Imagers as Sensors: Using Visible Light Images to Measure Natural Phenomena

Josh Hyman
Dissertation Proposal
UCLA Computer Science Department
josh@cs.ucla.edu

August 26, 2008

Abstract

There exist many natural phenomena where direct measurement is either impossible or extremely invasive. We propose using imagers as sensors by constructing a procedure that uses images to obtain approximate measurements of these phenomena. This procedure, composed of state-of-the-art computer vision, image processing, and statistical learning algorithms, will be evaluated in the context of a specific application and shown to be general through multiple instantiations. We show through application, that many of these algorithms make unacceptable assuptions about their input. We will describe a methodology that can be used to augment existing algorithms, making them robust to field conditions present in ecological applications. In this paper, we rigorously define the proposed procedure and begin to evaluate its accuracy in the context of an example application.

1 Introduction

There are many important natural phenomena that traditional sensors cannot measure directly. For example, accurately measuring a plant's rate of photosynthesis (release or absorption of CO_2) requires encasing part or all of the plant in a chamber. Then, air entering and leaving the chamber is compared to measure instantaneous CO_2 flux. Such measurement is error prone and must be frequently calibrated, making long term deployment difficult. Additionally, such an apparatus is clearly too bulky and invasive to be used in a field environment.

Visible-light imagers represent a very powerful and untapped sensing modality. When direct measurement is difficult, imagers are the missing input required to accurately model natural phenomena. Images are typically avoided in traditional sensing applications because they produce large quantities of uncalibrated data. The form of calibration required for an imager-based ecological sensor is dissimilar to that of typical sensors; there is no conveniently accessible reference that can be used to calibrate an imager used as a CO_2 sensor, for example. We aim to use state-of-the-art computer vision, image processing, and statistical learning algorithms to build a two step imager calibration process (Figure 1) that can be evaluated in the context of a specific application. First, image features must be extracted that are both domain relevant and immune to changing field conditions. Second, these features are used to model the signal of interest as measured in a controlled laboratory environment. Additionally, this model will also be informed by traditional ecological sensors suggested by structure

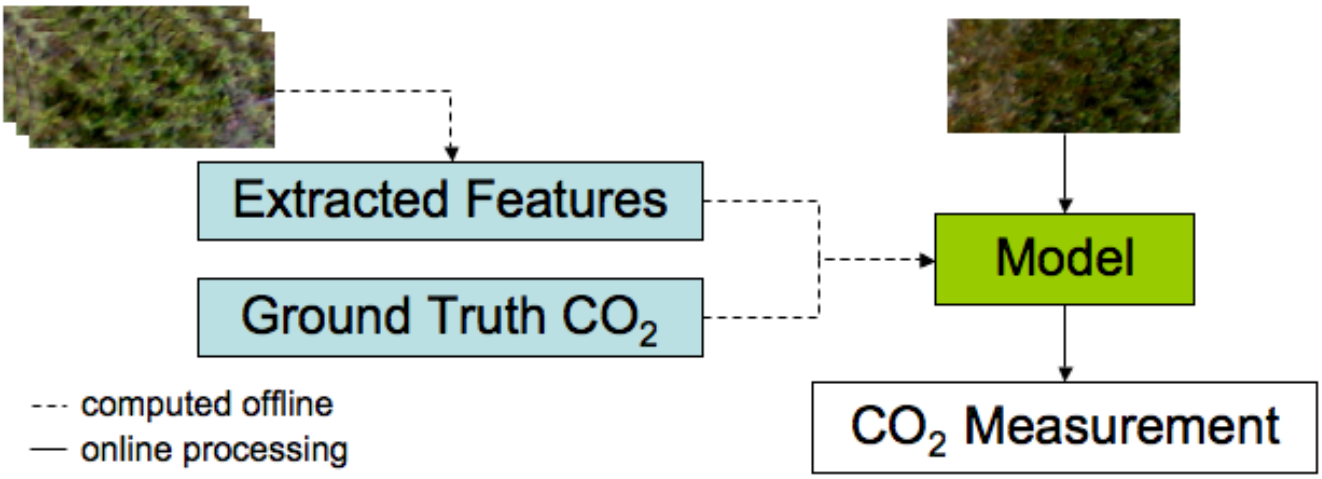


Figure 1: This diagram shows the two parts of imager calibration. First, field condition invariant image features are extracted from images. Second, laboratory data is acquired as ground truth for off-line the signal modeling. Together, both elements are used to calibrate the imager.

of the ecological system. Since each process instantiation is somewhat application specific, we intend to show, though multiple applications, that the process itself is general.

In order to perform meaningful feature extraction, we must account for the spectral response characteristics of the CCD (charge coupled device) or CMOS (complementary metal-oxide semiconductor) sensor as well as the spectral power distribution (SPD) of the incident light. The general form of this calibration, known as color constancy [33], has traditionally been difficult. Various computer vision applications, such as object recognition and image segmentation, would benefit if such calibration could be performed accurately in general. In our applications, we are free to fix the location of the observer (relative to the subject) as well as the subject itself. In particular, we can produce accurate models of both in the incident illumination and subject's spectral reflectance. These simplifying assumptions make this specific instantiation of the color constancy problem more tractable.

Once invariant image features have been extracted, they must be correlated to the signal of interest. Deriving such a correlation requires the construction of a model based on experimentally acquired data from imagers as well as co-located traditional sensors. In the case of photosynthesis measurement, temperature, PAR (photosynthetically active radiation), and rainfall sensors are of particular use. Including traditional sensors, in addition to the imager, has two important benefits. First and foremost, we can use sensing modalities that are correlated with the phenomena to increase prediction accuracy. Second, by incorporating field-deployable sensors into the model, we can more easily reason about the model's accuracy under field conditions.

We choose to use color-based image features because they are relevant to many ecological applications. Statistical models using color-based image features can be trivially built by quantizing the observed color distribution, making each unit an independent variable. However, the introduction of many variables into a regression algorithm causes a polynomial increase in runtime. We show how to take advantage of the fact that these variables are actually dependent to make the runtime linear in the number of image features and potentially increase the model's accuracy. Further, by considering the variables to be dependent, we reduce the number of input dimensions, helping to avoid the curse of dimensionality [5] that would be present when considering these features to be independent. Performing regression on high-dimensional, highly structured input, such as color distributions, is one complication that has come up immediately. We expect that other such problems will arise while

generalizing our procedure.

Once devised, we must evaluate the prediction accuracy of the model. This is somewhat complicated by the fact that ground truth data is unlikely to be available. By using a combination of laboratory experimentation, internal consistency checks, and other environmental cues we can leverage domain relevant information to evaluate our results. The design of the laboratory experiments is especially important. We must capture environmental signals easily measured in the field so as to corroborate the laboratory findings. Though each instantiation of this evaluation is unique to the application, we will show, through multiple applications, that the nature of this evaluation technique is general.

One particular application of ecological interest is the measurement of a drought-tolerant moss, *Tortula princeps*. This moss has the interesting ability to hibernate when conditions are not favorable for its growth. Ecologists are curious why it is not more prevalent in dry climates for which it seems well suited. This particular problem lends itself nicely to the use of imagers because it has been shown [50] [19] that a plant's photosynthetic respiration is related to its spectral reflectance. Further, this and other moss can be reliably modeled in the laboratory and are quite representative of other higher-order plant species [42]. Previous work [40] showed that the photosynthetic respiration of plants would remain at its maximum if it weren't limited by the ambient temperature or the availability light and moisture. Though temperature and light sensors can easily be deployed, moisture measurements are far more complex. Simple thermocouples on the surface of the plant are insufficient. Instead the plant must be destructively measured by removing it from its habitat and its weight compared against a reference dry weight. Thus, for continuous measurement in the field, an imager is an ideal choice of sensor.

1.1 Imager-Based Sensing Applications

There is a large class of sensing applications that can make use of calibrated imagers. We define the subset of applications we consider using a series of suggestive questions. Though this list is not exhaustive, it describes the application characteristics leveraged by our process.

Is an imager the most natural sensor for the phenomena?

For many applications, an imager is the most natural sensor of the phenomena. For example, detecting birds flying past an imager [29] or counting the number of eggs in a nest [1]. Alternatively, the target signal could be logically encoded in image features that are not easily discerned by a human. Thus, using a traditional vision approach, like object detection or image segmentation, is a non-starter. This criteria defines whether the solution will employ mostly computer vision or signal processing techniques. For this work, we consider applications for which an imager is *not* the most natural sensor. This choice enables us to consider the space of solutions that require what we call *applied vision*. This entails applying the physics-based modeling of image formation, as developed by the vision community, to help calibrate the imager and model natural phenomena.

Which spectrum of light is measured by the imager?

Though all frequencies of light from infrared to ultra-violet are of some ecological interest, the visible range has been found to be particularly useful for measuring many phenomena [19] [27] [20] [38] [12]. In this work we confine our measurement to the visible range. This has the additional benefit of allowing for the use of commercially available digital imagers. Current digital imagers use CCD or CMOS sensors that are most sensitive to light in the visible range (400nm – 700nm), and their dynamic range is bandwidth limited by various filters [49]. Intrinsically, these sensors have a dynamic range that extends beyond the visible range into near-infrared as well as ultra-violet and could, in principle, be used as a sensor for those spectra as well.

Event detection or process estimation?

Interest in ecological phenomena broadly falls into two categories: event detection and process estimation. We choose to focus on process estimating, attempting to derive a continuous signal as a function of the imager's output. There are a variety of application-dependent simplifying assumptions that can be made when estimating these processes. For example, they typically have memory, implying that the target signal is continuous. This suggests that the time dimension of the model's input can be used to reduce the prediction error. As a result, our estimation procedure is much different than trying to use an imager to detect vegetation in a field [36] or weeds in a potato patch [48]. Still, imager calibration techniques are common to both event detection and process estimation.

Which image features are extracted?

There are many features that could be extracted from images. Some features, like texture, are somewhat independent of lighting by nature. Other features, like color, intensity, or radiance are significantly affected by the incident illumination. We consider applications where the relative spectral reflectance of the subject over the visible range, a metric related to color, is the most informative feature. This feature has been widely used in everything from soil versus vegetation classification [36] to detecting the presence of clouds [47].

Is ground truth data available for field imagery?

Modeling these systems is greatly simplified if the target signal can be measured in the field for a short period, producing ground truth for future modeling. Though this is sometimes possible, we consider the more general case where such field measurement is not possible. In these cases, laboratory experiments must function as surrogates for data collection in the field. As such, they must be shown to sufficiently capture the space of important inputs to the ecologically process. The resulting model must be evaluated using properties of the ecological system, to place a bound on prediction error. For example, previous work [40] attempting to measure the rate of plant photosynthesis used the fact that plant growth (as measured by leaf area) is related to the integral of carbon uptake over time (the result of photosynthesis).

Modeling a single- or multi-valued signals?

Ecological processes can be defined by a set of measurable responses to their environmental input, some of which may be dependent on one another. We focus on applications that are interested in single-valued signals that are partially dependent on various easily-measurable environmental inputs. In the future we plan to consider applications that require a single signal be predicted over a 2-dimensional area. One particular application is predicting soil surface temperature throughout a meadow. We initially aim to model single-valued signals because they are a simpler regression targets and are easier to validate in the field.

Vantage point: in-situ or remote imaging?

Remote imaging, from satellites or planes, has produced excellent insight into large scale ecosystem processes [51]. However, even with high resolution imagers, single image pixels may represent tens or hundreds of meters. As a result, their predictions are necessarily general as acquiring ground truth to calibrate these measurements is difficult. We choose to focus on local measurements acquired from in-situ imagers because they have the potential to be more accurately calibrated using laboratory experiments. This work defines a process of imager calibration targeted at in-situ imagers similar in spirit to processing techniques used to derive meaning from remote imagery. The data acquired from in-situ imager deployments can then be used by the remote sensing community as ground truth to further refine their predictions.

Are domain-relevant sensors co-located with the imager?

Ecological processes are affected by a variety of different inputs, some more easily measured than others. For example, the rate of plant photosynthesis is known to be affected by the availability of temperature, light, and moisture [40]. We can easily measure temperature and light using traditional sensors that can be deployed in the form of micro-meteorological stations. Further, by measuring these signals both in the field and during laboratory experiments, we can more easily reason about the model's accuracy. For these reasons, we choose to focus on applications that have meaningful co-located sensors in field deployments.

What is the expected sample frequency?

Currently, ecologists can sample environments of interest at a monthly or weekly frequency. As a result, models based on these data collection efforts cannot make predictions with greater than one month resolution. We attempt to collect data multiple times an hour, allowing predictions based on these data to have a resolution of hours. This allows ecologists to study an entirely new set of phenomena that occur over the course of a few days. For example, measuring the effect of a summer rain event on moss photosynthesis. These events are known to be important, but have traditionally been incredibly time-consuming to measure.

1.2 Application Driven Innovation

Building image-based ecological sensors is a driving force for innovation in both sensor networking and computer vision. Traditionally, sensor networking has always endeavored to solve real application problems and innovate by adapting best-of-breed algorithms to the specific task at hand. Similarly, our approach to building an image-based sensor leverages the best available vision algorithms and innovates in areas where those techniques perform poorly. We anticipate further innovation through the reuse of this process for different ecological applications.

1.3 Contributions

- **Application evaluated image-based sensor toolkit:** Define a procedure to correlate images to ecological signals of interest using a series best-of-breed computer vision, image processing, and statistical learning algorithms. We will evaluate the prediction accuracy of this procedure in the context of a specific application, showing how to leverage intrinsic properties of that particular instantiation of the process. Further, we will show the generality of the procedure though its use in multiple applications. Using this procedure we intend to measure photosynthesis of a drought-tolerant moss, *Tortula princeps*, to help ecologists understand its habitat requirements and long-term grown trends.
- **High-dimensional, highly structured data as regression inputs:** Devise the set of alterations required to allow standard non-linear regression techniques to make use of high-dimensional, highly structured data. This will significantly decrease their worst-case complexity and potentially increase their accuracy.
- **Field-robust algorithms and methodology:** We have found that the many of the best-of-breed algorithms make unacceptable assumptions and require modification. For example, the JPEG compression appears to interfere with the *re-lighting* algorithm described in Section 3.3. More generally, we intend to articulate a methodology for making algorithms robust to the field conditions present in ecological applications.

1.4 Putting the work in context

Our proposed procedure has many characteristics in common with research in the agricultural engineering field. This research attempts to use images to monitor crop health, increasing yield by detecting problems quickly. For example, detecting weed growth in crop fields has long been a problem for the agriculture industry. A review of recent literature [9] suggests that there has been a shift from using remote imagers to in-situ imagers. Though remote imagers, typically in aircraft, have success detected weeds when the patches are dense and uniform in color, they have trouble detecting small patches because of their low resolution. In contrast, in-situ mobile imaging devices have had more success detecting smaller patches of weeds growing amongst the crops. One particular system developed by Slaughter et. al. [48] approaches this classification problem from first principles. Like our formulation, they use the physical model of image formation in an attempt to build lighting invariant color features. Additionally, they used shape and texture features as suggested by their application.

Techniques for dealing with natural lighting conditions from a machine vision perspective are discussed in a series of works by Marchant et al. [36] [35] [34]. Set in an agricultural context, they attempt to modify existing vision algorithms to better distinguish soil from vegetation. Similar to our procedure, they choose to use the distribution of sensor values as input their model since the possible subjects have very different spectral reflectance characteristics. In their case, the model was a binary classifier since they were interested in a binary signal. They approximate the spectra of daylight using an idealized black-body radiator and approximate the spectral sensitivity of the camera sensors as impulse functions. A transformation was constructed, based on a ratio of sensor responses and other factors, which rendered their images sufficiently independent of changing illumination. Finally, they showed that this transform effectively separated field imagery of soil and vegetation. Unlike our goal, this and the previous system attempt to build binary classifiers of the image's subject. However, the techniques they use (further discussed in Section 3) are applicable to our process formulation.

The goal of our work is similar in to much of the research in the environmental monitoring field. Unlike our work, however, typical research in that discipline performs rudimentary analysis of the images, ignoring the physical models of image formation. A representative work is the environmental monitoring system described by Crimmins et. al. [12]. It attempts to measure relative vegetation coverage by producing a "greenness" signal from a sequence of images. This signal is simply computed using the difference between the mean value of the color channels. They show that even this simple image feature tracks the increase in plant coverage over a three month period relatively well. However, this feature began to loose stability once the image became shaded by the canopy's growth. In fact, the system that inspired our work [20] used a similarly simple feature (average red-to-green ratio) to predict moss CO_2 uptake. However, they were only able to accurately predict relatively large values of CO_2 uptake. They posited that for small values of the target signal, there was less variation in the image and thus a simple average ratio was insufficient.

The remote sensing community approaches this sensing problem from a signal processing perspective. Common practice in these works has been to devise an image feature that is linearly related to the signal of interest. To some extent, this is the inverse of the environmental monitoring field's approach; instead of hypothesizing a feature using domain knowledge and measuring the correlation, they derive a feature which is defined to be well correlated. For example, the Dark Green Color Index (DGCI) [27] attempts to measure the species and health of commercial turf grass fields. The Damage Sensitive Spectral Index (DSSI) [38] tries to measure the damage to a wheat crop caused by weather or insects. By far, the most common feature is the Normalized Difference Vegetative Index (NDVI) [46] and its two close derivatives, the Soil-Adjusted Vegetation Index (SAVI) [23] and the Atmospherically Resistant Vegetation Index (ARVI) [28]. These indexes attempt to measure how much live, green vegetation is present in an image. Our work tries to strike a balance between the data-driven approach of the remote sensing community and the theory-driven approach of the environmental monitoring

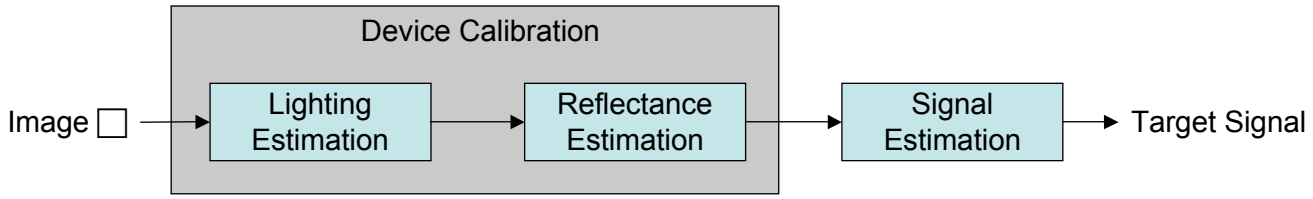


Figure 2: The process we propose consists of the two logical parts depicted here: device calibration and signal estimation.

community. We do this by imparting structure to the procedure rooted in theory while allowing the models to adapt to the data.

Similar hybrid approaches are seen in the remote sensing literature. For example, the satellite-based Multiangle Imaging SpectoRadiometer (MISR) [32] system attempts to detect the presence of clouds and cloud thickness using visible light and infrared imagery. They used image features based on radiance (the reflective characteristic of the subject) because they know that there is a significant difference between in energy reflected by land and the energy reflected by clouds. More recent analysis [47] of data produced by that satellite produced binary classifiers based on SVMs to identify features feature were most useful when attempting to distinguish ice sheets from clouds; both of which have very similar radiance.

Like MISR, the river morphology measurement system developed by Legleiter et. al. [31] takes a hybrid approach to feature selection. The purpose of this system was to derive the depth of a river channel from visible-light imagery. Informed by the application, they choose to extract the log-ratio of color band pairs. These pairs were selected such that one band had much greater attenuation in water than the other. As a result, the log-ratio of these values is sensitive to the river's depth and less effected by suspended sediment. A model, calibrated against field measurements, was then derived by linearly transforming the log-ratio.

Our work leverages the best that these distinct communities have to offer, producing a image processing toolkit suitable for in-situ imagers. The features we select are domain relevant but the phenomena are modeled using non-linear techniques. The sensing system we propose takes advantage of inexpensive, readily deployable, visible light sensors. Compared to satellite- or plane-based remote imagers, they have much higher temporal and spatial resolution. Unlike the trend in the agricultural engineering community, we believe that remote and in-situ imagery can easily work harmoniously to measure natural phenomena. A combined approach where data from a remote sensing applications can trigger the deployment of localized in-situ imagers would be mutually beneficial. This takes advantage of the significant strengths of each technology: the large coverage area of remote imagers, and the higher spatial resolution of in-situ imagers.

This work also draw heavily from the compute vision and statistics community. We discuss work related to image formation and illumination modeling in Section 3 and Section 3.2 respectively. Additionally, we discuss work related to reformulation of non-linear regression to use high-dimensional, highly structured data as inputs in Section 4.

2 Procedure Overview

Calibrating an imager for use as a sensor requires two fundamental steps: device calibration and target signal modeling. The process we propose, depicted in Figure 2, is constructed from a series of models that eventually produce the signal of ecological interest. We choose this configuration both

because it is suggested by the physical model of image formation and because it allows us to easily reuse existing algorithms. The interface between device calibration and signal estimation models is the relative spectral reflectance of the subject. As previously discussed, we assume that this feature is relevant to the signal of interest.

The state-of-the-art vision algorithms we use to form the various models were formulated independently. Some are formulated in a regression context, and others are formulated in a Bayesian context. Additionally, they make vary physical assumptions about the lighting, subject, and camera. As a result, their combination in our framework a bit awkward. In future work, we intend to sort out this inconsistency, placing all the stages of our procedure on consistent theoretical ground.

The remainder of the paper is organized as follows. Section 3 discusses lighting modeling and estimation as well as reflectance modeling and estimation. Section 4 discusses the procedure for estimating a target signal using the subject's relative spectral reflectance. Finally, Section 5 evaluates this procedure in the context of an application.

3 Device Calibration

The purpose of device calibration is to undo the effect of changing environmental conditions on the image formation process. In particular, we would like to accurately reconstruct the relative spectral reflectance of the subject given color features extracted from an image. Formally, image formation is composed of three components: the spectral power distribution (SPD) of the incident light $E(\lambda)$, the relative spectral reflectance of the surface $S(\lambda)$, and the spectral response of the imaging device's sensor $R(\lambda)$. There are two types of spectral reflectance. Light that reflects directly off the surface is known as interface reflectance, usually seen as the spectral highlight off of a glossy surface. Light that enters the surface and interacts with colorant particles is known as body reflectance [52]. Assuming the surface is matte or Lambertian, having only body reflection, the response of the imager's kth sensor to a *(lighting, surface)* pair over the spectral range w is defined by Equation 1.

$$r_k = \int_w E(\lambda)S(\lambda)R_k(\lambda)d\lambda \tag{1}$$

For typical visible light imagers, $w = (400nm, 700nm)$ specifying the visible range, and $k = 3$ corresponding to the red, green, and blue sensors in the imager. Since common commercial imagers intend for their output to be consumed by humans, having only three color sensors is reasonable; human color vision was determined to be a 3-dimensional space by color matching experiments [13]. That is, the use of three orthogonal sensors can represent most[1] of the gamut of human color vision. However, for the class of problems defined in Section 1.1, we are not interested the resulting human-perceived color; instead, we are interested in $S(\lambda)$, the relative spectral reflectance of the matte surface contained in the image.

This formulation is a bit simplistic. In particular it doesn't capture second-order effects attributed to the camera's lens, shutter speed, and aperture. We assume that the lens' distortion is uniform across the image and that the shutter speed an aperture are set such that the sensor is not saturated. An effect we can't ignore is JPEG image compression [11]. This compression algorithm is both lossy and has a spacial component, considering multiple adjacent pixels at a time. We consider the effects of JPEG compression on this model in Section 5.2.

[1] Any basis defined by human-visible colors cannot represent the entire gamut of human vision using positive coefficients. This fact can be described geometrically. The projection of the 3-dimensional human color gamut onto a plane of uniform brightness (chromaticity space) results in a convex polygon [21]. In this plane, the basis functions are represented by points, and the space of all colors represented by their linear combination using positive coefficients is a triangle. There is no triangle composed of points within a convex polygon that contain all points within that polygon.

3.1 Modeling Illumination and Relative Spectral Reflectance

We build a 3-dimensional linear model for the surface reflectance of the subject using principle component analysis (PCA) [45]; this results in a set of basis functions $\mathbf{B}$ and their weights w. We can write this in matrix notation (Equation 2) if we discretize the spectral range into n bins; $\mathbf{B}$ is a $n \times 3$ matrix, w is a 3×1 weight vector, and $\hat{S}(\lambda)$ is a $n \times 1$ vector that estimates of the surface's spectral reflectance. Since we are considering outdoor ecological applications, we can apply previous work [26] that has similarly defined a 3-dimensional linear model for daylight (Equation 3) using PCA.

$$\hat{S}(\lambda) \approx \mathbf{B_s} w_s \tag{2}$$

$$\hat{E}(\lambda) \approx \mathbf{B_e} w_e \tag{3}$$

Initially, we build the lighting and reflectance models independently. In the future, we intend to build these models iteratively because simply modeling each independently is sub-optimal [37]. In particular, the reflectance model can be designed to best fit the areas of most change when illuminated by different relevant spectra. Similarly, the measured spectral sensitivity of the imager can be incorporated to reduce the model's emphasis on wavelengths for which the imager has minimal sensitivity.

Once we have models for illumination and relative reflectance, we must mitigate the effect of the camera's shutter speed and aperture on $R_k(\lambda)$. Changing the shutter speed and aperture results in the image being under- or over-exposed. We assume that this effect is uniform across the sensor and that the sensor is never completely saturated (avoiding the loss of information). By using 2-dimensional chromaticity coordinates, instead of the raw 3-dimensional color coordinates, we can compensate for this uniform change in brightness. The chromaticity space is the projection of the 3-dimensional color space onto a plane of uniform brightness, and thus mitigates the effects of exposure. The chromaticity space we choose is the x and y dimensions of the xyY color space as defined by CIE [43].

$$\begin{aligned} r &\approx \hat{E}(\lambda)\hat{S}(\lambda)^T R(\lambda) \\ r &\approx (\mathbf{B_e} w_e)(\mathbf{B_s} w_s)^T \mathbf{R} \end{aligned} \tag{4}$$

Our resulting model for image formation (Equation 4), has six unknowns: the w_e and w_s weight vectors. As described, this system is under constrained since we only have two equations as defined by the chromaticity coordinates produced from the three sensors available in commercial imagers. Thus, we must estimate both w_e and w_s using the distribution of chromaticity coordinates present in the image. We proceed by estimating these values in sequence. First, we estimate w_e to produce the illuminant's spectra. Then, we transform the image to place it under a reference illuminant. From this "registered" image, we estimate w_s resulting in $\hat{S}(\lambda)$, an estimate of the subject's spectral reflectance. We assume that the same camera is used to produce all of the analyzed images, and thus the effect of $R_k(\lambda)$ on the final pixel value is constant across all images. Thus, we need only compensate for $E(\lambda)$ when creating the registered image.

3.2 Estimating Incident Lighting

There are a number of lighting estimation techniques suggested by the literature [3], each making different assumptions about the lighting and subject present in the image. Since our applications may have a fixed set of possible illuminants (for example, a subset of daylight illuminants) and typically have a single subject, we would like leverage that information during lighting estimation. Depending on the nature of the application we can use either the Color by Correlation [16] algorithm or the

Gamut Mapping algorithm [17] [14] [2]. The Color by Correlation algorithm assumes knowledge of both the subject as well as all possible illuminations. As a result, it is only capable of predicting illuminations that it has "seen" before. In contrast, the Gamut Mapping algorithm only assumes knowledge of image's subject. Consequently, it can predict an infinite set of possible illuminants.

By leveraging more application specific information, the Color by Correlation algorithm has been shown to slightly out-perform [22] the Gamut Mapping algorithm (both easily out-perform other more simplistic algorithms). Thus, the trade-off between these two algorithms is simply generality versus accuracy. If the set of possible illuminants can be defined, the Color by Correlation algorithm is superior. If not, we must turn to the Gamut Mapping algorithm. Since the choice of algorithm is application dependent, we present a short explanation of each here.

Color by Correlation

The Color by Correlation algorithm computes a correlation matrix representing the probability that given illuminant was present in a particular image. Each column of the matrix is possible illuminant, and each row is the probability that a particular chromaticity coordinate would be observed for surfaces under that particular illumination. Since the chromaticity space has infinite extent, the space is quantized to make building a correlation matrix feasible.

Producing the log-likelihood is a simple application of Bayes' rule. For a given illuminant E and a given set of observed chromaticities C_{im}, Equation 5 defines the probability that E was the illuminant for C_{im}. If we assume that the prior probabilities for E and C_{im} are uniform, all illuminations and surfaces are equally likely, then Equation 5 simplifies to Equation 6.

$$Pr(E|C_{im}) = \frac{Pr(C_{im}|E)Pr(E)}{Pr(C_{im})} \tag{5}$$

$$Pr(E|C_{im}) \propto Pr(C_{im}|E) \tag{6}$$

Further, we note that $Pr(C_{im}|E)$ is simply the product of the probability of observing each chromaticity c (Equation 7). Finally, if we take the logarithm of both sides (Equation 8), we get the same value as produced by multiplying the correlation matrix to a particular image's binary chromaticity vector. The binary chromaticity vector of an image is 1 for every value that is present in the image, and 0 elsewhere.

$$Pr(E|C_{im}) \propto \prod_{\forall c \in C_{im}} Pr(c|E) \tag{7}$$

$$log(Pr(E|C_{im})) \propto \sum_{\forall c \in C_{im}} log(Pr(c|E)) \tag{8}$$

There are two major shortcomings of this algorithm as suggested by [3]; both are related to the assumption that the set of possible illuminants is fixed. First, the set of observed chromaticity coordinates may suggest that none of the illuminants are possible. Second, the algorithm cannot predict a mixture of known illuminants. To solve the first problem, they suggest smoothing the frequency distribution of the chromaticity coordinates using a Gaussian filter. However, this still requires us to train the algorithm using an illuminant set that has complete coverage of all possible illuminants. As we suggested earlier, if such a set cannot meaningfully be produced, then the Gamut Mapping algorithm is a better choice for lighting estimation.

Gamut Mapping

The Gamut Mapping algorithm assumes a set of known surfaces defined by the convex hull of their combined color gamut under a known illuminant. However, it makes no assumption about the set of possible illuminants to which those surfaces may be subjected. In this context, the gamut is defined to be the set of all color coordinates that can be produced by the given surfaces, under a given lighting, with a given camera [17]. More recent approaches [14] measure this gamut in chromaticity space making it more robust to illumination intensity. This algorithm attempts to derive a transformation (or change of basis) to map the observed gamut under unknown illumination to the measured gamut under a known illumination.

$$E_{ref} = \begin{bmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{bmatrix} E_{measured} \tag{9}$$

These transforms, as represented by their diagonal matrices (Equation 9), define the change in whitepoint between the reference illumination and the unknown illumination. The whitepoint of an illuminant is the color coordinate for a pure white Lambertian surface as viewed under that illuminant. Thus, this transform is equivalent to determining the properties of the unknown light source with respect to the reference. In general, lighting transformation matrices (as represented by Equation 4) are not purely diagonal. However, *von Kries coefficient law* tells us that the diagonal values are most influential.

Unfortunately, there is no unique transform because we incorrectly assumed that the unknown illuminant's gamut was *equal* to the measured gamut. In fact, the measured gamut is a proper subset of the unknown gamut we we have only one sample image's gamut under that illuminant. This causes there to be several transforms that map the unknown illuminant to the reference illuminant. How to chose the "best" transform from this set has been contested in the literature. All solutions involve choosing the "average" solution, which has slightly different meaning depending on the exact problem formulation [2].

Since the Gamut Mapping algorithm can potentially produce any white point as output, it is clearly more general than the Color by Correlation algorithm. However, it makes the assumption that the camera's sensors are sufficiently narrow bandwidth that Equation 1 can be simplified to Equation 10. That is, they can be modeled as impulse functions at some wavelength λ_k, typically the center wavelength of the camera's sensor.

$$r_k = E(\lambda_k)S(\lambda_k) \tag{10}$$

This assumption is clearly not true of typical cameras. A technique known as sensor sharpening [4] attempts to map a camera's wide bandwidth sensors to narrow bandwidth (sharpened) sensors. Additionally, von Kries coefficient law is also somewhat unrealistic. However, it has been shown [15] [54] that for "reasonable" illuminants (such as daylight), it appears to hold.

3.3 Changing Illumination

After we've estimated the lighting present in a given image, we must transform the images to be under some reference illuminant. We call this operation *re-lighting* the image. Since we are considering outdoor phenomena, we choose D_{65} [41] as the reference illumination; D_{65} is an approximation of daylight defined by CIE [43]. Producing a re-lighting transform when using the Gamut Mapping lighting estimation algorithm is trivial: we simply compute the reference gamut from images illuminated by a D_{65} source and use the resulting diagonal lighting transform.

Building a re-lighting transform from the output of the Color by Correlation algorithm requires that we produce a lighting transformation matrix. Like Gamut Mapping, we assume that the camera's sensors are impulse functions at the sensor's center wavelength. We can define the lighting transformation T_{light}, in terms of Equation 10, as Equation 11.

$$\begin{bmatrix} E_1(\lambda_R)S(\lambda_R) \\ E_1(\lambda_G)S(\lambda_G) \\ E_1(\lambda_B)S(\lambda_B) \end{bmatrix} = T_{light} \begin{bmatrix} E_2(\lambda_R)S(\lambda_R) \\ E_2(\lambda_G)S(\lambda_G) \\ E_2(\lambda_B)S(\lambda_B) \end{bmatrix}$$

$$T_{light} = \begin{bmatrix} E_1(\lambda_R)/E_2(\lambda_R) & 0 & 0 \\ 0 & E_1(\lambda_G)/E_2(\lambda_G) & 0 \\ 0 & 0 & E_1(\lambda_B)/E_2(\lambda_B) \end{bmatrix} \tag{11}$$

We choose these center wavelengths to be $\lambda_R = 620nm$, $\lambda_G = 530nm$, and $\lambda_B = 450nm$; these are close to the center wavelength of the sensors on typical digital cameras [15]. The Color by Correlation algorithm produces $E_2(\lambda)$ and we have already assumed that $E_1(\lambda)$ is the standard D_{65} illuminant. To re-light the image, we need only compute the diagonal lighting matrix and then transform each of the image's pixels individually.

This formulation only works if we specified $E_1(\lambda_k)$ and $E_2(\lambda_k)$ in absolute terms. However, the spectral power distribution of an illumination is typically normalized such that $E(\lambda_{560}) = 100$ (as is the case for the D_{65} specification). This has the effect of multiplying T_{light} by β as defined in Equation 12.

$$\beta = \frac{E_2(\lambda_{560})}{E_1(\lambda_{560})} \cdot 100 \tag{12}$$

The β term can be factored out of the resulting transformed image if we use chromaticity coordinates instead of absolute color coordinates. This is intuitively true since chromaticity coordinates are designed to be independent of brightness, the effect for which β compensates. Further, using chromaticity coordinates is a reasonable requirement as we have already leveraged chromaticity coordinates to produce a brightness invariant image for lighting estimation.

3.4 Estimating Relative Spectral Reflectance

We can now estimate the weights w_s for the relative spectral reflectance basis functions B_s (see Equation 2) derived by using PCA. Unlike lighting estimation, however, we have less insight into the relationship between relative spectral reflectance and the chromaticity coordinates. Accordingly, we choose to estimate the parameters of our relative spectral reflectance model using non-linear regression. The input to this non-linear regression will be the 2-dimensional chromaticity coordinates. Similar to the Color by Correlation algorithm, we quantize the chromaticity space into $n \times n$ bins, using each as feature in our predictive model. These features are stable between images since we previously corrected for changes in illumination using the re-lighting transform.

Our previous work [24] showed that using this technique produced reasonable results for laboratory data. The dataset used in that work had consistent illumination since all images were taken under controlled laboratory lighting. As a result, it did not require the images to be chromatically registered using a re-lighting transform. Instead of using the x and y dimensions of the xyY color space, that work used the H and S dimensions of the HSV color space. Like xyY, HSV is a deformation of the RGB color space that extracts the brightness (the V dimension) from the chromaticity (the H and S dimensions). Unlike the formulation described here, the target signal was directly modeled from the quantized chromaticity.

4 Modeling the Target Signal

Data about the relationship between the relative spectral reflectance, other co-located sensor, and the signal of interest can then be gathered in a laboratory experiment. Such experiments, typically suggested by the domain science, must be sufficiently realistic such that derived models have predictive power in the field. These input data are both high-dimensional and highly structured. Conventional non-linear regression techniques are not well equip to take advantage of the structure in these data. We show how to modify one particular algorithm, regression trees, to leverage this structure.

We would like to model the relationship between the relative spectral reflectance and the target signal using regression trees [8] because of insight we gain from the resulting model's form. However, performance of the regression tree algorithm is $O(nk^2)$ where n is the number of features and k is the number of training examples. In particular, the algorithm is searching for the best $(feature, threshold)$ pair using some error metric E. This requires us to consider each $(feature, threshold)$ pair, of which there are $O(nk)$, and compute E for each. Computing E requires considering whether each example resides on the correct side of the threshold, an $O(k)$ operation.

Using the histogram of the relative spectral reflectance with h buckets as input, our runtime becomes $O((n+h)k^2)$. The histogram, which is logically one feature, increases the runtime by $O(hk^2)$. This begins to dominate when we include multiple such histograms for sequential samples in time to take advantage of the temporal continuity of our target signal. By approximating the input data as a location-scale distribution, we can reduce this incremental runtime to $O(h+k)$.

A member of a location-scale family can be written as as Equation 13 where m_i is the mean, s_i is the standard deviation, and $X(t)$ is the unit distribution in that family; the normal distribution is an example of a location-scale family [25].

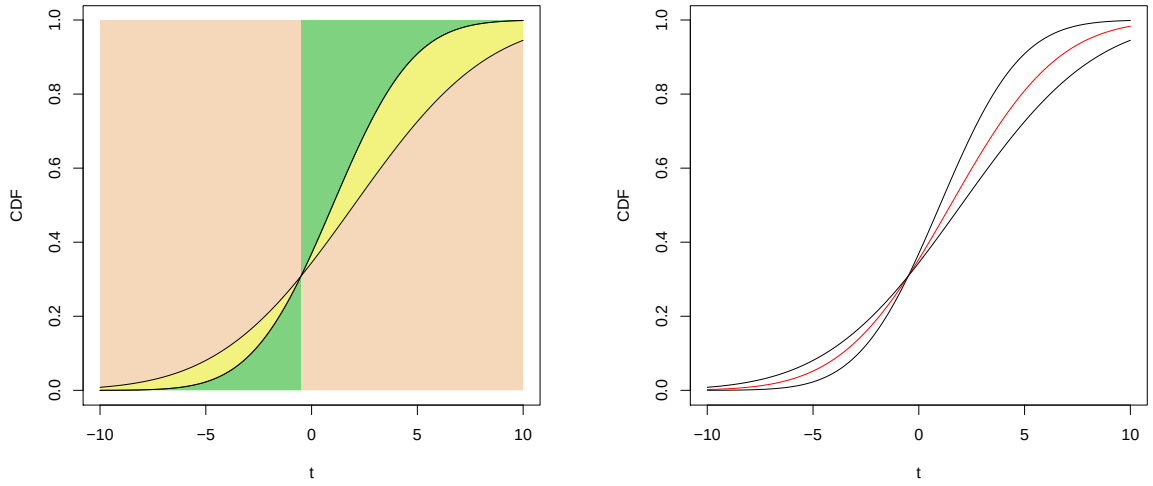
$$t = s_i X(t) + m_i \tag{13}$$

We can then approximate our relative spectral reflectance by some member of this family using an Expectation-Maximization (EM) algorithm [18]. Instead of the histograms themselves, the inputs to the regression tree become cumulative distribution functions (CDF) of the fitted distribution. We observe that any pair of CDFs will only cross once and thus have only three meaningful $(feature, threshold)$ pairs, as seen in Figure 3(a); here *feature* is a synonym for histogram bucket. More generally, a set of k CDFs will cross at most k^2 times producing at most $(k+2)^2$ meaningful $(feature, threshold)$ pairs. Thus, our incremental runtime is now independent of h but and has been increased to $O(k^3)$ (recall that we must compute E at each possible result).

4.1 Distributions As Input

Since all of the CDFs are monotonically increasing and continuous, we need not consider all k^2 intersections. The selection of a $(feature, threshold)$ pair is equivalent to the selection of a member from the location-scale that passes through that chosen location, called a *threshold CDF*, as seen in Figure 3(b). Thus, we need only consider when other examples cross our threshold CDF. Other intersections can be ignored because they do not effect E; the order of curves with respect to the chosen threshold CDF is unchanged. Further, since the error only changes upon an intersection with the threshold CDF, the error can be incrementally recomputed in constant time because only a single example needs to be updated. With the use of a threshold CDF and an incrementally computed error, we can reduce the runtime to $O(k+hk)$. We don't achieve $O(k+h)$ complexity because we must consider each other example at each histogram bucket to find possible intersections.

To find the intersection of two distributions in the same location-scale family, we can invert Equation 13 and set them equal to each other (Equation 14). Then we can easily solve for their intersection as function of their parameters (Equation 15).



(a) Meaningful *(feature, threshold)* values

(b) Threshold CDF

Figure 3: Two CDFs (plotted in black) from the normal location-scale family. The shaded sections of 3(a) show the three different threshold values (areas shaded in like colors represent the same logical threshold). The red curve in 3(b) represents a *threshold CDF* for the yellow threshold.

$$\frac{t - m_1}{s_1} = \frac{t - m_2}{s_2} \tag{14}$$

$$t = \frac{m_1 - m_2}{s_2 - s_1} \tag{15}$$

4.2 Efficiently Finding the Intersection

If we impose some relationship between the m_i and s_i, we can find the "next" intersection in constant time. Alternatively, we can model the samples using a weighted mixture of two location-scale distributions from the same family; this mixture itself represents a new location-scale family. In this case, we can rewrite Equation 13 as Equation 16. Inverting that equation and setting it equal to another mixture with a different weight (fixing the mean and standard deviation) and yields Equation 17. Finally, we can solve for the intersection (Equation 18).

$$t = \alpha(s_i X(t) + m_i) + (1 - \alpha)(s_i X(t) + m_i) \tag{16}$$

$$\frac{t - \alpha_1 m_1 - (1 - \alpha_1)m_2}{\alpha_1 s_1 + (1 - \alpha_1)s_1} = \frac{t - \alpha_2 m_1 - (1 - \alpha_2)m_2}{\alpha_2 s_1 + (1 - \alpha_2)s_1} \tag{17}$$

$$t = \frac{s_1 m_2 - s_2 m_1}{s_2 - s_1} \tag{18}$$

We see that the intersection doesn't depend on α. This means that either all curves intersect at the same point, or don't intersect at all (while simplifying from Equation 17 to Equation 18, we may have divided by zero, making the intersection undefined). As a result, there are at exactly $(k + 2)$ meaningful thresholds for k samples. Thus, the runtime of adding a histogram feature with h buckets

has an incremental worst case runtime of $O(h + k)$. In future work, we intend to generalize this technique to work with more complex model formulations and other non-linear regression algorithms.

4.3 Model Validation

For the ecological systems we consider, there is no meaningful way to capture ground truth in the field. As a result, we must validate the predictions of our model through other means. This requirement can be partitioned into two types of validation. First, we would like to ensure that the magnitude of individual predicted measurements are accurate. Second, we want to ensure that the process we are modeling progresses at a reasonable rate through time. Any validation procedure is likely to be highly application specific. However, there are certain best-practices that can be applied in general.

Previous ecological work [40] that studied CO_2 uptake in other plants encountered a similar problem. Their solution was to relate their predictions to an expectation of plant growth. In laboratory experiments, they correlated their model's estimates with an increase or decrease in leaf count. This approach takes advantage of domain specific information to validate the model: a net CO_2 gain should result in more leaves and a net CO_2 loss should result in fewer leaves. Validating the model simply required measuring leaf count in the field and comparing to the model's prediction. An extension of this approach could ensure internal consistency of the model by measuring net CO_2 gain during a period where no leaves were lost or created. During these periods, we expect there to be a net zero gain in CO_2.

In general, this metric suggests that easily observable characteristics of the system (either visual cues or other deployed sensors) can be used to validate prediction accuracy. This seems tautological: if such metrics existed, we would use them to help model the system. However, we are interested in the absolute instantaneous value of the signal, and this type of metric essentially measures the integral of that signal's value over time.

This form of validation metric attempts to remove absolute error from our predictions but contains no time component. As suggested earlier, the processes we'd like to model have memory and produce continuous signals over time. Factoring in the time component again requires application specific cues. For example, for measuring photosynthesis in moss, we know that such activity only happens while the moss is hydrated. The moss become hydrated briefly at dawn (from morning dew) and after a rain. Further, the duration of hydration mostly depends on the ambient air temperature and relative humidity. Thus, we can use the time of day or quantity of rain in addition to temperature and humidity to estimate the duration of active photosynthesis. Again, this estimates an orthogonal signal but allows us to evaluate the accuracy of our model.

A final approach is to build more realistic (less controlled) laboratory experiments. This requires that we be able to measure the signal of interest in simulated field conditions, which may not always be possible. Following the moss example, we could perform the photosynthesis measurement outside under natural light with uncontrolled (but representative) temperature and humidity. This type of experiment would provide us with measurements of the target signal that can be directly compared to the predicted values.

4.4 System Overview and Requirements

The process we have described is shown in Figure 4. It makes a few specific assumptions about the subject, the camera, and the application itself. We assume that the subject of the image can be modeled as a Lambertian surface, it is matte with no spectral highlights. Though not a very restrictive assumption, we assume that the illumination and reflection can be accurately modeled; this is the case for daylight and most natural surfaces. The specifics of the lighting-related assumptions made by the illumination estimation algorithms we employ were discussed earlier. We assume all images were taken

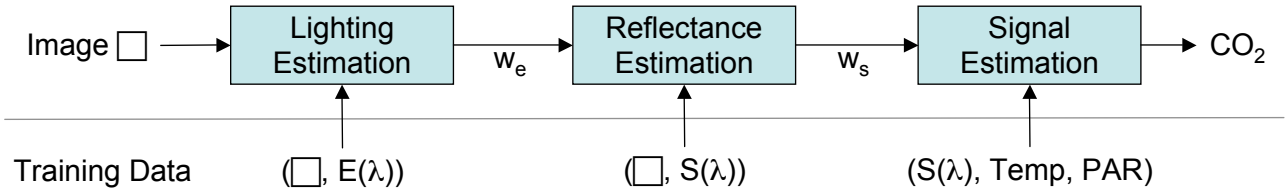


Figure 4: A graphical representation of Equation 4 with the addition of signal estimation. The models (boxes on the top row) are trained using the data depicted below each. These data are experimentally acquired. This particular instantiation of the process predicts CO_2 uptake.

with the same camera and that the shutter speed and aperture were adjusted to avoid saturating the sensor (as it typically the case for most modern cameras). We ignore second-order camera effects, like lens distortion, assuming that they are uniform across the image and have little influence. Finally, we expect there to be other in-situ ecological sensors available that can be used both for modeling and model validation.

Our process is derived directly from the physical model of image formation, and is broken into stages. First, we estimate the lighting present in the scene using either the Color by Correlation or Gamut Mapping algorithm. Given the lighting, we can perform a change of basis to place the scene under a reference illuminant. Next, we predict the relative spectral reflectance of the surface in the transformed image. Finally, using co-located sensors and the predicted relative spectral reflectance, we estimate the target signal using a non-linear regression algorithm specifically modified for use with distributions as inputs.

The current formulation suggests that the parameters of the lighting, reflectance, and target signal models be computed in sequence. An alternate approach would be to estimate the parameters of all three models at once. Such an approach could conceivably do equally well, but discards seemingly important information. Namely, it doesn't explicitly attempt to account for the predicted lighting, possibly causing more error in reflectance estimation. Yet, such an approach is an interesting generalization since it is able to apply domain knowledge, in the form of the individual models, without any additional supervision. We intend to evaluate this alternate approach in future work.

5 Application-driven evaluation: a case study

As suggested earlier, estimating the photosynthesis of *Tortula princeps*, a drought tolerant moss, is an example where imagers can become very useful sensors. Previous work [40] has produced monthly estimates of photosynthesis for plants in the field. Using field-based imagers, we can easily produce hourly photosynthesis estimates. High temporal resolution is of particular interest for this application since this moss can begin photosynthesizing mere minutes after becoming hydrated after a long dry spell (Figure 5).

Previous work suggests that relative spectral reflectance of a plant is related to that plant's photosynthesis and overall CO_2 uptake [50]. This intuitively makes sense since greener plants are rich in chlorophyll, a reactive photo-pigment involved in carbon uptake [19]. Thus, we expect that the light reflected from an actively photosynthesizing plant would be related to the relative spectral reflectance of the chlorophyll molecule. In the following sections we endeavor to calibrate an imager to measure the relative spectral reflectance of this moss. In future work, we will use our modified regression tree algorithm, discussed in Section 4, to model CO_2 uptake given an estimate of the moss' relative spectral

(a) Hydrating the moss (at 16:45)

(b) Moss actively photosynthesizing (at 16:50)

Figure 5: Moss at James Reserve during July 2008 after a long dry period. 5(a) shows us hydrating the moss and it beginning to photosynthesize in the moist areas. After only 5 minutes, much of the moss is green and photosynthesizing as seen in 5(b).

reflectance.

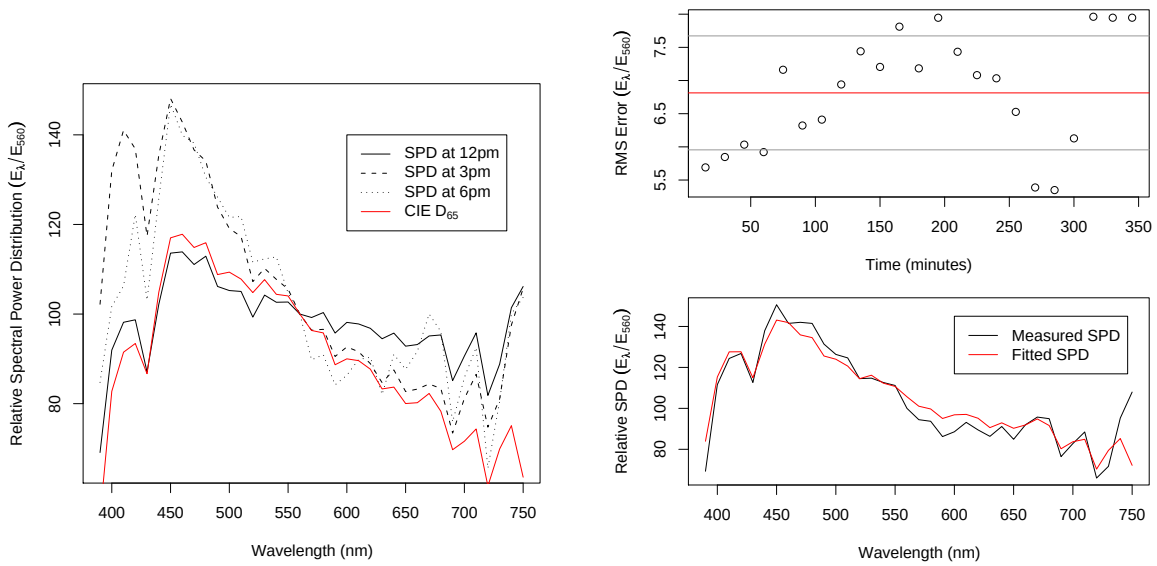
The goal of this ecological study is to determine the effect of short summer rain events on the moss' ability to survive. Ecologists hypothesize that short summer rain events are detrimental to the moss because it causes the moss to expend more carbon than is is able to uptake. Thus, it has been suggested that this moss is capable of surviving long hot summers as long as there is minimal rain during those periods. Using the estimates provided by our imager-based sensor, we can test this hypothesis.

5.1 Experimental Setup

This experiment attempts to performs two related functions. First, we aim to model the relative spectral reflectance of the moss as it drys over the course of a day. Second, we aim to show we can use a camera to accurately estimate the relative spectral reflectance of a fixed subject under realistic (and changing) natural illumination. We acquired a number of moss samples from the James Reserve, seen in Figure 5. We hydrated the moss and allowed it to dry for approximately 6 hours, from 12pm until 6pm. We collected samples of the illumination, moss' relative spectral reflectance, and images containing the moss and MacBeth Color Checker with an interval of 15 minutes. In total, 23 samples were collected. As mentioned earlier, there is a temporal component to the model validation. By allowing the moss to dry over a period, we attempt to include those temporal variation in our training data.

In order to measure both the incident illumination as well as the plant's relative spectral reflectance, we used a spectroradiometer (Licor 1800 [6]). To measure the absolute spectral power distribution of the incident illumination, we calibrated the response of the spectroradiometer using a reference tungsten illuminant, similar to the CIE A reference [43]. Similarly, the spectral reflectance of the plant was measured with respect to same tungsten illuminant. Samples of both the plant and the incident illumination were taken at 2nm increments from 390nm to 750nm.

Images of moss were taken using two standard consumer-grade cameras with their auto white-balance settings turned off. We used a Canon EOS 450D [10] to capture 10MP images in both RAW format and JPEG format; this camera represents a relatively high-end imager. Additionally, we used a Pentax Optio S5z [44] to capture 5MP images in JPEG format; this camera represents a lower-end imager. Each image contained both the moss sample as well the MacBeth Color Checker reference;



(a) Measured Illumination

(b) Illumination model accuracy

Figure 6: The spectral power distribution (SPD) of the illumination measured during the course of the experiment as well as the CIE standard D_{65} illuminant is shown in 6(a). The accuracy of the daylight model built by Judd et. al. [26] for our measured illuminants is plotted against time in 6(b), the red line is the mean and the grey lines are the first standard deviation. Below, we show the fit for the sample with the largest RMS error (the 21st sample at 315 minutes).

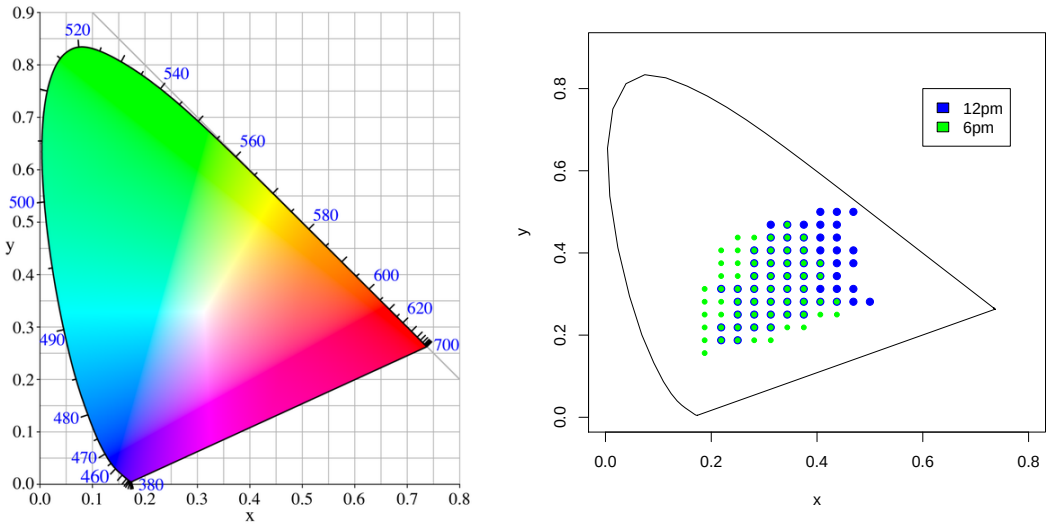
this chart contains 24 color swatches of known spectral reflectance.

5.2 Evaluation

We verified that the illumination we measured using the spectroradiometer was reasonable by comparing it to the CIE standard D_{65} illuminant as seen in Figure 6(a). Each of these spectra have been normalized such that $E(\lambda_{560}) = 100$. The CIE standard D_{65} illuminant is an approximation of daylight as measured around the northern hemisphere. We see that our measured spectra have the same characteristic shape as D_{65} although they are slightly bluer late in the day; they contain more power in the 400nm–500nm range than D_{65}. This similarity suggests that our measurements are producing reasonable spectra.

Using the daylight model derived by Judd et. al. [26], we computed the weights (w_e) of the basis functions (B_e), as defined in Equation 3, for our measured illuminants. As shown in Figure 6(b), the RMS error of this model does follow some time dependent trend through the course of the day. Initially, this might suggest that the model is missing some relevant information. However, we see that the fit for the example with the largest absolute RMS error still is quite good. This further confirms that our measurements are accurate. As we see in Figure 6(a) the lighting measurements are all quite similar in form. As a result, we choose to use the Color by Correlation algorithm described in Section 3.2. This algorithm requires that we convert the RGB color coordinates of our images into discretized chromaticity coordinates. We choose to use x and y dimensions of the xyY color space, as defined by CIE; the chromaticity gamut defined by this chromaticity space is shown in Figure 7(a).

To demonstrate the color shift caused by lighting, we plot the discretized chromaticity coordinates of the MacBeth Color Chart from the first and final samples of the experiment in Figure 7(b). Here we have chosen to partition the xy plane into a grid of 32×32 discrete chromaticity coordinates, as suggested by [16]. Each dot represents a discrete coordinate that contains at least one pixel. Since



(a) CIE Chromaticity Diagram (b) Chromaticity shift of the MacBeth Color Chart

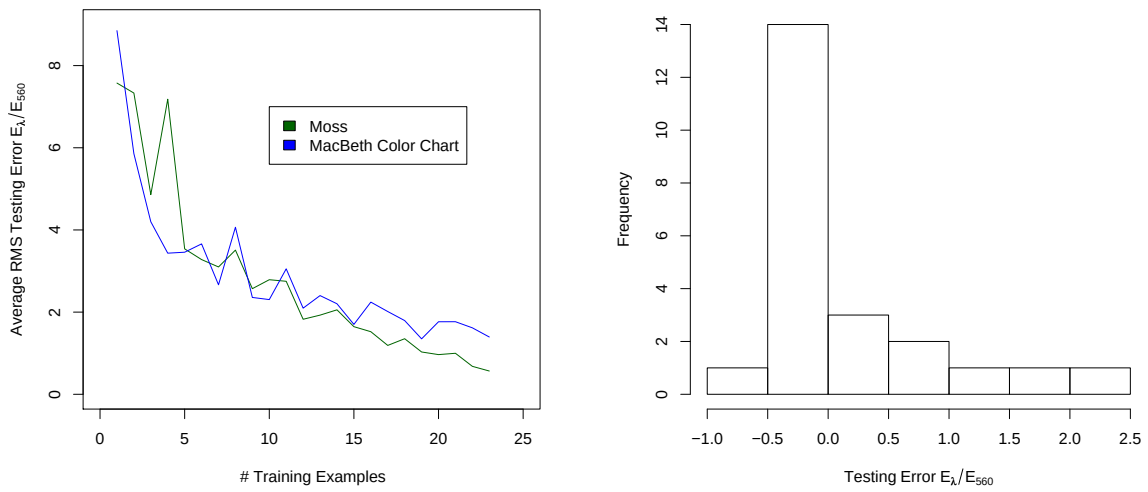
Figure 7: The standard chromaticity diagram shown in 7(a) [53], as defined by CIE [43]. 7(b) shows the chromaticity shift of the MacBeth Color Checker over the course of 6 hours illuminated by daylight. Both figures are shown in the xyY color space.

the image's subject is fixed, the shift in color can only be attributed to change in illumination. This is the effect we are attempting to remove.

The 2-dimensional chromaticity distributions of the sampled images are stored in matrices that we convert into row-major ordered vectors. Each vector is normalized by the number of pixels in the image and associated with the illumination measured using the spectroradiometer; these become the columns in the correlation matrix. This is a slightly simpler formulation of the Color by Correlation algorithm because we need not burden the model with the chromaticity distributions of other subjects under the same lighting; we have only one subject. To produce the log-likelihood that example image has been illuminated by particular illumination (Equation 8), we simply multiply the correlation matrix by the binary chromaticity vector. Recall, this binary vector is 1 for all chromaticity coordinates found in the example image and 0 elsewhere.

The training set for the Color by Correlation algorithm is selected at random from the set of experimentally obtained samples. We hand segmented the images from both cameras into an images containing only the moss and images only containing the MacBeth Color Checker. Figure 8(a) shows average RMS residual error between the predicted illumination and the measured illumination as a function of the training set size. Interestingly, for large training set sizes ($n >= 16$), the moss images had a slightly lower error than the images containing the MacBeth Color Chart. This is odd because the moss' reflectance is changing over time, where as the chart's reflectance is constant. In these cases approximately 70% of samples were used for training, so we believe this is simply an effect of over-training the model.

Though not shown, the model trained using both raw and JPEG images taken from the Canon camera produced similar residual error. This interesting result shows that JPEG compression has a minimal effect on the accuracy of the Color by Correlation algorithm when applied to these data. We can understand why by considering how JPEG compression works. First, it converts the image into the $YCbCr$ color space, which has two chromaticity dimensions and one brightness dimension similar



(a) Error with changing training set size

(b) Residual error with 12 training examples

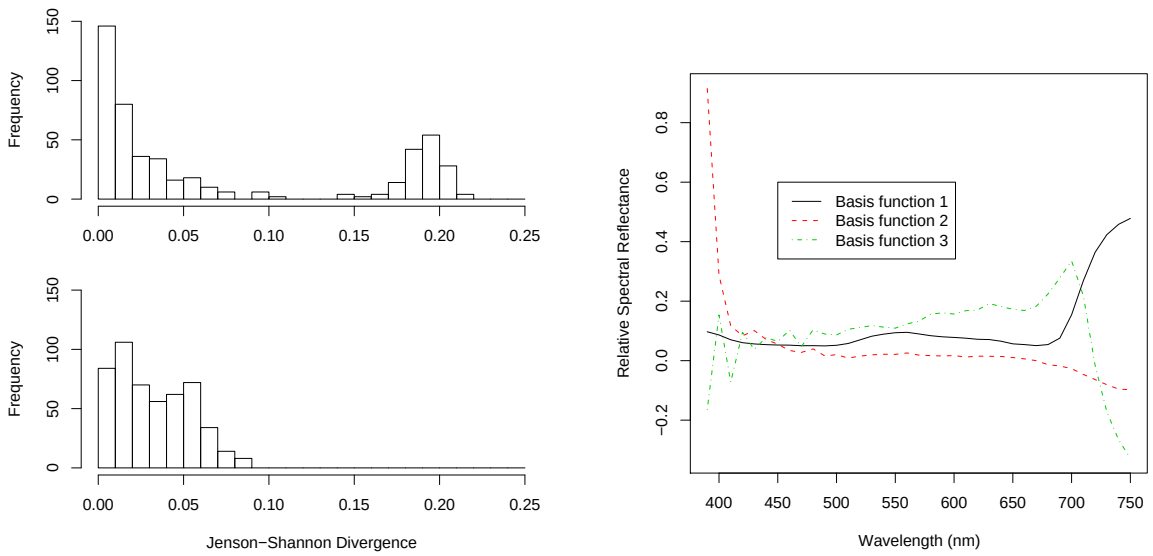
Figure 8: Error of the Color by Correlation model derived from images of moss and the MacBeth Color Checker taken with the Pentax Optio S5z camera under varying illumination. 8(a) shows the reduction in error as the size of the training set increases. 8(b) shows the histogram of testing error on moss examples for a training set size of 12 images.

to the xyY color space we used to train our model. Next, it computes the discrete 2-dimensional cosine transform of each 8×8 pixel block in the image. This produces the spatial frequency of colors within a given image block. Leveraging the fact that humans are more sensitive to lower frequency variations in color and brightness, JPEG compression discards some information about the high frequency signals, retaining most information about the lower frequency signals [11]. For both the moss and the MacBeth Color Chart, the spacial frequency in all three color dimensions is relatively low. This suggests that JPEG compression would have minimal effect on the chromaticity-based signals we are using to build our model.

Once we obtain an accurate estimate of the image's lighting we can correct for that illuminant using T_{light} (see Equation 11). We test this transform on our segmented images containing the MacBeth Color Chart because its spectral reflectance doesn't change (unlike the moss). To visualize the results of this transform, we choose to compute the 2-dimensional Jenson-Shannon Divergence (a symmetrical version of the Kullback-Leibler divergence [30]) of the discretized chromaticity coordinates. We compute this divergence for all pairs of examples and expect the divergences to small since the subjects are identical.

Histograms of these divergences are shown in Figure 9(a). As we can see, the lighting transformation compresses the distributions of divergences towards zero as expected. An unfortunate consequence is that it has also increased the variance among the previously well clustered examples. We hypothesize that this is caused by inherent error in estimating our sensors as impulse functions, a poor choice of center wavelengths, or color alterations resulting from JPEG image compression. Recall, we previously assumed that the camera's sensors were impulse functions (responding to a single center wavelength) such that we could compute a diagonal lighting transformation. In future work, we intend to try develop another algorithm, perhaps based on sensor sharpening, to further compress this divergence. Any improvement at this stage in our processing will improve the accuracy of our predictions.

After the images have been transformed we must predict the parameters of the relative spectral



(a) Pairwise divergence before (top) and after (bottom) re-lighting

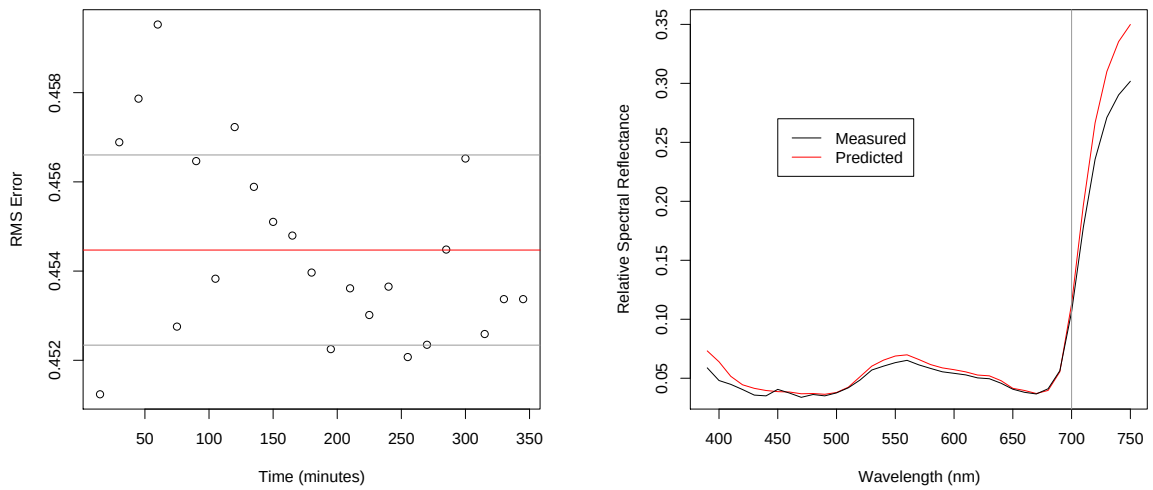
(b) Basis functions for relative spectral reflectance

Figure 9: 9(a) The Jenson-Shannon Divergence, before and after re-lighting, of all pairs of images containing the MacBeth Color Chard under varying daylight illumination. Optimally, all divergences would be zero after the lighting transformation. 9(b) The basis functions, as determined by functional PCA for the relative spectral reflectance of the moss as it dries over time.

reflectance model (shown in Figure 9(b)). We have chosen to use only the first three basis functions for our model because they account for 99.96% of the variance contained in the moss' measured relative spectral reflectance. The first basis function, plotted in black, represents the average spectral reflectance across all samples. The second and third basis functions, plotted in red and green respectively, show the type of variation seen. In particular, we see that there is significant variation in the blue (400nm – 450nm) and red (675nm – 750nm) parts of the spectrum. We expect some variation near 400nm because it is near the minimum wavelength our spectroradiometer can measure. It is not clear what caused the variation around 700nm. We suspect it was due to drift in the spectroradiometer's sensors during the course of the experiment.

Given this model, we must predict weights of these basis functions (w_s from Equation 4). We do this by training three regression-tree based models, one for each parameter, using the 2-dimensional chromaticity coordinates from the images previously registered by re-lighting. We trained this estimation model 12 samples, the same value which produced reasonable results for the lighting estimation. The RMS residual error of this prediction is shown in Figure 10(a). We can see that there is no meaningful spatial or temporal pattern in the error, suggesting the model captures most of the variation in the data. However, the magnitude of the error is rather large and we see some rather significant outliers. In comparison, the best possible values for w_s produce a mean RMS residual error of 0.0214, approximately 20 times smaller than the error produced by the spectral reflectance estimation model.

To better understand this error we plot the measured and estimated spectral reflectance for the largest outlier, sample 4 occurring at 60 minutes. As we can see in Figure 10(b), the fit is quite good. The vast majority of the error comes from wavelengths greater than 700nm. This error is somewhat expected since it is present in the model's basis functions as well as the original measurements. However, without evaluating the effect on the estimation of the target signal (CO_2 uptake), it is impossible to tell if this prediction accuracy is sufficient for our application. Regardless, there is



(a) Spectral reflectance model error

(b) Predicted spectral reflectance for observation 4

Figure 10: The RMS residual error of the spectral reflectance predicted by our procedure is shown in 10(a); the red line is the average error and the grey lines are the first standard deviation. In 10(b) we show the predicted spectral reflectance of the observation with the largest error (observation 4 at time 60 minutes).

certainly significant room for improvement. As suggested earlier, the error of this prediction can be reduced by improving the accuracy of the re-lighting transform.

5.3 Incorporating Prior Work

Prior ecological analysis [20] of this moss has produced very detailed measurements of CO_2 uptake in the laboratory. The moss was placed in a chamber under controlled temperature and lighting conditions. Ambient air was drawn through a series of tubes such that some passed through the chamber as a sample, and the rest was unaltered as a control. Using an infrared gas analyzer [7], the air that had passed through the chamber was compared to the control to compute the relative increase or decrease of CO_2 in the air by volume. The instantaneous CO_2 uptake of the moss ranged between -1 μmol m^{-2} s^{-1} and 6 μmol m^{-2} s^{-1} with an error of $\pm$ 0.5 μmol m^{-2} s^{-1}. Through the experiment, a florescent light was turned on and off at 12 hour intervals to simulate day and night. During the light periods, the moss was kept at 15 °C; during the dark periods, the moss was kept at 10 °C. In addition to the CO_2 and temperature measurements, PAR measurements and images of each sample were collected at 10 minute intervals. The experiment captured the moss' progression from hydrated to dry (known as a dry down) over the course of a eight days.

In addition to a series of moss dry downs, other measurements were taken to produce an environmental productivity index (EPI) [40] for this moss plant. This index posits that there are three factors that limit a plant from gaining carbon at its maximal rate: availability of light, availability of moisture, and suitable temperature. Further, it suggests that independently measuring the effect of each dimension on the plant's respiration is sufficient to reconstruct the plant's behavior in the field. For example, temperature was varied from -1 °C to 34 °C while the moss was kept moist and well lit. Then, by simply multiplying the limiting factors together we can predict the approximate percentage reduction in CO_2 uptake as compared to the moss' maximum absorption.

The problem with using this model in the field is the lack of information about the moss' moisture levels; temperature and PAR are easily measured. As a surrogate, we intend to use the moss' relative spectral reflectance, which has been shown to be correlated with moisture as well as CO_2 uptake [19]. Thus, to leverage the data previously collected, we must compute the relative spectral reflectance of the moss in the images during the dry down. Fortunately, we know the character of the incident light in the images as it was produced by a florescent bulb. By re-lighting the images to be under the D_{65} illuminant, we can use the reflectance model produced earlier to estimate the relative spectral reflectance of the sample. Finally, we can model CO_2 as a function of temperature, PAR, and relative spectral reflectance using our modified non-linear regression algorithm.

In the future, another experiment must be run to validate this model. In particular, the constant "day" and "night" temperature during the previous experiment is particularly unrealistic. The next experiment will attempt to measure CO_2 under more realistic (less controlled) conditions as suggested by Section 4.3. Such an experiment would be performed in the presence of representative natural light and ambient temperature. Because of the difficulty involved in running such an experiment, it will be of limited duration and provide less modeling utility as compared previous experiments. However, it will serve to produce a dataset we can use to evaluate the prediction accuracy of our models.

6 Other Applications

There are a number of applications that could benefit from using imagers as sensors. On example is estimating CO_2 uptake for an entire meadow already instrumented with NEON (National Ecological Observatory Network) [39] flux towers. These devices have a number of meteorological sensors in addition to a camera. Though the meadows typically contain more than one type of plant, it seems that simple color based image segmentation could partition the image accurately. Then, we could apply our existing methodology to the sub-images to measure the CO_2 uptake of each plant in the scene.

Like meadows, the soil beneath the forest canopy is hypothesized to emit and absorb a significant quantity of CO_2. As a result, ecologists have developed accurate models of soil CO_2 uptake based on soil surface temperature. However, we cannot measure soil surface temperature with the density required to produce an accurate prediction of CO_2 uptake of the soil. It is known that soil surface temperature lags air temperature (which we can easily measure) but is also significantly effected by direct sunlight. Using tower mounted cameras, we can image the soil below the forest canopy looking for sun flecks that have a distinctive color signature (they are saturated). Combining the localized sun flecks and ambient air temperature, we can predict a 2-dimensional map of surface temperature to feed into the CO_2 models.

Similar to the river depth measurement system [31], in-situ imagers can be used to measure particulate matter suspended in a river. When present in reasonable quantities, these particulates have an effect on the river's visible color. Measuring these particulates, many of which are contaminants, can help ecologists understand the health of the local ecosystem. Enabling ecologists to simply photograph the river, instead of performing direct measurement, will help increase the sampling coverage and suggest areas that will most benefit from more accurate measurement. This will help save effort, money, and time.

7 Work Plan and Timeline

- **Reformulate and further test re-lighting:** Collect more images and spectral power distribution curves under various natural light. Modify the re-lighting algorithm to be more robust to these changing conditions. – *2 Months*

- **Further generalize regression modifications:** Enable the use of arbitrarily complex mixtures of location-scale distributions from the same family. – *2 Months*
- **Complete evaluation of moss photosynthesis application:** Apply the procedure described here to one year of data collected by at James Reserve, producing a CO_2 signal. Evaluate the model's accuracy for this application using the model validation techniques we have described – *3 Months*
- **Build online system for predicting moss photosynthesis:** Provide measurements of the moss' CO_2 update given real time images captured at James Reserve – *1 Month*
- **Apply procedure to another application** Instantiate and evaluate our procedure targeting one of the applications mentioned in Section 6 – *6 Months*
- **Dissertation Writing:** *2 – 3 Months*

References

[1] Shaun Ahmadian, Teresa Ko, Sharon Coe, Mohammad Rahimi, Stefano Soatto, Michael Hamilton, and Deborah Estrin. A vision system to infer avian nesting behavior. Technical report, University of California, Los Angeles, 2007.

[2] Kobus Barnard. *Practicle Color Constancy*. PhD thesis, 2000.

[3] Kobus Barnard, Vlad Cardei, and Brian Funt. A comparison of computational color constancy algorithms. Part I: Methodology and experiments with synthesized data. *IEEE Transactions on Image Processing*, 11(9):972–984, 2002.

[4] Kobus Barnard, Florian Ciurea, and Brian Funt. Sensor sharpening for computational color constancy. *Journal of the Optical Society of America A*, 18(11):2728–2743, 2001.

[5] R.E. Bellman. *Dynamic Programming*. Princeton University Press, 1957.

[6] Licor Biosciences. `http://www.licor.com/env/Support/discontinued/li1800.jsp`.

[7] Licor Biosciences. `http://www.licor.com/env/Products/GasAnalyzers/li6262/6262.jsp`.

[8] L. Breiman, J.H. Friedman, R.A. Olshen, and C.J. Stone. *Classification and Regression Trees*. Chapman and Hall, 1984.

[9] Ralph B. Brown and Scott D. Noble. Site-specific weed management: sensing requirements – what do we need to see? *Weed Science*, 53(2):252–258, 2005.

[10] Canon. `http://www.usa.canon.com/consumer/controller?act=ModelInfoAct\&fcategoryid=139\&modelid=16303`.

[11] JPEG Compression. `http://www.w3.org/Graphics/JPEG/itu-t81.pdf`.

[12] Michael A. Crimmins and Theresa M. Crimmins. Monitoring Plant Phenology Using Digital Repeat Photography. *Environmental Management*, 41:949–958, 2008.

[13] Richard P. Feynman, Robert B. Leighton, and Matthew Sands. *The Feynman Lectures on Physics*, chapter 35: Color Vision. Addison-Wesley, 1963.

[14] Graham D. Finlayson. Color in Perspective. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18:1034–1038, 1996.

[15] Graham D. Finlayson and Steven D. Hordley. Color constancy at a pixel. *Journal of the Optical Society of America A*, 18(2):253–264, 2001.

[16] Graham D. Finlayson, Steven D. Hordley, and Paul M. Hubel. Colour by correlation: a simple, unifying approach to colour constancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23:1209–1221, 2001.

[17] D.A. Forsynth. A novel algorithm for color constancy. *International Journal of Computer Vision*, 5(1):5–36, 1990.

[18] C. Fraley and A.E. Raftery. Model-Based Clustering, Discriminant Analysis, and Density Estimation. *Journal of the American Statistical Association*, 97(458):611–632, 2002.

[19] A.A. Gitelson, Y. Gritz, and M.N. Merzlyak. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *Journal of Plant Physiology*, 160(3):271–282, 2003.

[20] Eric A. Graham, Michael P. Hamilton, Brent D. Mishler, Philip W. Rundel, and Mark H. Hanson. Use of a Networked Digital Camera to Estimate Net CO_2 Uptake of a Desiccation-Tolerant Moss. *International Journal of Plant Sciences*, 167:751–758, 2006.

[21] F.S. Hill. *Computer Graphics Using OpenGL*, chapter 12: Color Theory. Prentice Hall, second edition, 2001.

[22] Steven D. Hordley and Graham D. Finlayson. Reevaluation of color constancy algorithm performance. *Journal of the Optical Society of America A*, 23(5):1008–1020, 2006.

[23] A.R. Huett. A soil-adjusted vegitation index (SAVI). *Remote Sensing of Environment*, 25:53–70.

[24] Josh Hyman, Eric Graham, Mark Hansen, and Deborah Estrin. Imagers as sensors: Correlating plant CO_2 uptake with digital visible-light imagery. *Workshop on Data Management in Sensor Networks*, 2007.

[25] Norman Johnson, Samuel Kotz, and N. Balakrishnan. *Continuous Univariate Distributions*. Wiley, 2004.

[26] D.B. Judd, D.L. MacAdam, G. Wyszecki, et al. Spectral distribution of typical daylight as a function of correlated color temperature. *Journal of the Optical Society of America*, 54(8):1031–1040, 1964.

[27] D.E. Karcher and M.D. Richardson. Quantifying Turfgrass Color Using Digital Image Analysis. *Journal of Crop Science*, 43(3):943, 2003.

[28] Y.J. Kaufman and D. Tanre. Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. *IEEE Transactions on Geoscience and Remote Sensing*, 30(2):261–270, 1992.

[29] Teresa Ko, Stefano Soatto, and Deborah Estrin. Background Subtractions with Distributions. *European Conference on Computer Vision*, 2008.

[30] S. Kullback. *Information Theory and Statistics*. Wiley, 1959.

[31] Carl J. Legleiter, Dar A. Roberts, Andrew Marcus, and Mark A. Fonstad. Passive optical remote sensing of reiver channel morphology and in-stream habitat: Physical basis and feasibility. *Remote Sensing of the Environment*, 93:493–510, 2004.

[32] Y. Liu, J.A. Sarnat, B.A. Coull, P. Koutrakis, and D.J. Jacob. Validation of MISR Aerosol Optical Thickness Measurements Using AERONET Observations over the Contiguous United States. *Journal of Geophysical. Research*, 2003.

[33] Laurence T. Maloney and Brian A. Wandell. Color constancy: a method for recovering surface spectral reflectance. *Readings in Computer Vision: Issues, Problems, Principles, and Paradigms*, 1987.

[34] J.A. Marchant and C.M. Onyango. Shadow-invariant classification for scenes illuminated by daylight. *Journal of the Optical Society of America A*, 17(11):1952–1961, 2000.

[35] J.A. Marchant and C.M. Onyango. Spectral invariance under daylight illumination changes. *Journal of the Optical Society of America*, 19(5):840–848, 2002.

[36] J.A. Marchant, N.D. Tillett, and C.M. Onyango. Dealing with Color Changes Caused by Natural Illumination in Outdoor Machine Vision. *Cybernetics and Systems*, 35(1):19–33, 2004.

[37] David H. Marimont and Brian A. Wandell. Linear models of surface and illuminant spectra. *Journal of the Optical Society of America A*, 9(11):1905–1913, 1992.

[38] M. Mirik, G.J. Michels Jr., S. Kassymzhanova-Mirik, N.C. Elliott, V. Catana, D.B. Jones, and R. Bowling. Using digital image analysis and spectral reflectance data to quantify damage by greenbug (Hemitera: Aphididae) in winter wheat. *Computers and Electronics in Agriculture*, 51:86–98, 2005.

[39] National Ecology Observation Network. `http://www.neoninc.org`.

[40] P.S. Nobel. PAR, Water, and Temperature Limitations on the Productivity of Culitvated Agave Forcroydes (Henequen). *Journal of Applied Ecology*, pages 157–173, 1985.

[41] Ohta Noboru and Alan R. Robertson. *Colorimetry*, chapter 3.9: Standard and Supplementary Illuminants. Wiley, 2005.

[42] Melvin J. Oliver, Jeff Velten, and Andrew J. Wood. Bryophytes as experimental models for the study of environmental stress tolerance: *Tortula ruralis* and desiccation-tolerance in mosses. *Plant Ecology*, pages 73–84, 2000.

[43] International Commission on Illumination. `http://www.cie.co.at`.

[44] Pentax. `http://www.pentaximaging.com/files/manual/OptioS5z_web.pdf`.

[45] J.O. Ramsay and B.W. Silverman. *Functional Data Analysis*. Springer, 1997.

[46] J.W. Rouse, R.H. Hass, J.A. Schell, and D.W. Deering. Monitoring the Vernal Advancement and Retrogradation (Greenwave Effect) of Natural Vegetation. *Proceedings of the Remote Sensing Center*, 1974.

[47] T. Shi, E.E. Clothiaux, B. Yu, A.J. Braverman, and D.N. Groff. Detection of daytime arctic clouds using misr and modis data. *Remote Sensing of Environment*, 2006.

[48] D.C. Slaughter, D.K. Giles, and D. Downey. Autonomous robotic weed control systems: A review. *Computers and Electronics in Agriculture*, 61(1):63–78, 2008.

[49] Albert J.P. Theuwissen. *Solid-State Imaging with Charge-Coupled Devices*. Kluwer Academic Print on Demand, 1995.

[50] Zoltan Tuba, Zsolt Csintalan, and Michael Proctor. Photosynthetic response of a moss, *Tortula ruralisi*, ssp. *ruralis*, and the lichens *Cladonia convolta* and *C. furcata* to water deficit and short periods of desiccation, and their ecophysiological significance: a baseline study at present-day CO_2 concentration. *New Phytologist*, pages 353–361, 1996.

[51] Susan L. Ustin, Dar A. Roberts, John A. Gamon, Gergory P. Asner, and Robert O. Green. Using Imaging Spectroscopy to Study Ecosystem Processes and Properties. *BioScience*, 54(6):523–534, 2004.

[52] Brian A. Wandell. *Foundations of vision*. Sinauer Associates, 1995.

[53] Wikipedia. `http://en.wikipedia.org/wiki/CIE_1931`.

[54] J.A. Worthey and M.H. Brill. Heuristic analysis of von Kries color constancy. *Journal of the Optical Society of America A*, 3(10):1708–1712, 1986.