: Blue bands of the VIS spectrum. OA: overall classification accuracy, UA: user accuracy, and PA: producer accuracy.

Figure 3 Two examples of single image classification results with the pixel-based random forest classifier on non-calibrated data: Left: NIR-RG (near infrared, red, green), and Center: RGB (red, green and blue) images of the 80x80 geometrically corrected plots. Right: Classified images for shrub, non-shrub, ground, and other cover classes,

Figure 4 Relationship between overall accuracy and the proportion of data used for training. Loess-smoothed overall accuracies (mean ± standard deviation) for training and remaining subsets is plotted against the portion of images used for training the classifier for different calibration models; Delaunay triangulation (DT), 1st order polynomial (linear, M2) and exponential (M8, M9) with all the colours from the Macbeth colour checker or only with grey series (M2_g, M8_g, M9_g).

Figure 5 Effect of the calibration models on classification transferability of all ground classes. Each block shows classification overall accuracies (OA) of different portions of images used for training the classifier. Images were either non-calibrated data (RD) or calibrated with different implementations of calibration models: Delaunay triangulation (DT), 1st order polynomial (linear, M2) and exponential (M8, M9) with all the colours from the Macbeth colour checker or only with grey series (M2_g, M8_g, M9_g). We assessed if calibration methods significantly improved transferability of the classifier by performing a Dunnett's test with the non-calibrated image set as control group.

Figure 6 Effect of the calibration models on classification transferability of the shrub class. Users' and producers' accuracies (UA and PA) computed as averages across all portions of the training images. Images were either non-calibrated data (RD) or calibrated with different implementations of calibration models: Delaunay triangulation (DT), 1st order polynomial (linear, M2) and exponential (M8, M9) with all the colours from the Macbeth colour checker or only with grey series (M2_g, M8_g, M9_g). We assessed if calibration methods significantly improved transferability of the classifier by performing a Dunnett's test with the non-calibrated image set as control group.

References

Berns RS, Taplin LA, Nezamabadi M, Mohammadi M, Zhao Y (2005) Spectral imaging using a commercial colour-filter array digital camera. ICOM Committee for Conservation. Triennial meeting, 14th, The Hague, Netherlands, pp. 743-750

Bold KC, Wood F, Edwards PJ, Williard KWJ, Schoonover JE (2010) Using photographic image analysis to assess ground cover: a case study of forest road cutbanks. Environ Monit Assess 163:685-698

Booth DT, Cox SE (2008) Image-based monitoring to measure ecological change in rangeland. Front Ecol Environ 6:185-190

Bricher PK (2012) Methods for mapping the tundra vegetation of sub-Antarctic Macquarie Island. Dissertation, School of Geography and Environmental Studies University of Tasmania

Chen ZH, Chen WJ, Leblanc SG, Henry GHR (2010) Digital Photograph Analysis for Measuring Percent Plant Cover in the Arctic. Arctic 63: 315-326

Cimpoi M, Maji S, Kokkinos I, Mohamed S, Vedaldi A (2014) Describing Textures in the Wild. 2014 IEEE Conference on Computer Vision and Pattern Recognition. https://doi.org/10.1109/CVPR.2014.461

Engler R, Randin CF, Thuiller W, Dullinger S, Zimmermann NE, Araujo MB, Pearman PB, Le Lay G, Piedallu C, Albert CH, Choler P, Coldea G, De Lamo X, Dirnbock T, Gegout JC, Gomez-Garcia D, Grytnes JA, Heegaard E, Hoistad F, Nogues-Bravo D, Normand S., Puscas M, Sebastia MT, Stanisci A, Theurillat JP, Trivedi MR, Vittoz P, Guisan A (2011). 21st century climate change threatens mountain flora unequally across Europe. Glob Chang Biol 17:2330-2341

Elmendorf SC, Henry GHR, Hollister RD, Bjork RG, Boulanger-Lapointe N, Cooper EJ, Cornelissen JHC, Day TA, Dorrepaal E, Elumeeva TG, Gill M, Gould WA, Harte J, Hik DS, Hofgaard A, Johnson DR, Johnstone JF, Jonsdottir IS, Jorgenson JC, Klanderud K, Klein JA, Koh S, Kudo G, Lara M, Levesque E, Magnusson B, May JL, Mercado-Diaz JA, Michelsen A, Molau U, Myers-Smith IH, Oberbauer SF, Onipchenko VG, Rixen C. Schmidt NM, Shaver GR, Spasojevic MJ, Porhallsdottir PE, Tolvanen A, Troxler T, Tweedie CE, Villareal S, Wahren CH, Walker X, Webber PJ, Welker JM, Wipf S (2012) Plot-scale evidence of tundra vegetation change and links to recent summer warming. Nat Clim Change 2:453-457

Finlayson GD, Trezzi E (2004) Shades of gray and colour constancy. The Twelfth Color Imaging Conference: Color Science and Engineering Systems, Technologies, Applications, CIC 2004, November 9, 2004, Scottsdale, Arizona, USA, pp. 37-41

Fischer T, Veste M, Eisele A, Bens O, Spyra W, Huttl RF (2012) Small scale spatial heterogeneity of Normalized Difference Vegetation Indices (NDVIs) and hot spots of photosynthesis in biological soil crusts. Flora 207:159-167

Foody GM (2002) Status of land cover classification accuracy assessment. Remote Sensing Environ 80:185-201

Gehler, P. & S. Nowozin (2009) On Feature Combination for Multiclass Object Classification. 2009 IEEE 12th International Conference on Computer Vision. https://doi.org/10.1109/ICCV.2009.5459169

Gehler PV, Rother C, Blake A, Minka T, Sharp T (2008) Bayesian color constancy revisited. 2008 IEEE Conference on Computer Vision and Pattern Recognition. https://doi.org/10.1109/CVPR.2008.4587765

Guay KC, Beck PSA, Goetz SJ (2015) Long-Term Arctic Growing Season NDVI Trends from GIMMS 3g, 1982-2012. ORNL DAAC, Oak Ridge, Tennessee, USA. http://dx.doi.org/10.3334/ORNLDAAC/1275

Hansen RR, Hansen OL, Bowden JJ, Treier UA, Normand S, Høye T (2016) Meter scale variation in shrub dominance and soil moisture structure Arctic arthropod communities. PeerJ 4:e2224

He KS, Bradley BA, Cord AF, Rocchini D, Tuanmu MN, Schmidtlein S, Turner W, Wegmann M, Pettorelli N (2015) Will remote sensing shape the next generation of species distribution models? Remote Sens Ecol Conserv 1:4-18

Hyman JM (2010). Imagers as Sensors: Using Visible Light Images to Measure Natural Phenomena Dissertation, University of California, Los Angeles

Jackowski M, Goshtasby A, Bines S, Roseman D, Yu C (1997) Correcting the geometry and color of digital images. IEEE Trans Pattern Anal Mach Intell 19:1152-1158

Johansen K, Sohlbach M, Sullivan B, Stringer S, Peasley D, Phinn S (2014) Mapping Banana Plants from High Spatial Resolution Orthophotos to Facilitate Plant Health Assessment. *Remote Sens* 6:8261-8286.

Karami M, Westergaard-Nielsen A, Normand S, Treier UA, Elberling B, and Hansen BU (2018) A phenology-based approach to the classification of Arctic tundra ecosystems in Greenland. ISPRS J Photogramm Remote Sens 146:518–529

Kercher SM, Frieswyk CB, Zedler JB (2003) Effects of sampling teams and estimation methods on the assessment of plant cover. J Veg Sci 14: 899-906

Kreslin R, Calvo PM, Corzo LG, Peer P (2014) Linear Chromatic Adaptation Transform Based on Delaunay Triangulation. Math Probl Eng. http://dx.doi.org/10.1155/2014/760123

Lengyel S, Deri E, Varga Z, Horvath R, Tothmeresz B, Henry PY, Kobler A, Kutnar L, Babij V, Seliskar A, Christia C, Papastergiadou E, Gruber B, Henle K (2008) Habitat monitoring in Europe: a description of current practices. Biodivers Conserv 17:3327-3339

Liu NF, Treitz P (2016) Modelling high arctic percent vegetation cover using field digital images and high resolution satellite data. Int J Appl Earth Obs 52:445-456

Luscier JD, Thompson WL, Wilson JM, Gorham BE, Dragut LD (2006) Using digital photographs and object-based image analysis to estimate percent ground cover in vegetation plots. Front Ecol Environ 4: 408-413

Masson-Delmotte V, Schulz M, Abe-Ouchi A, Beer J, Ganopolski A, González Rouco JF, Jansen E, Lambeck K, Luterbacher J, Naish T, Osborn T, Otto-Bliesner B, Quinn T, Ramesh R, Rojas M, Shao X, Timmermann A (2013) Information from Paleoclimate Archives. In: Stocker TF, Qin D, Plattner GK, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds) Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, pp. 383-464, doi:10.1017/CBO9781107415324.013

McCamy CS, Marcus H, Davidson JG (1976) A Color-Rendition Chart. J Appl Photogr Eng 2: 95-99

Menesatti P, Angelini C, Pallottino F, Antonucci F, Aguzzi J, Costa C (2012) RGB Color Calibration for Quantitative Image Analysis: The "3D Thin-Plate Spline" Warping Approach. Sensors 12: 7063-7079

Morueta-Holme N, Engemann K, Sandoval-Acuna P, Jonas JD, Segnitz RM, Svenning JC (2015) Strong upslope shifts in Chimborazo's vegetation over two centuries since Humboldt. Proc Natl Acad Sci USA 112: 12741-12745

Myers-Smith IH, Hallinger M, Blok D, Sass-Klaassen U, Rayback S .A, Weijers S, Trant AJ, Tape KD, Naito AT, Wipf S, Rixen C, Dawes MA, Wheeler JA, Buchwal A, Baittinger C, Macias-Fauria M, Forbes BC, Lévesque E, Boulanger-Lapointe N, Beil I, Ravolainen V & Wilmking M (2015) Methods for measuring arctic and alpine shrub growth: a review. Ear Sci Rev 140: 1-13

Myers-Smith IH, Elmendorf SC, Beck PSA, Wilmking M, Hallinger M, Blok D, Tape KD, Rayback SA, Macias-Fauria M, Forbes BC, Speed JDM, Boulanger-Lapointe N, Rixen C, Levesque E, Schmidt NM, Baittinger C, Trant AJ, Hermanutz L, Collier LS, Dawes MA, Lantz TC, Weijers S, Jorgensen RH, Buchwal A, Buras A, Naito AT, Ravolainen V, Schaepman-Strub G, Wheeler JA, Wipf S, Guay KC, Hik DS, Vellend M (2015) Climate sensitivity of shrub growth across the tundra biome. Nat Clim Change 5:887-891

Myers-Smith IH, Forbes BC, Wilmking M, Hallinger M, Lantz T, Blok D, Tape KD, Macias-Fauria M, Sass-Klaassen U, Levesque E, Boudreau S, Ropars P, Hermanutz L, Trant A, Collier LS, Weijers S, Rozema J, Rayback SA, Schmidt NM, Schaepman-Strub G, Wipf S, Rixen C, Menard CB, Venn S, Goetz S, Andreu-Hayles L, Elmendorf S, Ravolainen V, Welker J, Grogan P, Epstein HE, Hik DS (2011) Shrub expansion in tundra ecosystems: dynamics, impacts and research priorities. Environ. Res Lett 6:045509 (15pp)

Nabe-Nielsen J, Normand S, Hui FKC, Stewart L, Bay C, Nabe-Nielsen LI, Schmidt NM (accepted in sep. 2017) Plant community composition and species richness in the High Arctic tundra: from the present to the future, Ecol Evol 7: 10233-10242

Neeser C, Martin AR, Juroszek P Mortensen DA (2000) A comparison of visual and photographic estimates of weed biomass and weed control. Weed Technol 14:586-590

Neumann C, Weiss G, Schmidtlein S, Itzerott S, Lausch A, Doktor D, Brell M (2015) Gradient-Based Assessment of Habitat Quality for Spectral Ecosystem Monitoring. Remote Sens 7:2871-2898

Nielsen SS, von Arx G, Damgaard CF, Abermann J, Buchwal A, Büntgen U, Treier UA., Barfod AS, Normand S (2017) Xylem anatomical trait variability provides insight on the climate-growth relationship of Betula nana in western Greenland. Arct Antarct Alp Res49:359-371

Normand S, Randin C, Ohlemüller R, Bay C, Høye, TT, Kjær ED, Körner C, Lischke H, Maiorano L, Paulsen J, Pearman PB, Psomas A, Treier UA, Zimmermann NE, Svenning JC (2013) A greener Greenland? Climatic potential and long-term constraints on future expansions of trees and shrubs. Phil Trans R Soc B 368:20120479

Pearson RG, Phillips SJ, Loranty MM, Beck PSA, Damoulas T, Knight SJ, Goetz SJ (2013) Shifts in Arctic vegetation and associated feedbacks under climate change. Nat Clim Change 3:673-677

Post E, Forchhammer MC, Bret-Harte MS, Callaghan TV, Christensen TR, Elberling B, Fox AD, Gilg O, Hik DS, Høye TT, Ims RA, Jeppesen E, Klein DR, Madsen J, McGuire AD, Rysgaard S, Schindler DE, Stirling I, Tamstorf MP, Tyler NJ, van der Wal R, Welker J, Wookey PA, Schmidt NM, Aastrup P (2009) Ecological Dynamics Across the Arctic Associated with Recent Climate Change. Science 325:1355-1358

Quevedo RA, Aguilera JM, Pedreschi F (2010) Color of Salmon Fillets By Computer Vision and Sensory Panel. Food Bioprocess Tech 3:637-643

Ritchie GL, Sullivan DG, Perry CD, Hook JE, Bednarz CW (2008) Preparation of a Low-Cost Digital Camera System for Remote Sensing. Appl Eng Agric 24:885-894

Rodriguez-Galiano VF, Ghimire B, Rogan J, Chica-Olmo M, Rigol-Sanchez JP (2012) An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS J Photogramm Remote Sens 67:93-104

Rose RA, Byler D, Eastman JR, Fleishman E, Geller G, Goetz S, Guild L, Hamilton H, Hansen M, Headley R, Hewson J, Horning N, Kaplin BA, Laporte N, Leidner A, Leinagruber P, Morisette J, Musinsky J, Pintea L, Prados A, Radeloff VC, Rowen M, Saatchi S, Schil S, Tabor K, Turner W, Vodacek A, Vogelnaann J, Wegmann M, Wilkie D (2015) Ten ways remote sensing can contribute to conservation. Conserv Biol 29:350-359.

Signorell A (2017). DescTools: Tools for descriptive statistics. R package version 0.99.21 (2017-06-29). Retrieved from https://cran.r-project.org/package=DescTools

Tape K, Sturm M, Racine C (2006) The evidence for shrub expansion in Northern Alaska and the Pan-Arctic. Glob Chang Biol 12:686-702

Tichy L (2016) Field test of canopy cover estimation by hemispherical photographs taken with a smartphone. J Veg Sci 27:427-435

Villafuerte R, Negro JJ (1998) Digital imaging for colour measurement in ecological research. Ecol Lett 1: 151-154

Wang XZ, Zhang B, Guo ZH, Zhang D (2013) Facial image medical analysis system using quantitative chromatic feature. Expert Syst Appl 40:3738-3746

Wang XZ, Zhang D (2010) An Optimized Tongue Image Color Correction Scheme. IEEE Trans Inf Technol Biomed 14:1355-1364

Whiteside TG, Boggs GS, Maier SW (2011). Comparing object-based and pixel-based classifications for mapping savannas. Int J Appl Earth Obs Geoinf 13:884-893

Zlinszky A, Heilmeier H, Balzter H, Czucz B, Pfeifer N (2015) Remote Sensing and GIS for Habitat Quality Monitoring: New Approaches and Future Research. Remote Sens 7:7987-7994