Cloud shadow and uneven illumination detection and correction using the U-net architecture in near-surface images of complex forest canopies

Nathan B Goncalves[1,2], Bruce W Nelson[3], Adriana Simonetti[3], and Scott C Stark[1,2]

[1] Michigan State University, Department of Forestry, East Lansing, MI, United States,
[2] Program in Ecology, Evolution, and Behavior, Michigan State University, East Lansing, MI, USA,
[3] INPA National Institute of Amazon Research, Manaus, Brazil

Keywords: UAV, deep learning, cloud shadow, tropical forest, multispectral, RGB

Abstract

In humid tropical regions, irregular illumination and cloud shadows can complicate near-surface optical remote sensing. This could lead to costly and repetitive surveys to maintain geographical and spectral consistency. This could have a significant impact on the regular monitoring of forest ecosystems. A novel correction method using deep learning is presented here to address the issue in high-resolution canopy images. Our method involves training a deep learning model on one or a few well-illuminated/homogeneous reference images augmented with artificially generated cloud shadows. This training enables the model to predict illumination and cloud shadow patterns in any image and ultimately mitigate these effects. Using images captured by multispectral and RGB cameras, we evaluated the method across multiple sensors and conditions. These included nadir-view images from two sensors mounted on a drone and tower-mounted RGB Phenocams. The technique effectively corrects uneven illumination in near-infrared and true-color RGB images, including non-forested areas. This improvement was also evident in more consistent normalized difference vegetation index (NDVI) patterns in areas affected by uneven illumination. By comparing corrected RGB images to the original in a binary classification task, we evaluated the method's accuracy and Kappa values. Our goal was to detect non-photosynthetic vegetation (NPV) in a mosaic. The overall accuracy and Kappa were both significantly improved in corrected images, with a 2.5% and 1.1% increase, respectively. Moreover, the method can be generalized across sensors and conditions. Further work should focus on refining the technique and exploring its applicability to satellite imagery and beyond.

1. Introduction

Partial cloud shadows and uneven illumination in drone-based optical remote sensing, especially in the humid tropics, often force expensive, repeated surveys, hindering stable ecosystem monitoring (Luo et al., 2020). This issue compromises the distinct advantage that near-surface sensors have over orbital sensors (Richardson 2013), especially considering how cloud cover significantly impacts the latter (Schroeder et al., 2008). The problem is particularly challenging for near-surface remote sensing, which often requires capturing overlapping images of the same area to construct orthomosaics that are created by stitching several individual images for posterior analysis. This could often cause distortions in orthomosaics, jeopardizing the geographic and spectral/color consistency of the studied objects. Additionally, this problem can also cause artifacts in several fields, including time series analysis and land use change.

Generally, the techniques to detect and correct shadows fall into three main categories: geometrical methods, property-based methods, and more recently machine learning-based methods (Kang et al., 2018). Here, we propose a simple method to address this problem in the forest canopy context using a machine learning-based solution. Our method employs a single or few well-illuminated images to train a deep learning model (using the U-net architecture) (Ronneberger et al., 2015) that detects uneven illumination within all other images that comprise an orthomosaic output. The goal is to detect and correct the partial opacity of cloud shadows and uneven illumination in complex forest canopies, possibly improving orthomosaics and reducing the need for data reacquisition.

To empirically test this concept, we pose the following questions: 1) Does this simple correction yield qualitatively better results than uncorrected images? 2) Will corrected images perform better in a simple classification task than uncorrected ones? 3) How general is the method—can a model trained on a single image from one sensor be used to correct images from a different sensor? We present a case study using drone-derived RGB images, a multispectral sensor, and digital Phenological cameras to evaluate the proposed method.

2. Methods

The method begins with a single or few well-illuminated reference images without apparent cloud shadows and with even illumination, resampled to 128 x 128 pixels. These can be RGB digital images or any grayscale band from a multispectral sensor. Artificial multiplicative cloud shadows are randomly created, featuring linear, circular, and diagonal shadow patterns with 1500 combinations of shadow position, size, and opacity (Fig. 1A) simulating different light conditions. The intensity of the illumination varies randomly within specified limits and is applied across the image to create a gradient effect (Fig. 1A). This gradient is then smoothed with a Gaussian blur to ensure and mimic a natural transition of light. These masks are multiplied to the resampled image, generating artificially shadowed or unevenly illuminated images. It is assumed that this will create invertible shadows or unevenly illuminated configurations that approximate real-world shadow and illumination patterns. We utilize these 1500 randomly created artificial shadow masks and images to train a deep learning model built upon the U-Net architecture, which is widely used for semantic segmentation tasks (Ronneberger et al., 2015). For the U-net we use 512 filters in both upsampling and downsampling paths, using an ADAM optimizer.

We used the free resources of the Google Colab platform and only 25 epochs https://github.com/Nathanborg/Cloud_shadow_correction/). In the deep learning training context, the artificially shadowed (unevenly illuminated) versions of the reference images act as training data, and the corresponding invertible masks serve as the targets. Operationally, the task is framed as a regression problem to predict the masks (both shadowed and over-illuminated areas). Since the model is trained on one or a few images with similar irradiance/illuminations conditions, it also functions as a mean correction, approximating the average values of the reference images. These features allow the model to detect and correct illumination in any real-world image by normalizing them to a reference image. It accurately identifies and adjusts large shaded areas, such as true cloud shadows, while ignoring individual crown shadows, thus effectively reducing the impact of cloud shadow interference and uneven illumination. Once the trained model predicts the best multiplicative shadow mask for each real-world image rescaled to 128 x 128 pixels, the shadow mask is resampled up to the original image shadow size. The correction is then applied by dividing the original image by the resized shadow mask. This process is performed on the individual images before creating an orthomosaic in the case of drone imagery.

To test this concept qualitatively and address our first question, having an idea of the method's effectiveness, we applied the method in two distinct contexts: first, to near-infrared images (NIR, from red-edge Micasense sensor) impacted by uneven illumination and cloud shadows, assessing the subsequent effects on the individual images and in orthomosaics derived from the corrected images (Fig. 1B); and second, to RGB images collected with a FLIR Duo Pro R camera (Fig. 1C). For the NIR image training, we visually identified a single image free of cloud shadows with even illumination from a flight conducted in December 2022 over Manaus, Brazil. The flight conditions were partially cloudy, a typical scenario in the Brazilian Amazon during the wet season, where atmospheric conditions often rapidly change, resulting in uneven illumination. To this single image we applied the 1500 artificial shadow patterns and a U-Net model was trained to predict the shadow masks. The training process takes about 8 minutes for 25 epochs for 3-channel images and about 6 minutes for single grayscale images in the T4 GPU on google Colab. The trained model is then used to correct all other images within the flight as explained above in the case of the grayscale near-infrared image. A similar procedure was also used to train the RGB bands of the FLIR Duo Pro R camera; however, in this case, a tensor (input_size = (128, 128, 3)) incorporating all three channels was used for training.

We further compared geographically matched multispectral orthomosaics, both uncorrected and corrected, in the context of a common problem in remote sensing: spectral unmixing. Spectral unmixing is a process used in remote sensing to decompose the pixel reflectance values in an image into fractions of constituent materials, referred to as endmembers (Ponzoni, 2019). We collected five pure endmembers of each shadow, wood and leaves in both corrected and uncorrected radiance images (Images were transformed in radiance using Micasense utils - https://micasense.github.io/). For this task, we utilized the NIR, Red, and the normalized difference vegetation index (NDVI) as inputs. Using the R package RStoolbox (Leutner et al., 2024) and the Multiple Endmember Spectral Mixture Analysis (MESMA) function we determined the percentage of each component per pixel. The unmixed pixels were then plotted in a false-color composite with red representing non-photosynthetic vegetation (NPV), green indicating vegetation, and blue denoting shadow.

To quantitatively test the method and address our second question in a real-world scenario, we analyzed a total of 9 images with cloud shadows and uneven illumination captured by the FLIR Duo Pro R RGB camera (Fig. 3). These images were examined before and after correction in a binary classification problem involving Non-Photosynthetic Vegetation (NPV) versus Green Vegetation and Shadow detection. This is a common problem in ecological and agricultural applications of near surface remote sensing. We analyzed the uneven illuminated images, applied our correction and created a mosaic with the corrected images, mirroring it with a mosaic of the uncorrected images to ensure that every pixel in both mosaics was matched. We mosaicked the 9 images using the same 1% linear stretching (Fig. 3 A, B). We then manually collected 500 pixels visually identified as NPV (Non-Photosynthetic Vegetation) and 500 pixels identified as green vegetation or shadows. This visual identification was done using the uncorrected mosaic to avoid any sort of bias with the person identifying the pixels. After that, the same mirrored pixels were retrieved from the corrected mosaic. We calculated two indices: the Green Chromatic Coordinate (GCC) (Eq. 1) as per Woebbecke et al. (1995), and the Wood index (WDX) (Eq. 2), to increase the band space utilized alongside the R, G, and B channels for training and testing the classifier model. For this purpose, we selected the Random Forest algorithm, a widely-used machine learning algorithm for classification tasks, particularly in tabular data (Liaw & Wiener 2002).

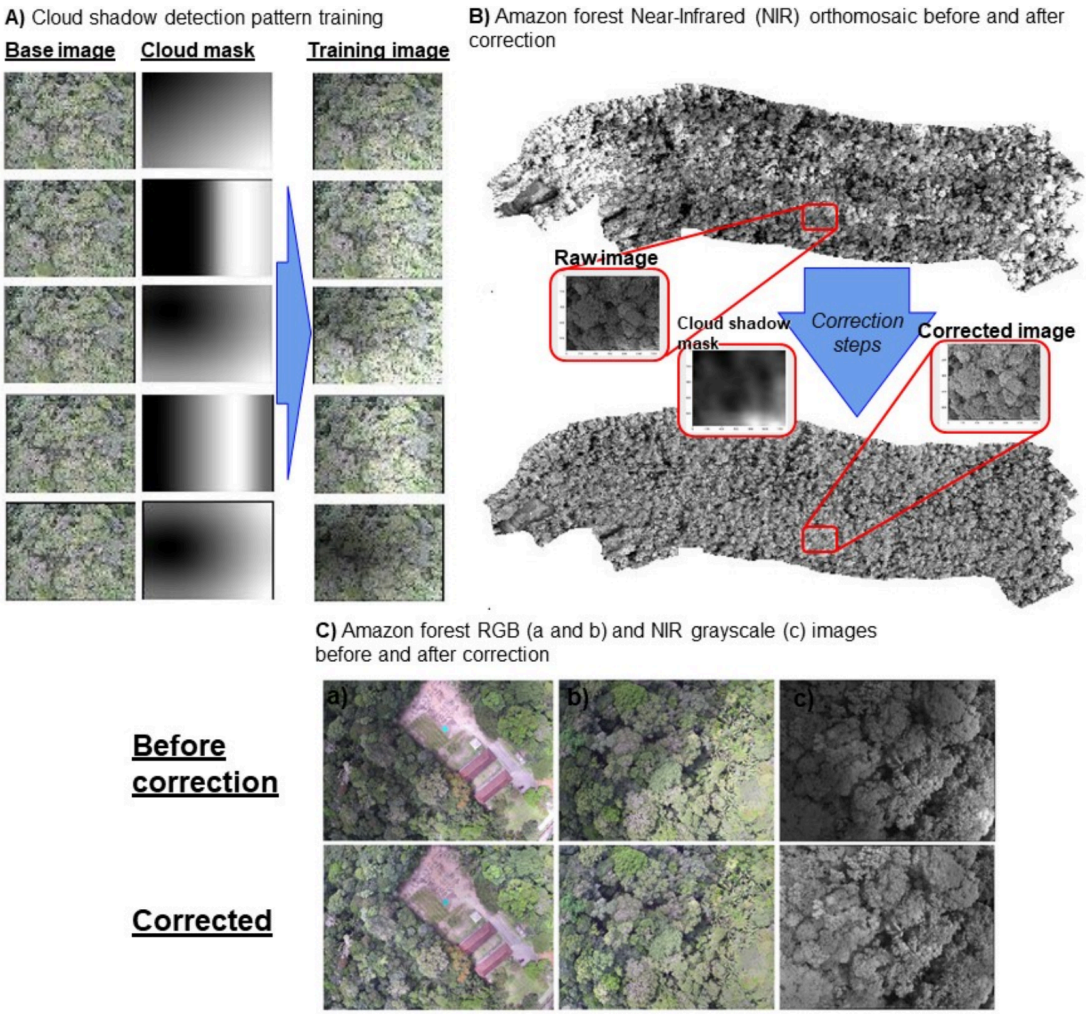

Figure 1. A) Schematic example of the training process, specifically, the creation of a large number of candidate artificial shadow shapes to match natural shadows that can then be removed by inversion of the best-matched predicted illumination/shadow, B) Near-infrared orthomosaic reflectance image before and after correction of its input images. C) Examples of individual images before and after correction.

$$GCC = G/(R + G + B) \quad (1)$$

$$WDX = (B + 2 * (R)) - G \quad (2)$$

Where G = Green,
B= Blue
R = Red channels

We then conducted an iterative process for the validation process, consisting of 200 iterations of the 1000 collected pixels for each mosaic. During each iteration, 70% of the data is used to train a random forest model, while the remaining 30% was used for testing, for both the corrected and uncorrected image mosaics. The iterative approach was applied to evaluate the generalizability of the classification model across 200 iterations. Each iteration involved a random stratified partitioning of the dataset into training (70%) and testing (30%) subsets using the createDataPartition function from the caret package in R (Khun 2008). For this study, the hyper parameters of the random forest were not specifically tuned and were left at their default settings. Predictive accuracy and the Kappa statistic were calculated for each iteration by applying the trained model to the corresponding testing subset and assessing the predictions against the actual target labels using a confusion matrix. This iterative testing framework aimed to robustly assess the model's performance across varying subsets of the data. We then compare the Kappa and accuracy values between the corrected vs uncorrected pixels to determine the impact of image correction on classification performance. We performed t tests to test differences in the averages accuracies and kappa. Additionally, we compared histograms between the corrected and uncorrected mosaics to visually assess the distribution of pixel values and the effect of the correction process.

To address our third question—whether a model trained with one sensor can be used to correct images from other sensors and different contexts, such as varying viewing angles—we utilized a model trained with near-nadir images from the Flir Duo Pro R RGB, collected by a drone, to correct Phenocam imagery captured by a Wingscape Pro model time lapse camera installed at the top of a 66-meter tower in Manaus, Brazil. This imagery had uneven illumination and a completely oblique view. We also qualitatively tested whether training with the Phenocam's own images resulted in a visually better correction.

3. Results

Our approach of training a U-Net model on a few or single images using artificial shadowing showed qualitatively excellent results in detecting and correcting uneven illumination caused by cloud shadows on unseen untrained images for both the NIR grayscale images and the RGB images. That is evident when comparing a NIR orthomosaic before and after correction (Fig. 1B), for those original images affected by uneven illumination. In addition, the model trained on a three-band RGB image could even be applied successfully to an area without forest (Fig. 1C). For the grayscale images from multispectral Micasense cameras, we also corrected both the NIR and Red bands separately to calculate the NDVI. We then compared how the widely used NDVI varies with and without correction in the orthomosaic context. Although NDVI is known to be less affected by Bidirectional Reflectance Distribution Function (BRDF) effects or shadow effects compared to other indices, differences in illumination can still cause several artifacts. However, the corrected images show homogeneous NDVI patterns(Fig.2).

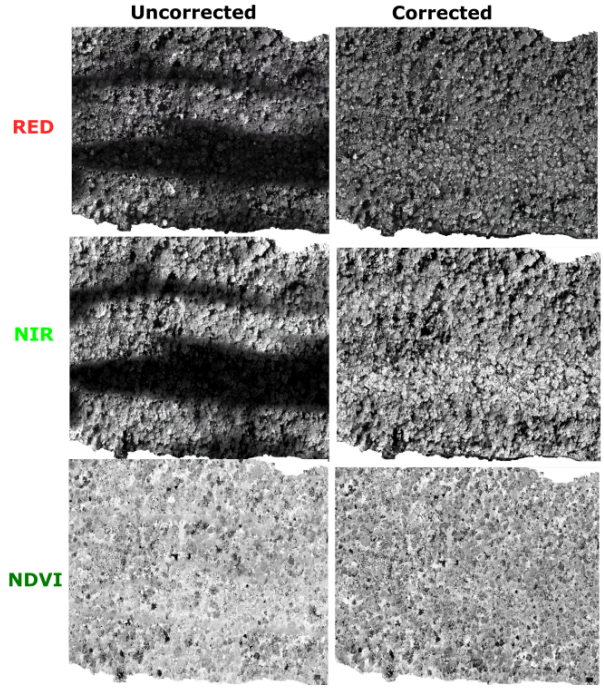

Figure 2. Corrected and uncorrected orthomosaics were composed of Red, NIR, and NDVI images from the Micasense RedEdge sensor collected in December 2022. The images are presented in radiance with a consistent 1% linear stretch applied. In this example, a single well-illuminated image was used to train the model for both Red and NIR corrections. Separate models were used to correct the Red and NIR images for each band individually.

The unmixing process (Fig. 3) demonstrates how uneven illumination can interfere with a simple numerical analysis where the objective is to retrieve the proportion of materials in pixels.

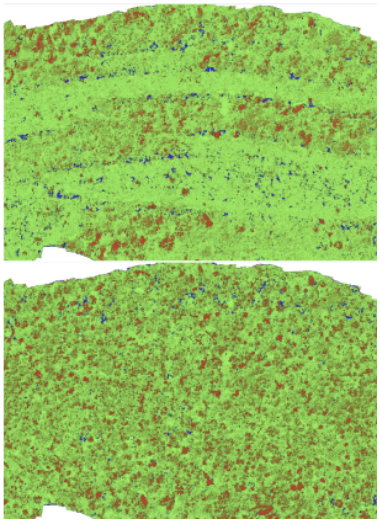

Figure 3. Corrected and uncorrected false color RGB (Red - NPV, Green - vegetation, Blue - shadow) unmixed pixels with the same 1% linear stretching. This figure illustrates how the corrected images could be a problem for uneven illumination in multispectral sensors in common unmixing problems.

Correcting these inconsistencies could also enhance the accuracy of orthomosaics, as shown in Figure 4, where duplicate objects are evident in images with uneven illumination.

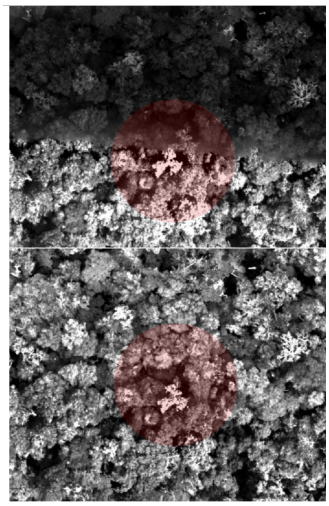

Figure 4. Corrected and uncorrected Red radiance images are shown in this figure, illustrating how distortion can be problematic under uneven

illumination in multispectral sensors, especially when stitching is necessary to create the orthomosaic. The red circle highlights how the leafless tree appears doubled in the uncorrected orthomosaic.

Furthermore, we demonstrate that a simple model, trained on a single Nadir RGB image from a FLIR Duo Pro R sensor, can effectively correct images even when those images have an oblique view such as in Phenocam images analyzed here (Fig. 5B), reducing or eliminating saturated pixels. However, a model trained on a single well-illuminated Phenocam image appeared to provide even better visual correction (Fig. 5C).

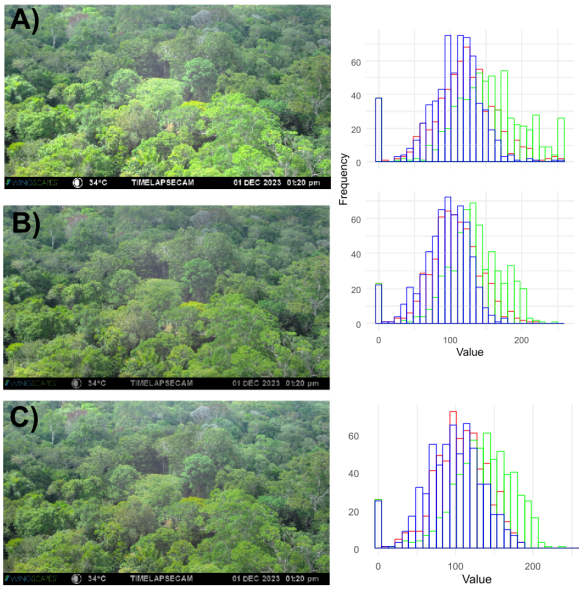

Figure 5. A) Uncorrected Phenocam RGB image. The associated histogram on the right side represents the distribution of RGB values in the uncorrected image. B) Phenocam (Wingscapes TimelapseCam Pro) RGB image corrected using a model trained with a single drone FLIR Pro Duo pro RGB camera. The histogram reflects the RGB value distribution post-correction, indicating improved color balance compared to the uncorrected image. C) RGB image corrected using a model trained with a single well-illuminated Phenocam image. The histogram exemplifies the RGB values distribution after this specific correction method.

For the quantitative analysis, corrected pixels yielded better results than the uncorrected pixels presenting significantly higher Kappa and accuracy values for this specific binary classification task of NPV detection in the canopy ($p < 0.01$, Fig. 6). On average kappa values were 2.5% higher and accuracy only a modest increase of 1.1% (Fig. 7).

Several tests using up to 9 images were also conducted to train the models for both RGB and multispectral contexts. These tests yielded slightly better results compared to analyses conducted using a single image. To simplify data presentation, we opted not to include these results.

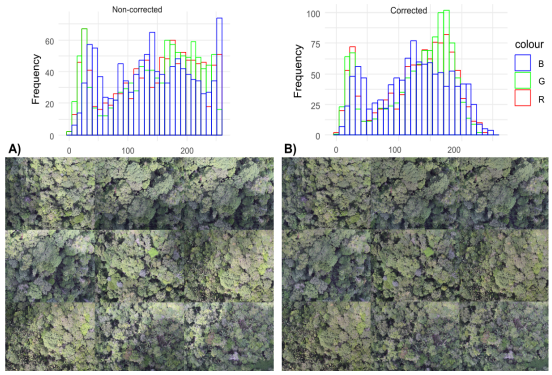

Figure 6. Comparison of uncorrected (A) vs corrected mosaic (B) and their corresponding histograms. In this example one RGB well illuminated image is used to train a single model. It is important to note that every image is corrected prior to mosaicing. All images have the same linear 1% stretching.

4. Discussion and Conclusion

We have addressed the three posed questions regarding our image correction method. Firstly, we have confirmed that our correction markedly improves the qualitative aspects of images compared to their uncorrected versions for both RGB digital cameras and multispectral cameras. Secondly, these enhanced images exhibit better performance in a basic binary classification task. Lastly, we have shown the possible broad applicability of the method; a model trained on a single image can attenuate uneven illumination in images from different sensors. These findings highlight the method's advantages, including increased classification accuracy, reduced distortions in orthomosaics, and the possibility of using a broadly applicable model across different imaging contexts.

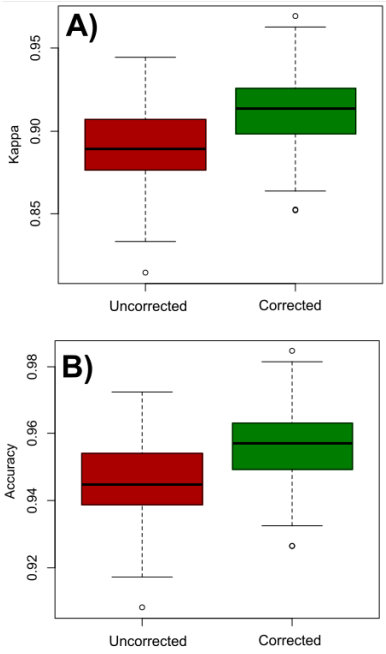

Figure 7. A) Boxplots of Kappa and B) accuracy values for the iteration of the random forest binary classification task for the RGB images mosaic (see Fig. 6).

In some cases, however, over- and under-correction has occurred (Fig. 2, NIR corrected orthomosaic), indicating areas for refinement. These less optimal results could be for several reasons: our processing constraints, we have not experimented with more complex architectures, limiting the models to a maximum of 512 filters for the encoder and decoder in the U-Net architecture, and using lower resolution training images, as we resampled the images to 128x128 pixels. Also, the simplicity of only using one image or a few images could make it difficult for the model to generalize and identify new patterns in untrained data. For instance, we have noticed that in the grayscale images context, trees with a very bright radiance are sometimes recognized as high illumination areas, thus causing an artifact or an over-correction (Not shown).

These artifacts prompted us to test the method for both RGB and multispectral contexts using up to 9 well-illuminated training images. These tests showed apparently better results compared to analyses performed with a single image avoiding the artifacts mentioned above. However, for the sake of simplicity in data presentation, we decided not to include those multiple-image results here. We also found that training RGB images using the three channels together produced more consistent colors than training with individual channels. However, we have not tested how the method deals with inputs of several bands in a multispectral context and should be further investigated.

Our findings also indicate that a model trained on one type of sensor, such as nadir drone imagery, can apparently correct oblique Phenocam images with uneven illumination, confirming the model’s versatility across different sensors and conditions. This illustrates the model's capacity to learn and adapt to various shadow patterns, "learning" the general underlying illumination patterns in complex canopies. Further exploration is needed to assess the method's impact on the spectral/color integrity of images in the context of more complex problems such as multiple-classes classification and tree segmentation.

We also believe this approach has the potential to extend to detecting and correcting cloud shadows or uneven illumination in satellite imagery. Another interesting application of the technique in satellite remote sensing. For example, could be empirical BRDF corrections. In this context, training images taken at certain irradiance levels and sun-sensor angles could be used to correct untrained images in other sun-sensor configurations and normalize them for further analysis. Despite the promising results from our qualitative and quantitative tests, further testing in various contexts is necessary. For example, it is important to determine whether incorporating a larger set of images or enhancing artificial shadow variability could improve the model's robustness.

6. References:

Kang, X., Huang, Y., Li, S., Lin, H., Benediktsson, J.A., 2018. Extended Random Walker for Shadow Detection in Very High Resolution Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*, 56(2). doi.org/10.1109/TGRS.2017.2755773

Kuhn, M., 2008. Building Predictive Models in R Using the caret Package. *Journal of Statistical Softwar*e, 28(5), 1–26. doi.org/10.18637/jss.v028.i05

Leutner, B., Horning, N., Schwalb-Willmann, J., Mueller, K., 2024. RStoolbox: *Remote Sensing Data Analysis*. R package version 1.0.0, CRAN.R-project.org/package=RStoolbox

Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. *R news*, 2(3), 18-22.

Luo, S., Li, H., Shen, H., 2020. Deeply supervised convolutional neural network for shadow detection based on a novel aerial shadow imagery dataset. *ISPRS Journal of Photogrammetry and Remote Sensing*, 167, 443-457.doi.org/10.1016/j.isprsjprs.2020.07.016

Ponzoni, F.J., 2019. Spectral Mixture for Remote Sensing.

Richardson, A.D., Klosterman, S., Toomey, M., 2013. Near-surface sensor-derived phenology. In Phenology: An integrative environmental science (pp. 413-430). Dordrecht: Springer Netherlands.

Schroeder, W., Csiszar, I., Morisette, J., 2008. Quantifying the impact of cloud obscuration on remote sensing of active fires in the Brazilian Amazon. *Remote Sensing of Environment*, 112(2), 456-470.doi.org/10.1016/j.rse.2007.05.004

Woebbecke, D.M., Meyer, G.E., Von Bargen, K., Mortensen, D.A., 1995. Color indices for weed identification under various soil, residue, and lighting conditions. Transactions of *ASAE*, 38(1), 259–269. doi.org/10.13031/2013.27838

Zhu, Z., Wang, S., Woodcock, C.E., 2015. Improvement and expansion of the Fmask algorithm: Cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images. Remote Sensing of Environment, 159, 269-277.doi.org/10.1016/j.rse.2014.12.014