CENTER FOR EMBEDDED NETWORKED SENSING

Imagers as sensors:

Correlating plant CO_2 uptake with digital visible-light imagery

Josh Hyman, Eric Graham

Mark Hansen, Deborah Estrin

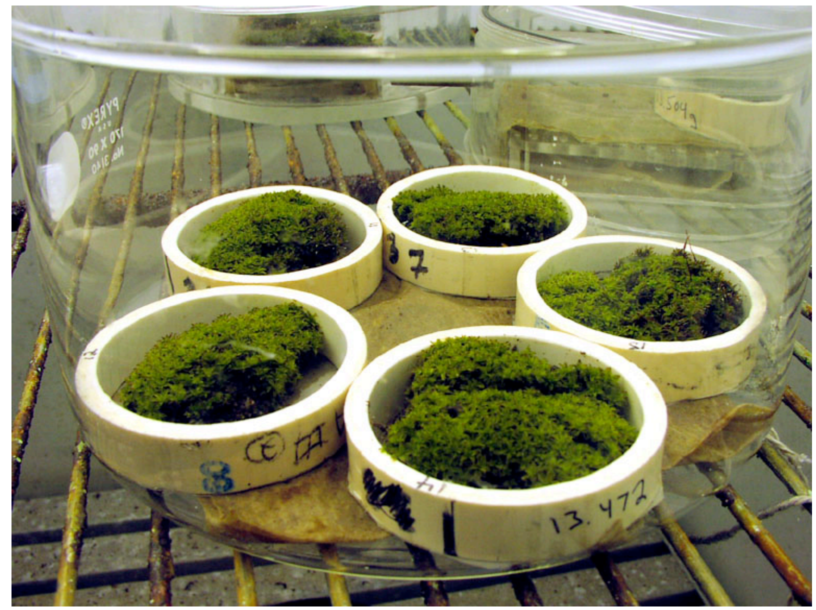
Center for Embedded Networked Sensing

University of California, Los Angeles

UCLA USC UCR CALTECH UCM

The problem

- Estimate CO_2 uptake of a drought tolerant moss
  - we measure moss because there is extensive biological knowledge to inform our model
- Build model from ground truth and images gathered under laboratory conditions

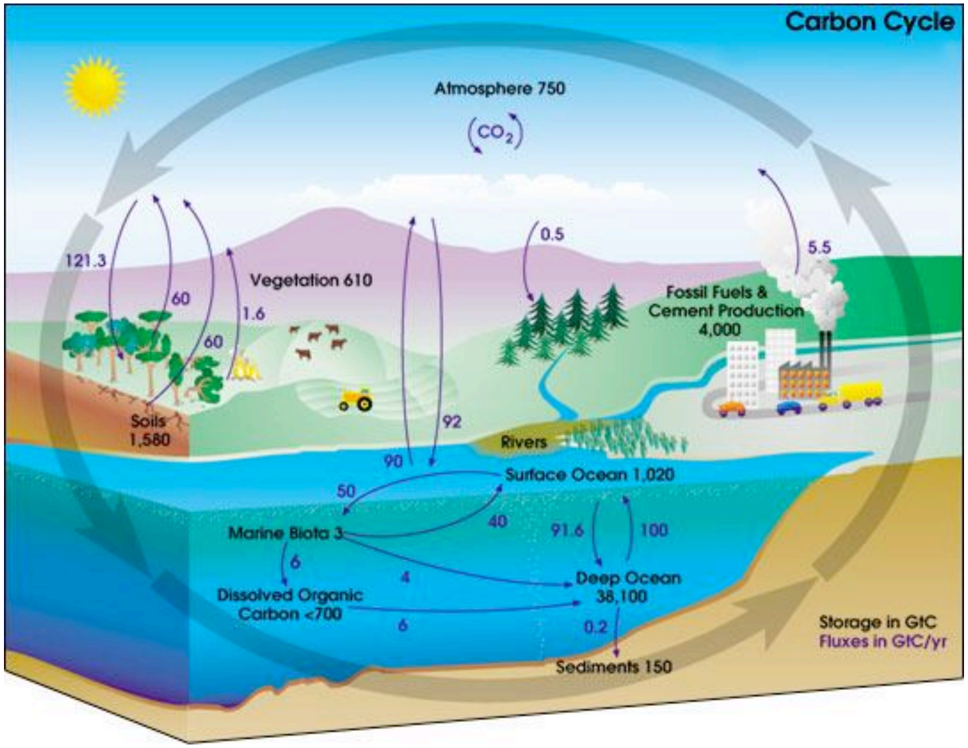


Such a model adds value to MossCam database which contains historical images of this plant in the field

Why measure CO_2 uptake?

CO_2 uptake is the quantity of CO_2 (ppm/m^2/sec) absorbed or released by a plant during photosynthesis

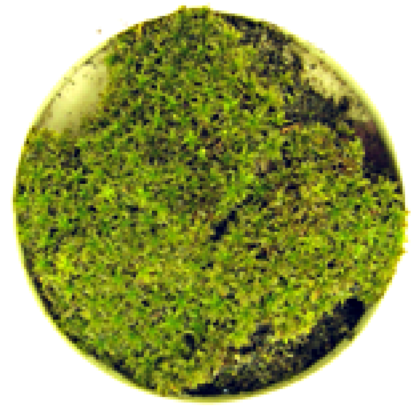
- Dense measurement of plant CO_2 uptake can be extrapolated to entire forests
- Understanding the CO_2 uptake of an entire forest can help model global CO_2 levels



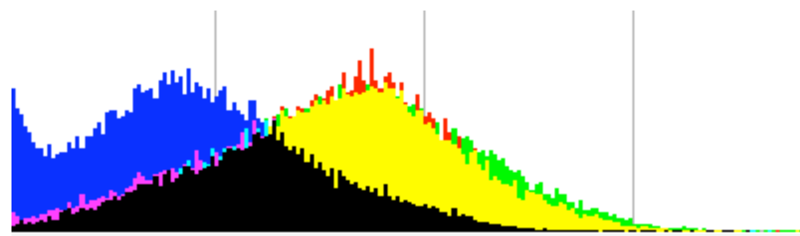
Source: http://en.wikipedia.org/wiki/Carbon_cycle

Using applied vision

- Not general vision
  - ground truth not present in the images themselves
  - humans can't discern CO_2 from images of plants
- Applied vision
  - use features similar to those used by general vision
  - features as input to a model which learn the underlying phenomena



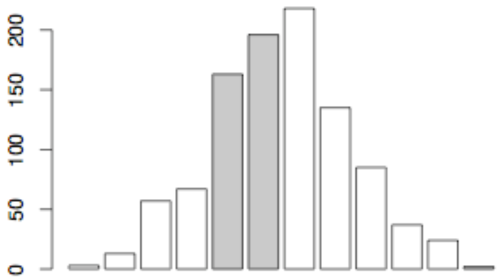
Original Image



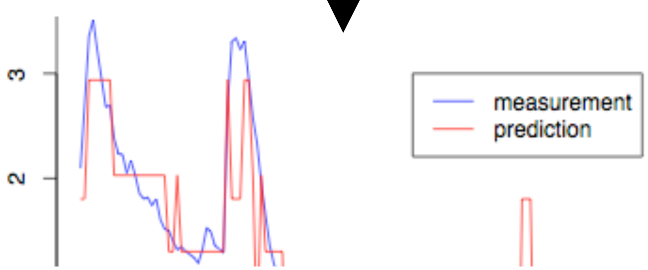
Extracted Color Histogram

Method for building sensing imagers

Start with images taken of a biological event stored in a database



Compute domain relevant image feature set



Find the most correlated image features and generate a model from this set to predict the biological event

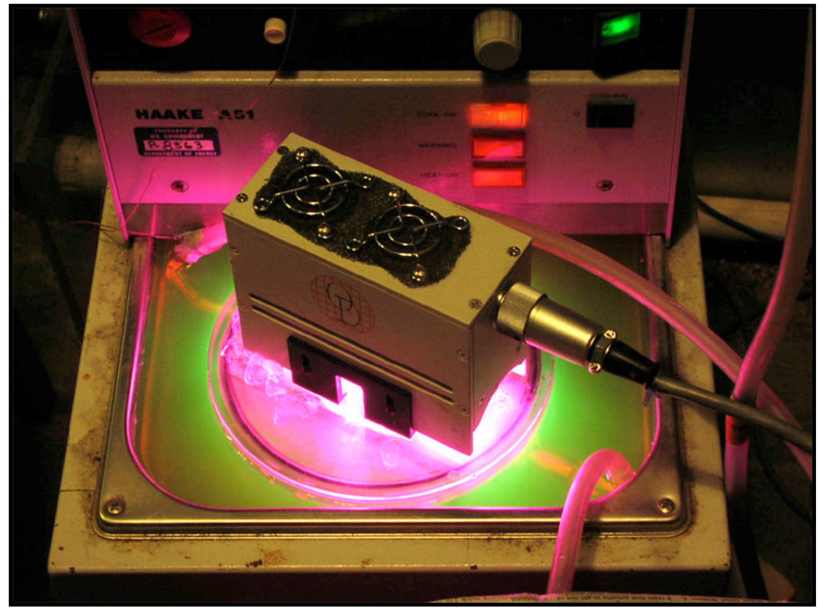
Experimental design

Procedure:

1. Water the moss
2. Cycle light on and off
3. Repeat until the moss is dormant

Gathered Data:

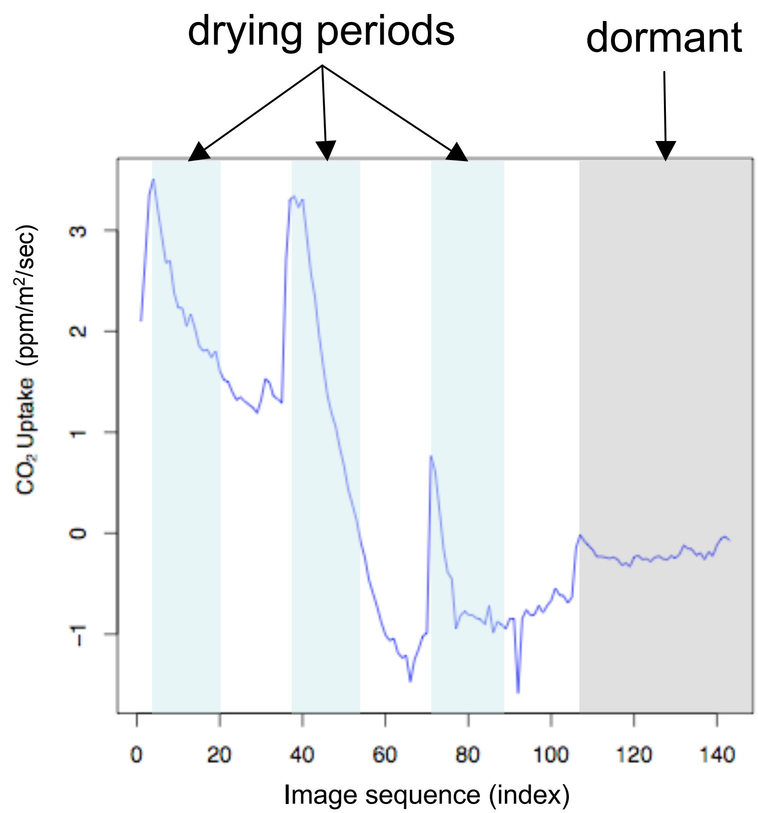
1. Ground truth: the difference between CO_2 in the intake versus the exhaust
2. Images of the moss plant which coincide with the CO_2 uptake measurements



- Controlled lighting
- Controlled moisture
- CO_2 measured using spectroscopy (measurement error: 0.1 ppm)

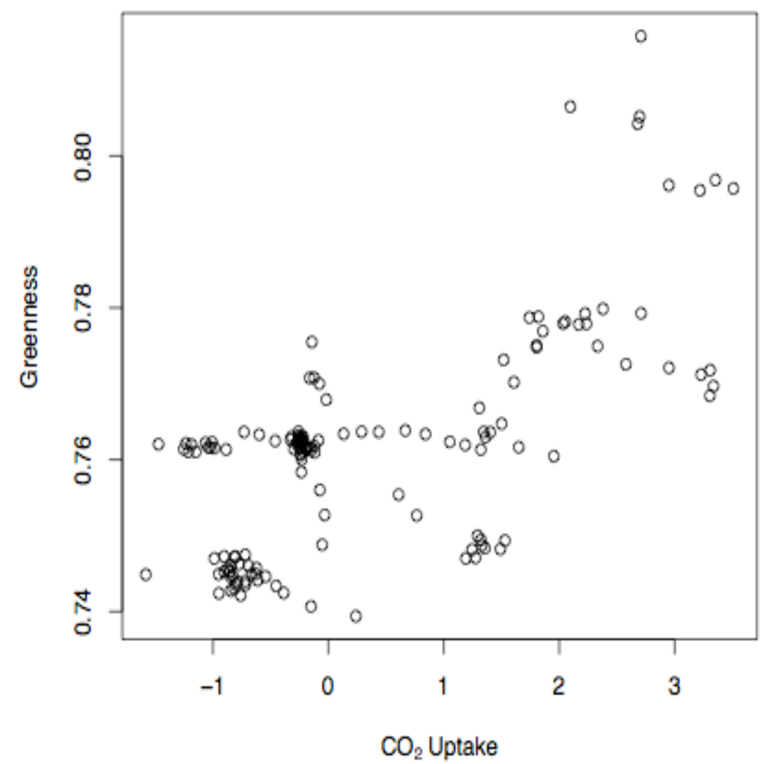
Characteristics of CO_2 uptake

- Drying cycle
  - water evaporates when warmed by light
  - redistributes water when dark
- Once completely dry, becomes dormant until it is watered
- CO_2 uptake dependant on moisture



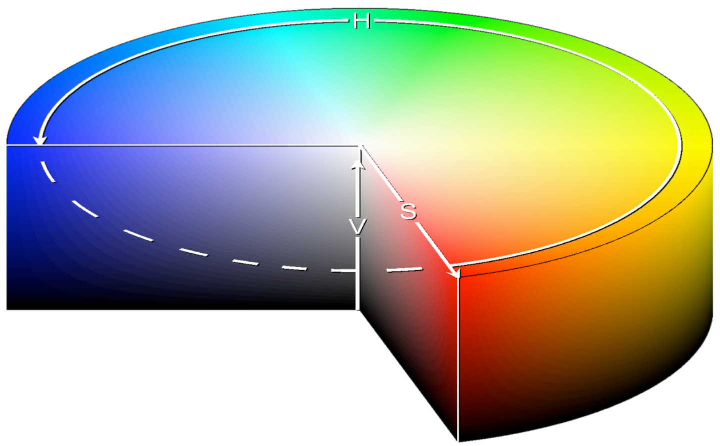
What does domain knowledge suggest?

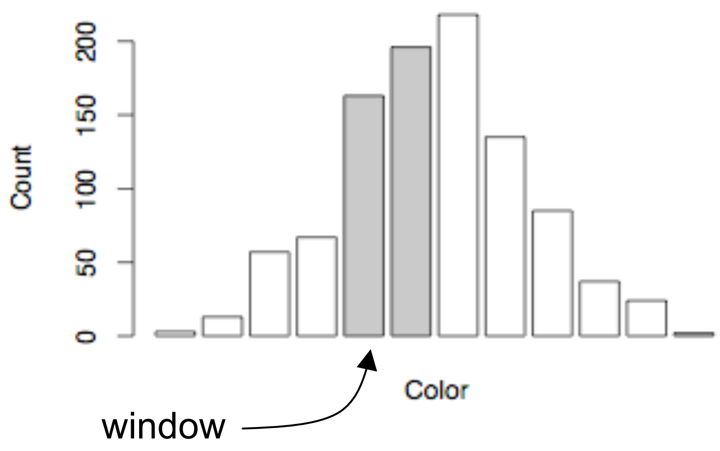
- Domain knowledge suggests that color (specifically greenness) is a good predictor
- Compute Euclidian distance between image's color histogram and a reference green histogram
- Plot shows the beginnings of a trend



Extracting image features

- Compute HSV (Hue, Saturation, Value) histogram
  - more stable than RGB
  - inexpensive to compute
- Compute a set of variable sized windows, grouping similar colors





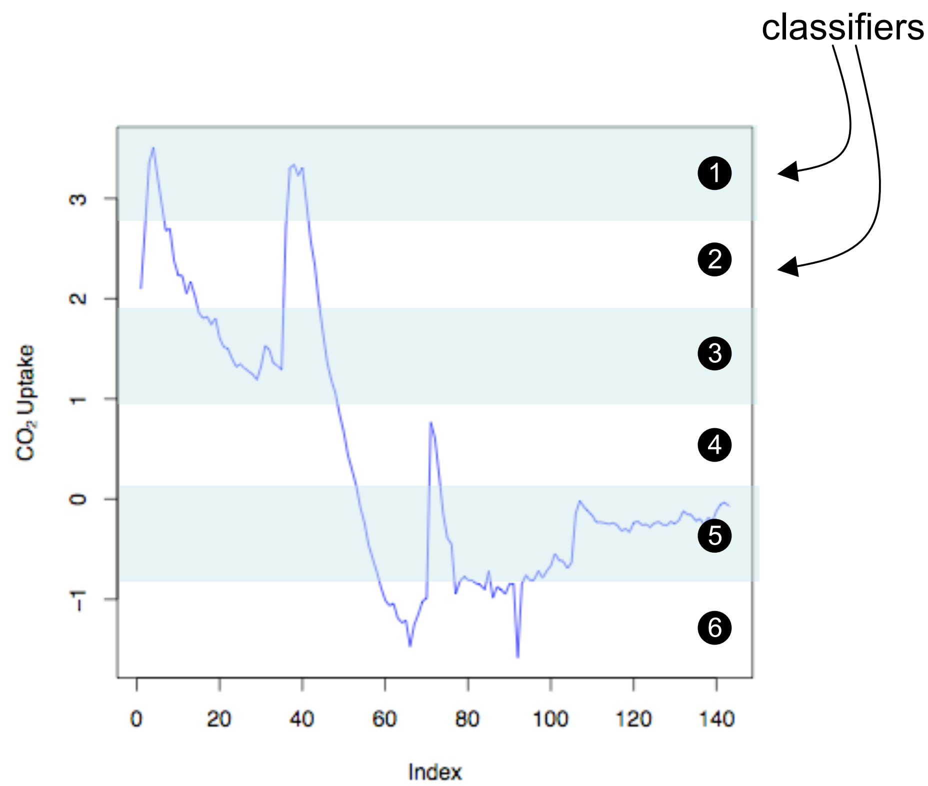
Building a model

- We consider building 2 different kinds of models based on the data collected in the lab
  - Classification based model
  - Regression based model
- The models are trained on features extracted from the images and are trying to estimate the coinciding CO_2 measurement

Building a model: Classifier

Build 6 binary classifiers

- Divide values into 6 equal-sized bands
- Train 6 SVM based classifiers using all generated features



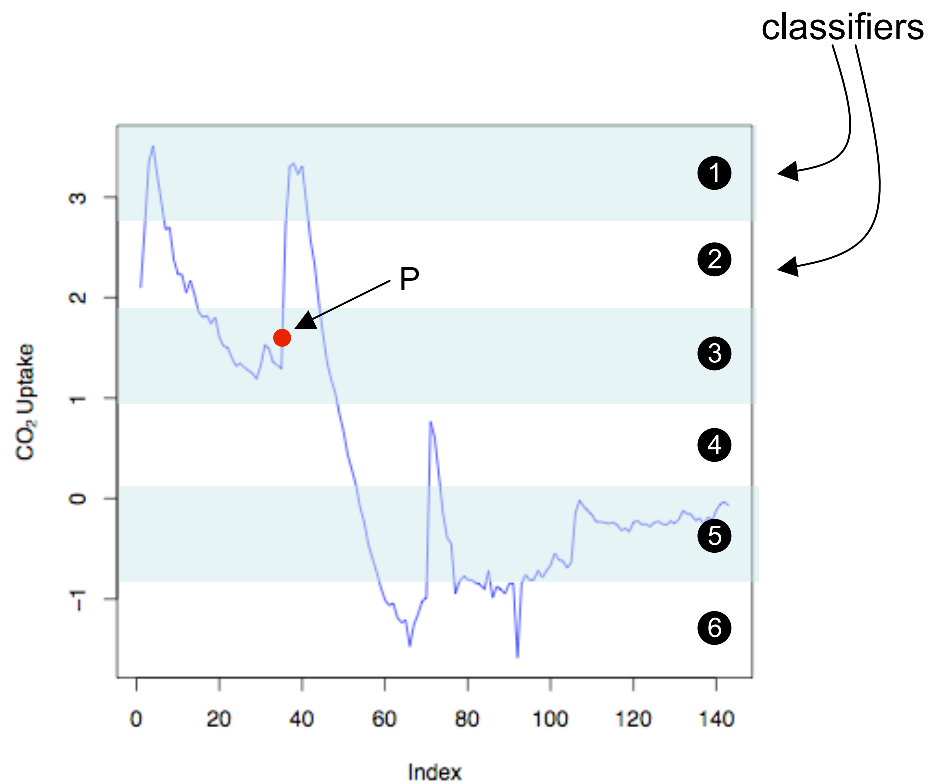
Why 6 classifiers?

- More than 6 results in over-fitting
- Less than 6 leads to insufficient accuracy

Building a model: Classifier

Classifying data

- A point P is in class K if the Kth binary classifier responds most strongly
- The point P is assigned the median value of class K
- Response is the distance to the hyper-surface separating class K from all other points



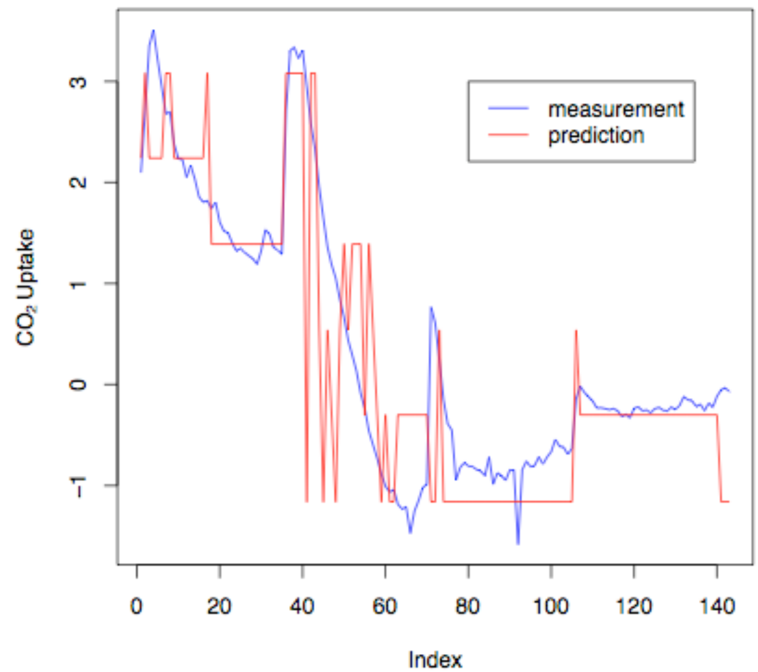
Classifier	1	2	3	4	5	6
Response	-2.19	1.54	3.08	2.33	-1.37	-3.21

Results: Classifier

Reasons for high error

- too few classes to cover all drying states
- many classes had too little training data
- classes of fixed size, doesn't reflect reality

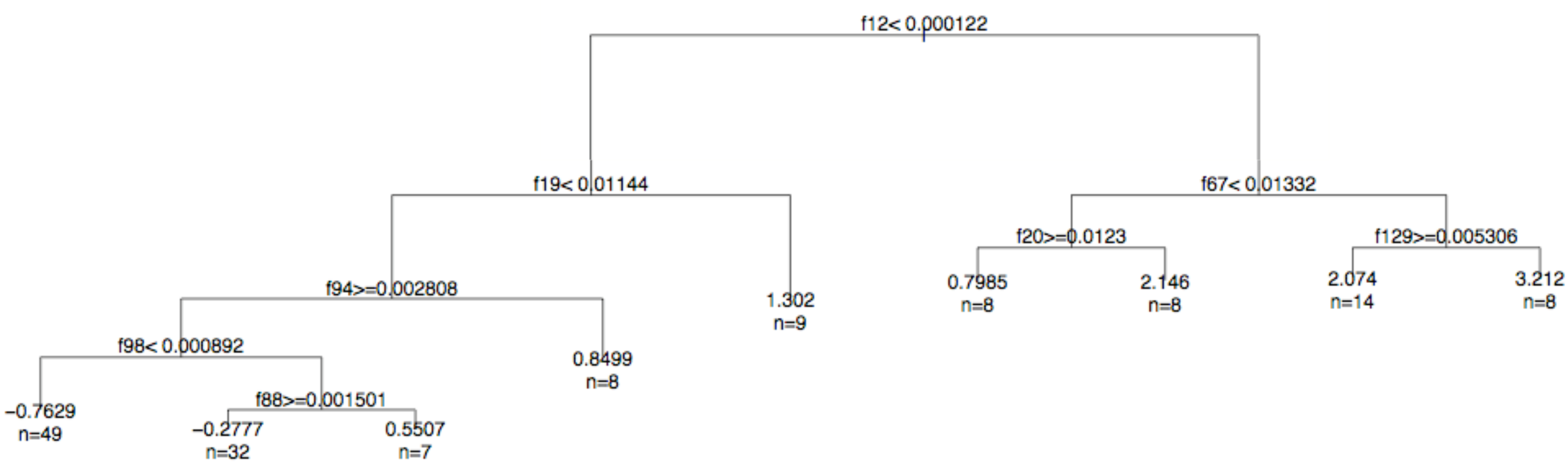
Not particularly meaningful to compare distance in different features spaces



Sensor measurement error	0.10 ppm
Acceptable error	0.50 ppm
Model's RMS error	0.74 ppm

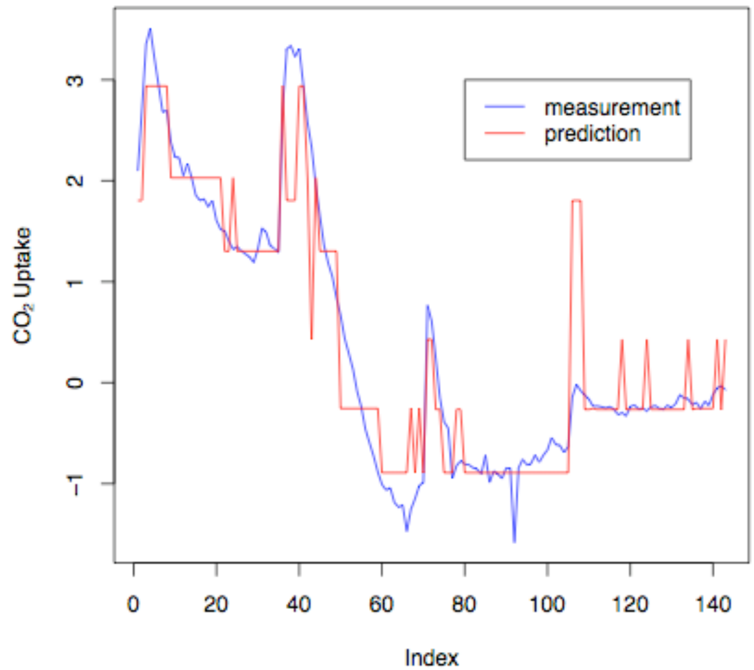
Building a model: Regression

- Recursively choose a feature and corresponding threshold that increases purity
- At each leaf node, all remaining points assigned the average of the corresponding CO_2 uptake value
  - Effectively creates variable sized pseudo-bands based on similarity



Results: Regression

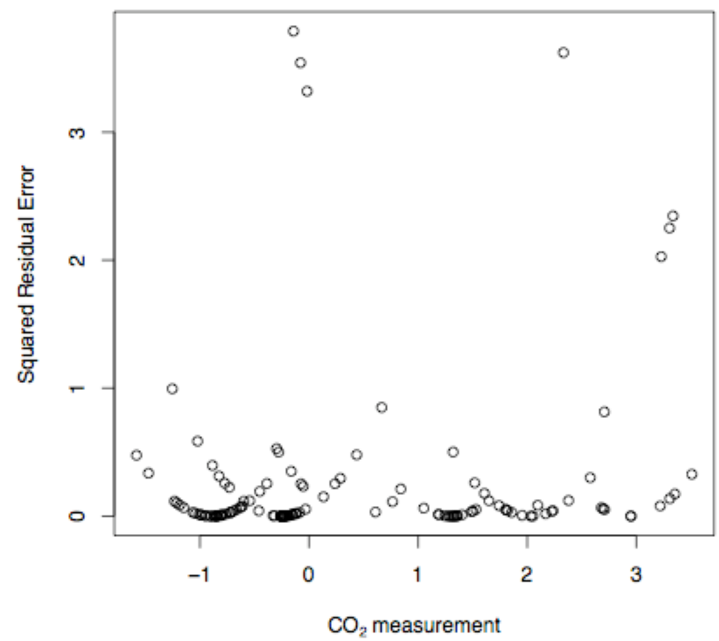
- Count (9) and width of "bands" chosen intelligently
  - More accurate prediction
  - Data adaptive rather than statically chosen
- Errors
  - Final state color similar to drying moss color, causing confusion
  - Unique or rarely seen values have larger error



Sensor measurement error	0.10 ppm
Acceptable error	0.50 ppm
Model's RMS error	0.49 ppm

Error Analysis

- The squared error is parabolic-shaped because one value is assigned to each “band”
- There is 1 parabolic error curve for each “band” in the model



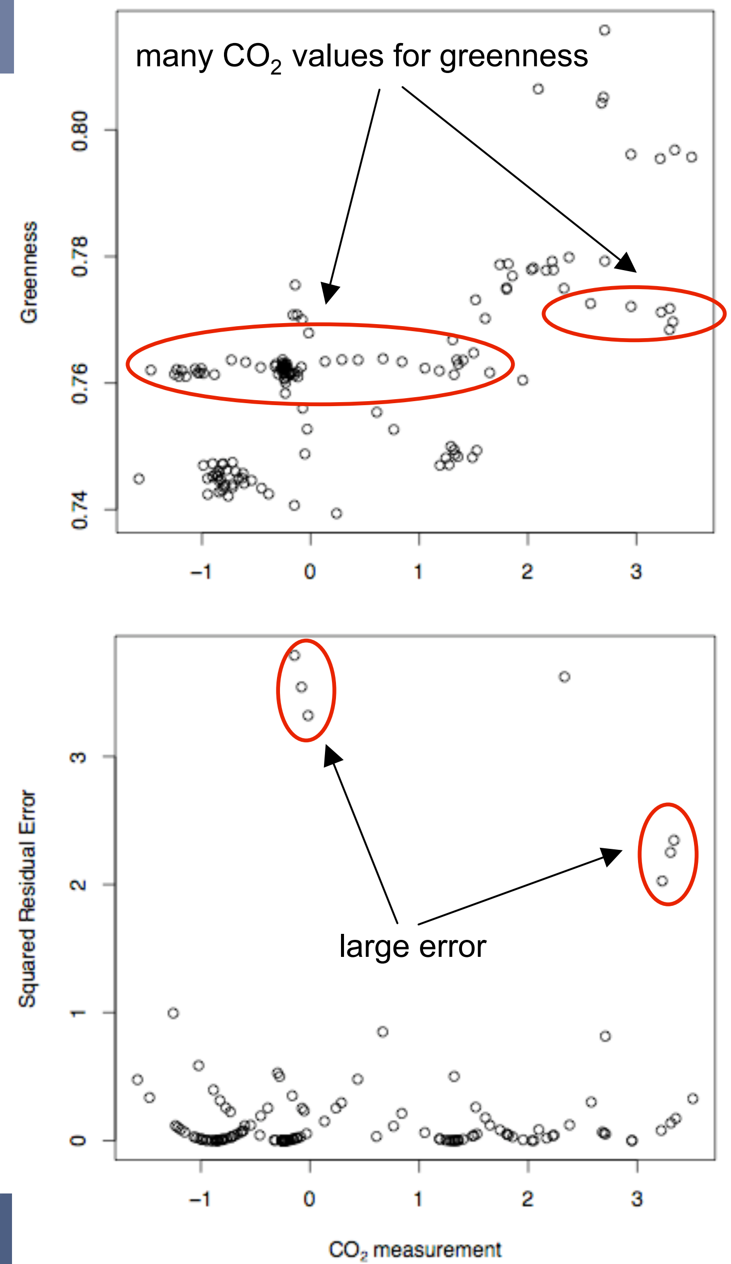
Error Analysis

CENTER FOR EMBEDDED NETWORKED SENSING

- Likely locations of high error revealed by rudimentary greenness measure
- Many different CO_2 measurements for a greenness value makes prediction more difficult

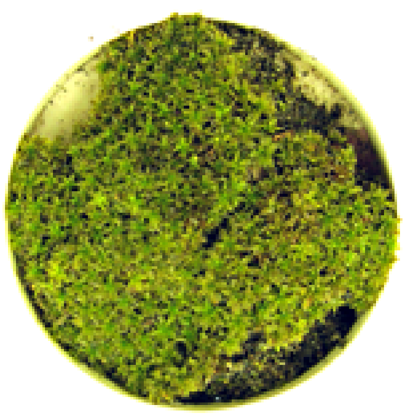
Biological reason:

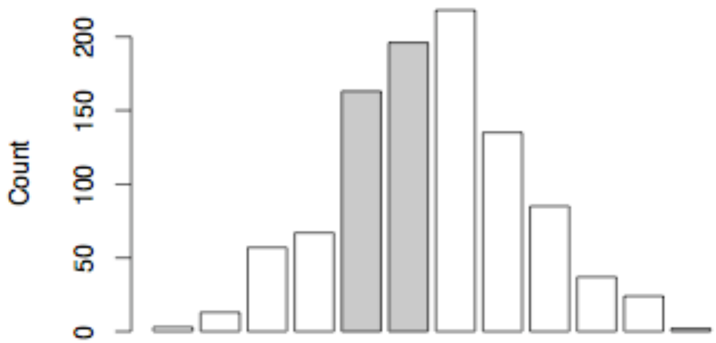
- Same approximate color for different stages of drying
- Moss drying stage dictates CO_2 uptake



Computational requirements

- Feature computation
  - time and space linear in the number of pixels
  - constant factor improvement by down-sampling HSV histogram





Computational requirements

- SVM model application
  - requires complex model application

$$\max \sum_{i=1}^{n} \alpha_i - \sum_{i,j} \alpha_i \alpha_j c_i c_j \mathbf{x}_i^T \mathbf{x}_j \text{ subject to } \alpha_i \geq 0.$$

  - significant storage required for a large set of support vectors used in the model
- Regression tree model application
  - requires only numerical comparison
  - minimal storage needed to store tree structure

Model comparison

	Classification Model	Regression Model
Feature Computation Complexity	Low	Low
Storage Requirements	High	Low
Application Complexity	Medium	Low
Testing RMS Error	0.74 ppm	0.49 ppm

Overall, the regression model performed better than the classification model with respect to many different axes of comparision

Isn't this like Remote Sensing?

- Uses in-situ imagers
  - rapid deployment
  - more dense image coverage (sub-pixel resolution in comparison)
  - can faster react to environmental changes
- Complementary measurement
  - an in-situ imaging sensor deployment can be triggered by events captured by remote sensing
- Minimal post-processing
- Models simple enough to be applied on sensors themselves eventually

Source: http://www.geo.mtu.edu/

Future Direction

- Push computation onto sensor nodes
  - perform model application on sensor nodes
  - turn imagers into first class sensors capable of producing biological measurements directly
- Extend to field imagery
  - changing lighting (eg. clouds) effects color
  - must compensate for variable white balance
- Use external sensors to inform model
  - temporal nature of drying could be used if rain events can be detected
  - PAR (light) sensor can help fix white balance

Questions?