B41K-0196: Standardizing PhenoCam Image Processing and Data Products.

Tom Milliman[1], Andrew Richardson[2], Stephen Klosterman[2], Josh Gray[3], Koen Hufkens[2], Donald Aubrecht[2], Min Chen[2], Mark Friedl[3]

1 – *University of New Hampshire; 2 – Harvard University; 3 – Boston University*

PhenoCam Project Overview (http://phenocam.unh.edu)

Vegetation phenology controls the seasonality of many ecosystem processes, as well as numerous biosphere-atmosphere feedbacks. Phenology is also highly sensitive to climate change and variability. Our data are derived from conventional, visible-wavelength digital camera imagery collected through the PhenoCam network. Time series data derived from this imagery can be used for phenological model validation and development, evaluation of satellite remote sensing data products, and studies of climate change impacts on terrestrial ecosystems. The PhenoCam Network archive contains:

- 246 Sites (194 active)
- 5.2+ Million images
- 800+ site years of (raw) data
- Sites in 9 different ecosystems/biomes
- All available for download to registered users!

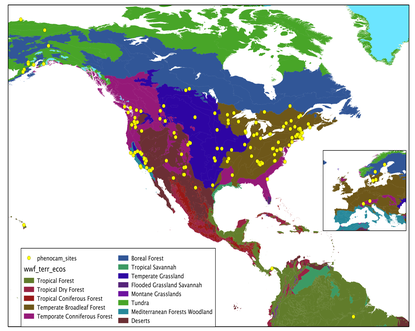

Figure 1: Map of PhenoCam sites with WWF Biomes

The Problem

- Each user downloads and stores the imagery.
- Processing the images while straightforward is time-consuming and must be repeated by each user.
- Any issues encountered when processing the images (shifts in field of view, data gaps, color balance changes, etc.) must be discovered and accounted for by each user.
- Comparing results from different processing procedures becomes complicated.

The solution

- Apply "standard" processing and provide resulting "standard" products (in addition to image downloads).

Standard Image Processing

The standard processing will:

- Provide mask(s) designating a region-of-interest (ROI) for each site. Multiple ROI's per site may be provided depending on vegetation types in the image.
- Account for field-of-view (FOV) shifts by using lists of these masks with associated valid date/time ranges where each mask covers roughly the same ROI.
- Provide image statistics calculated for each image/mask pair. (See Ref. 1 for details.)
- Pull in ancillary information like vegetation type, solar zenith angle, exposure, snow cover, etc.

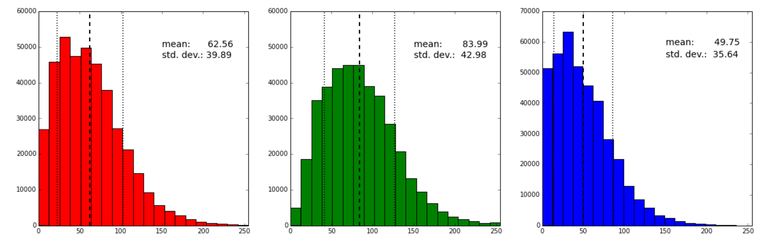

Figure 3: Distributions for the red, green and blue pixel DN values for the ROI.

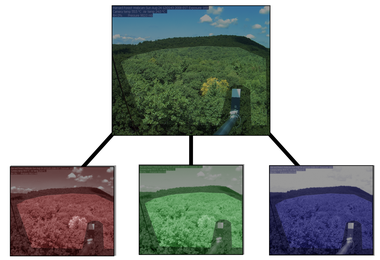

Figure 2 – Image and mask pair applied to R, G, and B color planes.

In addition to the mean and std. dev. we calculate a 'gcc' value for each image/mask pair defined as:

$$gcc = \frac{g_{mean}}{r_{mean} + g_{mean} + b_{mean}}$$

Here the means are taken over the pixels in the ROI. To better characterize the distribution of digital number values (DN) over the ROI we also collect the 5,10,25,50,75,90,95th percentiles and correlation coefficients between color DNs in the ROI.

[1] http://nbviewer.ipython.org/github/tmilliman/phenocam_notebooks/blob/master/Standard_Processing_ROI_Stats/PhenoCam_ROI_stats.ipynb

Standard Product Overview

The "standard" products, updated daily, will be provided for each ROI/site pairs:

- **An ROI timeseries (or "all image") CSV**. Each line summarizes the ROI statistics for a single image and mask. Only minimal image filtering will be done so there will be one record for most images in the archive. (http://phenocam.unh.edu/webcam/tools/roi_timeseries_format/)
- **1-day and 3-day GCC summary CSV's**. The ROI time series will be filtered (solar elevation angle > 5°, brightness thresholds applied, etc.) and then each record will be 1-day or 3-day summary statistics for the time period. (http://phenocam.unh.edu/webcam/tools/gcc_file_format/)

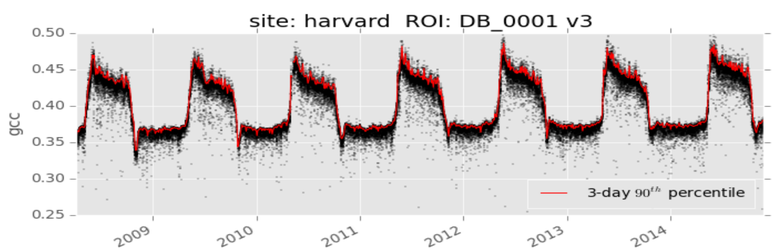

Figure 4: This figure was produced using only the standard products. Black dots are gcc values for each image in the ROI timeseries and the red line is from the 3-day GCC summary file.

Future Work

- **Software Tools for Standard Products**
  By using a relatively simple CSV format, visualization and analysis can be easily accomplished using a wide variety of existing software tools. Software developed by the PhenoCam community for these products can be easily shared. We will be working on providing software for the standard products such as:
  - Plotting utilities
  - Transition date extraction and growing-season length extraction
  - Outlier detection
  - Improved filtering for snow, atmospheric conditions (fog, rain, clouds), FOV obstructions, etc.
- **Improved Metadata**
  These products will become more useful if we can provide additional information like local site information (quality, species, biome/ecosystem type, local contacts, publications, etc.)
- **Citizen Science**
  An associated project is investigating the possibility of using citizen scientists to help identify other features in the images which might be useful auxillary data such as snow cover, flowering events, budburst, etc.
- **Links to Related Data**
  - Climate data (e.g. Daymet[2])
  - Other sources of vegetation remote sensing (e.g. MODIS, Landsat, etc.)

[2] Scripts for extracting Daymet data from the ORNL DAAC for PhenoCam sites have been posted here: http://daymet.ornl.gov/web_services.html

The PhenoCam project is generously supported by the National Science Foundation, through the Macrosystems Biology Program, award EF-1065029. Additional support is provided by the USGS, through grant number G10AP00129, and the Northeastern States Research Cooperative.