

How to access PhenoCam data and imagery: A tutorial

Andrew Richardson Northern Arizona University 13 November 2023

How to access PhenoCam data and imagery: A tutorial

Andrew Richardson Northern Arizona University V2 16 August 2021 V3 13 November 2023 (with feedback from Natasha Wesely)

To the reader

This is a work in progress. If something needs updating, clarification, etc.—please let us know.

Objective

This tutorial is designed to help you learn how to access, browse, and download data and imagery collected in ecosystems around the world and made publicly available by the PhenoCam Network (https://phenocam.nau.edu/)

We welcome your comments and feedback. Please email Professor Andrew Richardson directly, Andrew.Richardson@nau.edu

What is PhenoCam?

Briefly, with the PhenoCam Network we are using digital camera imagery to track vegetation seasonality — its phenology — in a wide range of ecosystems across North America and around the world. Phenology is highly sensitive to year-to-year variation in weather, and long-term variation in climate, and has been found to be a robust indicator of the biological impacts of climate change.

Most participating network sites use a standard camera, the StarDot NetCam. This camera measures about 3" x 3" x 6" and it fits in a standard camera enclosure that is smaller than a shoebox.

Cameras record images from sunrise to sunset, every half hour, 365 days per year. The images are uploaded to the PhenoCam Network server for archiving, processing, and distribution.

We could look at the images and visually assess when leaves emerge or when they change color in autumn. But we can also treat the camera images the same way we would a multi-channel satellite image. In fact, digital cameras record images in three different "layers" or "channels" — the imaging sensor has a red channel, a green channel, and a blue channel. The red-greenblue additive color model says that we can reproduce any color $-$ as perceived by the human eye – through some combination of the three primary colors, and you can see that here: yellow is a mix of green light and red light, for example.

Remember that the primary colors for *light* (red, green and blue) are different from those for *pigments* (technically magenta, yellow and cyan but more commonly red, yellow and blue —the *subtractive* color model for pigments).

If this sounds confusing, keep in mind that this is exactly how your old-style TV set works: you may remember that if you looked up close at the screen, you would see that what appeared white from a distance was actually separate red, green and blue dots. If the blue dots were turned off, but the red and green dots were still on, you'd have yellow. And we can get a wider range of colors not just by turning each primary color on or off, but by adjusting its brightness or intensity.

You can easily play around with mixing red, green, and blue to produce a range of different colors. One way to do this is in Microsoft Word – if you click on the Font Color button on the toolbar, and then select "More Colors…", you can then select "RGB Sliders" in the pop-up window.

By adjusting each RGB slider you can produce almost any color you can imagine. Here, we've made a green the color of late-spring maple leaves by upping the green slider and holding back on the red and blue sliders.

There are also web pages that allow you to mix colors onscreen; try: https://csfieldguide.org.nz/en/interactives/rgb-mixer/ What the RGB color model means is that we can split our digital camera images into these three layers—red green and blue— and for a particular region within each image, quantify the intensity of each color, and develop a *quantitative* measure of canopy greenness. Our standard index, the green chromatic coordinate (GCC) is just the relative brightness of the green channel, normalized against the overall brightness of the red green and blue channels together.

$$
GCC = \frac{G \text{ intensity}}{R \text{ intensity} + G \text{ intensity} + B \text{ intensity}}
$$

Here's a plot of 6 years of GCC data for a deciduous forest. Greenness is low in the winter, rises steeply in spring, drops off gradually over the summer, and falls sharply in autumn. And these cycles repeat year after year, with some variation in the timing of green-up and green-down, depending on how the weather in that year unfolds, with warmer weather generally advancing spring green—up and delaying green-down. The timing of these seasonal changes can vary by 2 weeks or more from year to year in response to that weather.

From the GCC time series, we also estimate "green-up" and "green-down" transition dates, corresponding to the start and end of the growing season (in a deciduous forest, for example, these line up well with budburst or leaf out in spring, and leaf coloration and leaf drop in autumn).

To determine transition dates, we first estimate the minimum and maximum values of greenness over the course of the year, and from that we calculate the range.

We then determine the dates when our greenness curve hits 10%, 25%, and 50% of that range, both in spring and autumn – these percentages define our spring and autumn transition dates The 10% date corresponds about to when leaves are emerging in spring, and when they reach their maximum color in fall.

Navigating the PhenoCam Network Web Page

Now that you know what PhenoCam is doing and how we use the camera imagery, let's explore the PhenoCam Network web page.

The URL for the PhenoCam Network web page is:

https://phenocam.nau.edu/webcam/

Please note that this URL was updated during the summer of 2022; the old UNH URL no longer works. While the screen grabs below may still reference UNH, you should always use the NAU address.

When you navigate your browser to that address, you will see the landing page shown below.

The "About" tab has some basic background information about the project team, why we are interested in phenology, and the science questions we are addressing with PhenoCam data.

If you click on the "Gallery" tab, you will be presented with thumbnail pictures for each site in the PhenoCam network. Sites are arranged alphabetically and grouped by "site type". Type I sites follow our standard protocol and we are directly in contact with the researchers at the site (typically the highest quality data). Type II cameras don't follow the standard protocol, but we are still in contact with researchers at the site. Type III cameras don't follow the standard protocol and we do not have contact with site personnel (typically lower quality data, with longer and more frequent data gaps).

You can scroll through the gallery to find a site of interest, based on what you see in the thumbnail, or if you know a site name, you can search for it directly (HF for "find in page" in Firefox or Safari on a Mac; or control-F in Windows) — in the example below, I have searched for all the cameras whose name starts with "harvard," for example:

Clicking on a thumbnail takes you to the "site page" for that camera. Here's the site page for the main Harvard Forest camera (site *harvard*), mounted on the EMS Tower and in operation since 2008. At the top of the page, the camera's site name is just a short name, typically all lowercase, that is used to identify the images from that specific camera. The "location" field provides more descriptive information about where the camera is actually located. The latitude, longitude and elevation are for the base of the structure the camera is mounted on, NOT the center of the camera field of view. The "Site Metadata" button brings up information about the direction the camera is pointing (most commonly N, for North), the mean annual temperature (MAT) and precipitation (MAP) $($ site is as recorded at the site, while daymet and worldclim are from the gridded DayMet and WorldClim data products, based on statistical interpolation of weather records from surrounding stations), the dominant species (as Latin names; you can easily find out the common name for a species by googling the Latin name), and the primary and secondary vegetation types at the site ($DB =$ deciduous broadleaf forest, $DN =$ deciduous needleaf forest, EB = evergreen broadleaf forest, EN = evergreen needleleaf forest, GR = grassland, AG = agriculture, SH = shrubland, TN = tundra, WL = wetland).

The large image on the site page is by default the latest image recorded by the camera (updated about every 20 minutes), while using the thumbnails across the bottom you can choose instead the mid-day image from any one of the last five days.

The map on the right, by default, shows a base map of Google satellite imagery. If you click the layers tab in the upper right, you will get a drop-down menu to select the background layer.

ared by Esri | DigitalGlobe, GeoEye, i-cubed, USD

The streets option will show placenames and roads, the topographic option will show shading and contours, and the imagery option will show a satellite image (default). Two land cover overlays, at different spatial resolutions, can be selected using the MODIS and NLCD options these show how the area around the site is classified (e.g. forest, grassland, urban, water) based on satellite imagery. This is useful for evaluating whether a given site is representative of the broader landscape. For example, with the MODIS overlay, you can see that Harvard Forest is mostly surrounded by deciduous broadleaf forest, with some mixed (deciduous/evergreen) forest as well.

Selecting the "Browse Images" button will take you to a gallery like this, where the first mid-day image of each month is shown, grouped by year.

You can click a single year and then select a month to see a calendar like this, from which you can then select a given day.

When you click on a particular day, here May 15 2021, you will see all the images recorded throughout that day. Most cameras record images from 4 am to 8 pm, local standard time. An image is also recorded each night when the camera reboots itself. Hitting the (i) icon next to the time stamp for an image will bring up a text file with the metadata recorded along with the image (but for most users, this metadata is not of interest).

If you click on any image thumbnail, you will get a full-size version of the image in your browser. You can drag the image to your desktop, or copy it ($\operatorname{\mathsf{HC}}$ or Copy from the Edit menu on a Mac; right click then select "copy" in Windows) to paste into another document (e.g. PowerPoint), or save it to another location ((HC or "Save as" from the File menu of Firefox or Safari).

Back on the main page for the *harvard* site, there are also two buttons under the Google map labeled "ROI Timeseries". ROI stands for "region of interest" (a selected area within an image that includes a particular vegetation type), and these buttons will take you to graphs derived from color information for that selected region within the camera's field of view. The first two letters for an ROI name correspond to the vegetation type, e.g. DB for deciduous broadleaf forest. The DB 1000 ROI for Harvard Forest is focused on the deciduous trees in the image foreground. You can see the mask (representing the ROI) that is used to identify this vegetation, which is shown at the bottom of the page.

The time series of data calculated for this mask is displayed on the graph at the top. Remember that GCC is the green chromatic coordinate—a measure of how bright the "green" signal is across the mask, relative to the total brightness of the all the colors that comprise the digital image (red, green, and blue-). We have found this GCC index to be a reliable metric that we can use to characterize phenology. When GCC is high, the foliage is very green. A low value of GCC can mean several things, most typically brown, senesced (dead) foliage (as in grasslands in the

dormant season) or a barren canopy (as in a deciduous forest in winter). But, maple forests in New England and the Great Lakes Region also have very low GCC when autumn colors are at their peak.

With the sliders across the bottom, you can adjust the date range that is depicted on the graph. Here, we're looking at just January 2013-March 2014. If you mouse over a point on the graph (this can be difficult to do accurately when multiple years are shown), the date and GCC value for that date will be shown on the graph. If you click on a point on the graph, you will be taken to the browse image page for that day, exactly the same as above.

The "Provisional Data" link will also allow you to download a zip file that includes (1) color information for the region of interest, calculated for every image in the archive; (2) one- and 3 day data products (we typically record images every half hour, but phenological changes don't occur that quickly so we aggregate the data to produce data products that are more easily used; the 1-day data are used to generate the time series graph shown above, for example); (3) "green-up" and "green-down" transition dates (roughly representing the start of spring and end of autumn, calculated using the methods described earlier) derived from the GCC time series. The format of these data files is described in the 2018 paper in *Scientific Data* by Richardson et al. (open access at https://www.nature.com/articles/sdata201828).

The standard images recorded at each site are visible-wavelength pictures based on the red, green and blue additive color (RGB) model used to create and store digital images. Each of these primary colors corresponds to a layer in the image, and together these three layers can be used to reproduce a scene as perceived by the human eye. But, many of the cameras (those identified with IR in the image gallery) also record a visible plus near infrared (NIR) image. NIR wavelengths are beyond what can be seen by the human eye, and healthy vegetation is highly reflective in the NIR. The *harvardems2* camera, mounted on the same tower as the *harvard* camera used in the above examples, provides a nice example. This is the site page for the *harvardems2* camera, and at the top you can see the "show IR view" link. In visible wavelengths, the canopy is quite dark, and green in color.

In the visible + NIR image, the deciduous canopy is almost luminous because it reflects more than 50% of the incoming near infrared light (about half of the sun's radiation is in the visible range, from 400-700 nm, while half is in the NIR, from 700 to 2500 nm). This compares with about 10% of the visible-wavelength light that plant canopies reflect.

You can browse through the visible+NIR image archive just as you can browse the visiblewavelength image archive.

In the examples above, we have looked at images and data for a typical deciduous broadleaf forest. But, the PhenoCam Network archive includes imagery from sites in many different climate zones, with different vegetation, and different seasonality.

You may be surprised that even evergreen conifer forests have a season signal in the PhenoCam data. For example, here's the GCC time series for the niwot3 site, in Colorado's Rocky Mountains. The big dips that occur during the winter months are when recent snow fall is covering the trees. Snow tends to show up as whitish-blue in the images, and the GCC of snow is much lower than the GCC of green needles.

You can also see that even when there isn't snow on the trees, greenness is low in January and might higher in June. That's because the amount and type of pigments in the evergreen needles is actually changing from season to season.

You can see that fairly clearly in the example below. The needles in July is bright green, compared to more of a reddish-green in January. So although these trees retain their needles year round, they are still changing as the seasons pass — and "evergreen" may be a misleading word!

Grasslands come in all different flavors. Here's the GCC time series for a cold grassland site *lethbridge*— in Alberta, Canada, where winter temperatures limit plant growth during the cold months. In most years, there is a well-defined peak in greenness which occurs when temperatures are warm and water is sufficient. In 2016 though, you can see evidence of a summer drought (dip in greenness followed by recovery). In 2019 and 2021 you can see additional evidence of water limitation because the grasses just didn't green up as much as in other years.

But tropical grasslands can look very different. This is from the *kamuela* camera on the Big Island of Hawaii. Greenness goes up and down, but not with the same rhythmic seasonality as at *lethbridge*. In Hawaii the green-up and green-down cycles can be very brief, or they can be prolonged. They are determined by water availability, not by temperature. Rain occurs frequently at this site, but rainfall can be erratic. The vegetation responds quickly to water inputs, but dries up quickly as well—and so regular rainfall is required to keep things looking green.

We don't yet have too many tropical sites in the PhenoCam Network, but since 2014 we've had the *elverde* camera at the Luquillo Experimental Forest Long-Term Ecological Research site in Puerto Rico. This is a very wet forest which receives almost 3000 mm of rain per year. The trees here are mostly evergreen broadleaf species which keep some leaves year-round. But there is still a strong seasonal cycle in greenness, which increases rapidly in early May, and then declines gradually over the following months.

The photos below show before (top) and after (bottom) the May increase in GCC; the new foliage is a bright green, particularly on the tree in the right foreground of the image.

This site is also interesting because the imagery shows the impacts of the 2017 hurricanes, Irma and Maria. The top picture shows the site in late August, before the hurricanes, and the bottom picture in early November. The tremendous reduction in leaf area that occurred as hurricane winds ripped leaves from the trees is apparent.

Before we get back to some of the other features available through the PhenoCam Network web page, let's look a one more type of vegetation—agricultural crops. This is GCC from the *bouldinalfalfa* camera in California. The rapid increases and decreases in greenness are the result of crop growth and harvesting, which occurs on about a monthly schedule.

The top picture shows the field two days apart: before harvest (top) and after harvest (bottom).

With careful browsing of the image archive, you may be lucky enough to find an image that includes the harvester

(https://phenocam.sr.unh.edu/data/archive/bouldinalfalfa/2018/04/bouldinalfalfa_2018_04_23_124611.jpg)

The tabs at the top of the screen – "About", "Gallery", "Map", etc.—can be accessed from any page including the main landing page. The "Map" tab brings up a world map as shown below. You can use the + and – buttons to zoom in and out, and clicking on the map will display a hand icon which allows you to drag what is shown on your screen. When the map is zoomed out, the circles indicate the number of sites within a geographic region.

As you zoom in with the + button, individual sites start to appear, and you can click on a site symbol to bring up a thumbnail and basic metadata for that site. Clicking on the thumbnail will take you to the site page for that site.

The "Site Filters" button at the top of the map can also be used to search for sites by name, vegetation type, or species. For example, here's a map showing sites that have *quercus* (=oak) listed as one of the major species.

As described above, you can download images one at a time using the browse feature, and you can download processed data (which you can then import into Microsoft Excel or Google Sheets, for graphing or other analysis) using the "provisional data" links on each site's ROI page.

If you want to download multiple images at one time, you can use the "Image Data" link under the "Data" tab. Note that to access this tab you have to have a PhenoCam account (we do this so we can track downloads and report to our funding agencies how many different users are downloading data each year). An account is free and easy to set up, using the "Sign Up" button:

Once you're logged in, you can use the drop-down menu to select a site, and then enter start and end dates and start and end times for the period you want included in your download. Here's an example that will give you mid-day images (typically one per day) for an entire year (in this case, 349 images, and 120 MB). The images are compressed into a .zip archive, which is easily unzipped to a normal folder format (images are arranged by year/month) in the Mac OS.

There are a variety of other resources presented in the "More Info" tab. This includes some previously developed educational materials, under the "Education Resources" tab, which were developed in collaboration with Harvard Forest and the National Ecological Observatory Network (NEON).

We ask that everyone using PhenoCam data and imagery familiarize themselves with our Fair Use Data policy, which is found under "Fair Use Policy" in the drop-down menu of the "Data" tab.

Appendix 1: Archive Statistics

Graphs displaying the number of sites in the archive, as well as total image counts, are available under the "Archive Summary" tab under the "More Info" dropdown menu. These are updated nightly.

Appendix 2: The PhenoCam Explorer

Curated data sets are periodically released and are made available through the data archive of the Oak Ridge National Laboratory as well as through the PhenoCam Explorer web page (https://phenocam.nau.edu/phenocam_explorer/#shiny-tab-explorer_v2).

The Explorer offers some functionality that is not available through the standard PhenoCam Network page (https://phenocam.nau.edu/).

This is the landing page for the Explorer. From the dropdown "vegetation type" menu you can limit the analysis and presentation to specific vegetation types. The inset plot shows the distribution of sites in climate space, with mean annual temperature on the X axis and mean annual precipitation on the Y axis.

For example, out of a total of 1773 site-years in the entire V2 dataset, released in 2019, there are 643 site-years of data for deciduous broadleaf forests, across 112 sites.

Clicking on the "Plot and Download Data" tab brings up a navigation pane with data plotting options on the left, site names and associated information on the right, and a plotting window at the bottom. Entering "harvard" in the search box and hitting return gives a list of the site names that include the string "harvard". These are sorted alphabetically by site name, but, for example, the search order can be changed to be the length of the time series in days by clicking on the small arrows to the right of "days". The longest such time series is for the *harvard* camera, which was used above as an example, with 3934 days of data in the version 2 dataset. If you click on the site name, the data will be displayed in the graphing window.

The data are shown at a 3-day time step, which results in a smoother time series than 1-day data but this option can be changed by the drop-down "Frequency" menu. The black dots are the 3-day GCC values, while the green symbols of varying hue denote "green-up" dates in spring and brown symbols of varying hue denote "green-down" dates in autumn. These dates are calculated by first determining the seasonal amplitude in the GCC curve, and then identifying the dates when 10%, 25% and 50% of the seasonal amplitude are reached.

Clicking on the graph itself will display a tool bar in the upper right corner. From left to right, these tools enable you to

• download an image file containing this plot

• zoom in to a specified region of the plot (click magnifying glass, then use your mouse to select the region to zoom in on)

- pan left, right, up or down on a zoomed in region of the plot
- select a region to display data for using a box or lasso
- zoom in (+) or out (–) from the current selection
- reset the zoom to the full range of the plot
- reset the axes to default
- display reference lines that extend from the current mouse position to the x and y axes
- show the value associated with the nearest data point when the mouse hovers over the plot

By clicking on any of the plot symbols on the right, you can also turn plotting on or off for any of the data sets. Here we have turned off all but the smoothed curve fit to the data ("Gcc loess fit") and the 10% green-up and green-down dates.

This comparison shows the difference between 1-day and 3-day data, for 2013. You can see that there is more variability in the 1-day data, but for some ecosystem types where phenology is very responsive to environmental factors, this may be preferred over the smoother time series in produced by the 3-day aggregation.

The drop-down "Percentile" option allows you to select different ways that data from the halfhourly images are aggregated to a 1- or 3-day data product. The 90th percentile method basically tracks the upper envelope of the data, and is quite effective at reducing artifacts associated with rain and snow. But, a median (50th percentile) or mean (average) may sometimes give better results.

Under the "Plot Type" drop down menu you can choose alternative ways of displaying the data. For example, "Time Series (DOY)" plots all years of data on the same X-axis where the day of year, from 1 to 365, is on the X axis.

For sites like Harvard with many years of data, this is a nice way to easily visualize outliers and "unusual' years. Years are color coded and mousing over a particular curve will allow you to identify what year that is. Here the brown curve with early green-up is 2010. The dotted black line is the long term mean of all the data (LTM) and the shading around that curve corresponds to ±1 standard deviation (a measure of variability).

The "Time Series (RCC/BCC)" option will display the blue and red chromatic coordinates (BCC and RCC, respectively), which are analogous to the GCC that is more commonly used. Here we have turned off the BCC and GCC data, as well as the GCC loess smoother, and are showing only the RCC data. We've also zoomed in to January 2012 to January 2015.

There are a few things to point out about this plot. First, there is a small peak in canopy RCC just as leaves emerge in spring. Many juvenile leaves tend to be tinged with red, and this shows up in the camera data. As the leaves mature, their pigments change and the green of chlorophyll becomes dominant. Also, the peak in autumn color is very clearly shown in the data. The maximum RCC value in 2013 occurred on October 20 (mouse over the data points). The image for that day (located by browsing the archive) is included below.

Here's another example, from the *bbc7* camera located in the White Mountains of New Hampshire. In 2018, autumn colors peaked on October 8. Note that the RCC value is 0.49, vs. 0.40 for the *harvard* example. This indicates the canopy is much redder at the New Hampshire site, and as the attached image shows, the colors are quite spectacular.

For sites with long enough time series, the "Phenology Trends" option can be used to plot the long-term trend in spring green-up and autumn green-down transition dates. The goodness-offit (R2) of the trendline, and its slope, are indicated. Between 2008 and 2018, green-up was trending later at a rate of 0.6 d per year, while green-down was trending later at a rate of 1.26 d per year. This suggests an overall extension of growing season length over this period, driven by the delayed senescence (leaf-fall).

You can use the magnifying glass to zoom in on just the springtime data. 2010 was the earliest year for green-up during this 10-y period, while 2016 was the latest. Overall there was a range of 2 ½ weeks between earliest and latest green-up!

Finally, with the basic "Time Series" option selected, you can also elect to have satellite data superimposed over the PhenoCam data. EVI (Enhanced Vegetation Index) and NDVI (Normalized Difference Vegetation Index) are two commonly used indices calculated from satellite data that track vegetation greenness. Here we are using data from the MODIS satellite platform (https://modis.gsfc.nasa.gov/about/). The satellite data are at a comparatively coarse spatial (500 m) and temporal (16 day) resolution, but you can see that EVI (green line) lines up well with the phenological signals in the PhenoCam data, at least for the *harvard* site.

At the bottom of the window is a "Download Data" button. The organization and file structure for these data from Explorer are the same as the Provisional Data download from the ROI pages described earlier. The main difference is that the Explorer data have been curated and publicly released. The provisional data are more up-to-date (updated through the previous day's imagery) but may not have undergone rigorous quality control or expert evaluation.

Appendix 3: The PhenoCam API (Application Programming Interface)

The PhenoCam API is a work-in-progress tool for advanced users that facilitates access to PhenoCam data and imagery. It can be accessed through the following URL:

https://phenocam.nau.edu/api/

There is minimal documentation available here:

https://phenocam.nau.edu/api/docs/

For R users there is an API wrapper available at: https://github.com/phenocam/phenocamapi

There is an introduction of how to use the API wrapper at: https://www.neonscience.org/resources/learning-hub/tutorials/phenocam-api-intro

The online textbook for the NAU graduate class, INF 550: Environmental Informatics Using Research Infrastructures and their Data, has an R Markdown tutorial that uses the PhenoCam API:

https://github.com/katharynduffy/Environmental-Informatics-Using-Research-Infrastructuresand-their-Data/blob/master/05-PhenoCam.Rmd