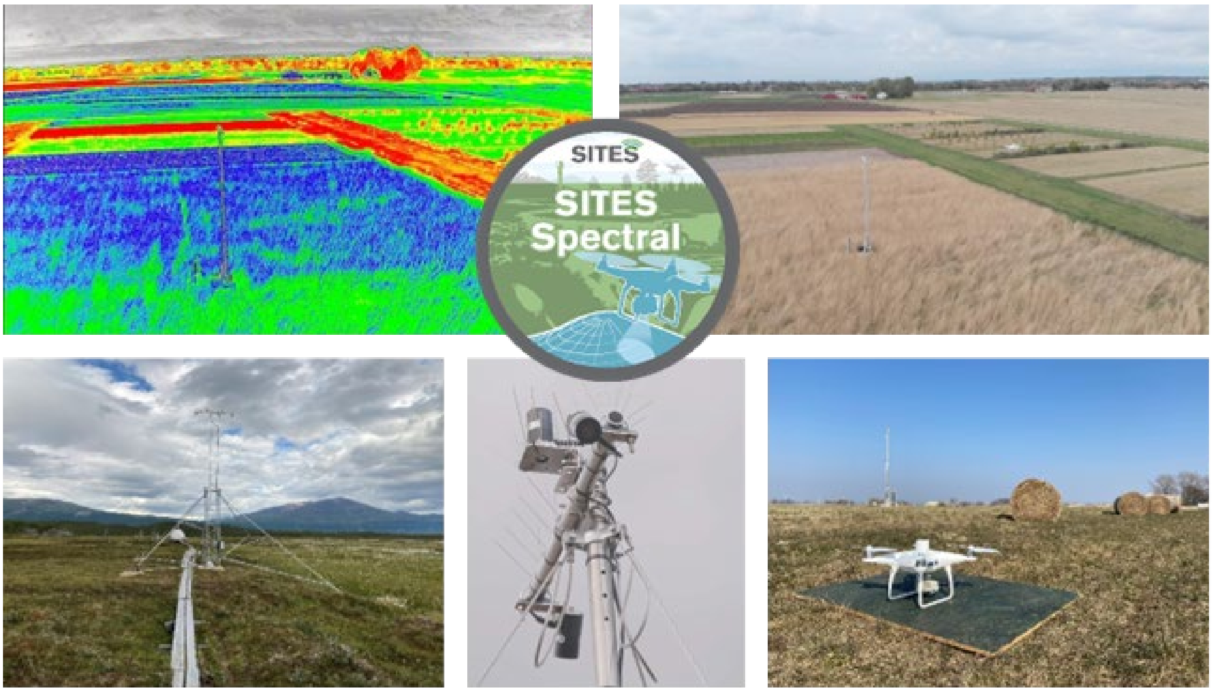
SITES Spectral

Algorithm Theoretical Based Document (ATBD)

Ver. 2023-03-31

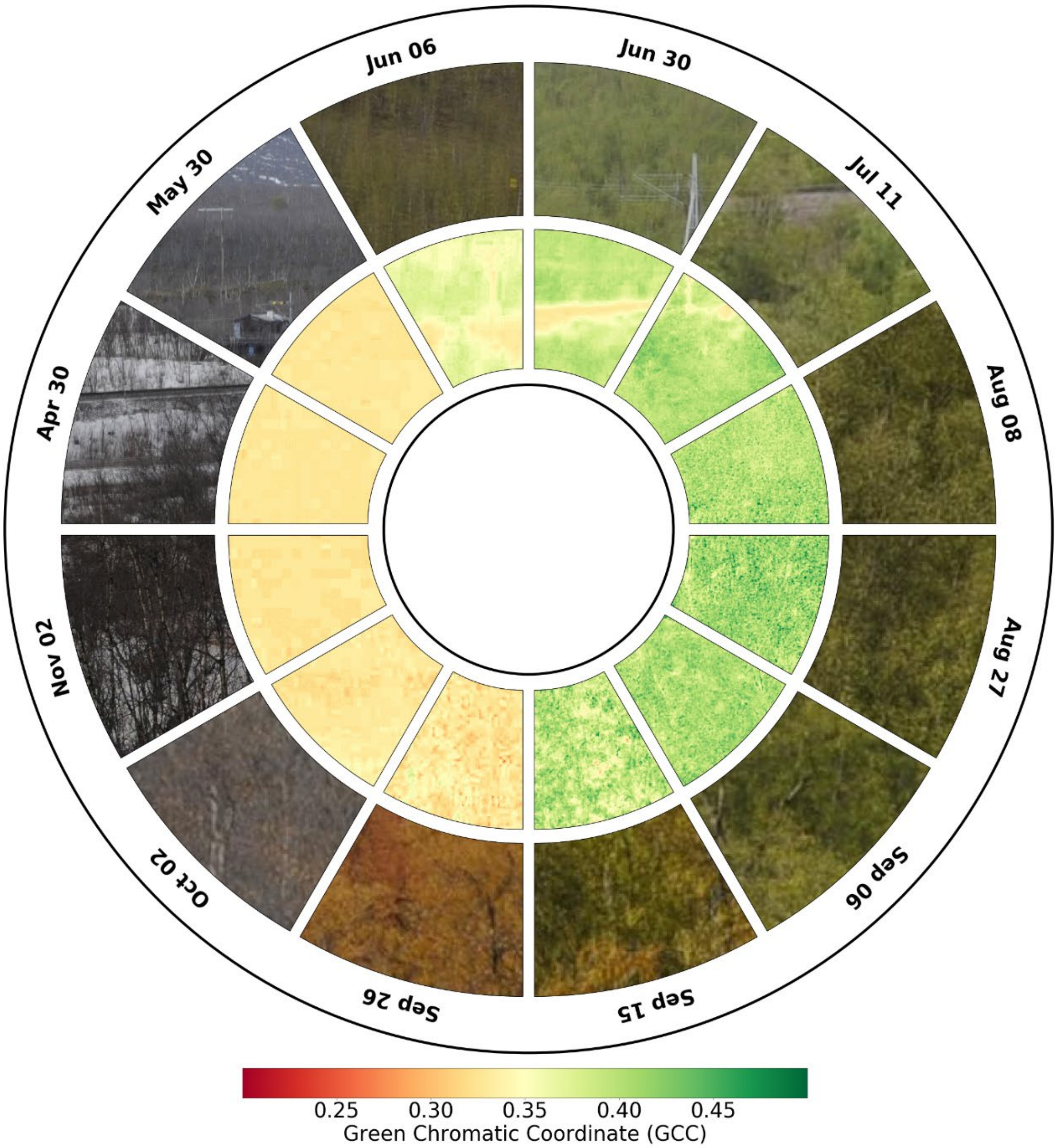


Shangharsha Thapa
Per-Ola Olsson
José Beltrán
Lars Eklundh

Department of Physical Geography and Ecosystem Science
Lund University

Table of Contents

Chapter 1. PhenoCam



1.1 SITES PhenoCam Network

The concept of digital repeat photography has been increasingly used for near surface remote sensing of vegetation (Keenan et al. 2014; Migliavacca et al. 2011) mainly because of its low cost, high temporal resolution, high-quality imagery, and automatic operation. In recent years, there have been significant progress in the imaging technolgy such as the commercially available digital cameras, often called PhenoCam, designed specifically for tracking the timing of different life cycle stages of plants such as when plants grow buds, leaf out, flower, fruit and die back, the science known as plant phenology. The study of phenology dynamics is highly sensitive to the interannual variation in weather, and also to the long term variation in climate, thereby helpful in understanding how the vegetations are responding to the climate change (Richardson et al. 2018a).

SITES Spectral has 15 such cameras set up at 8 SITES stations (*Figure 1*), located at different locations across Sweden recording images, every half hour, and all year long of different ecosystems. Most participating stations use a standard camera, the Mobotix camera while few stations have the StarDot NetCam. The images collected by these cameras are uploaded to the Lund University server for archiving, processing, and dissemination of different data products to the SITES data portal.

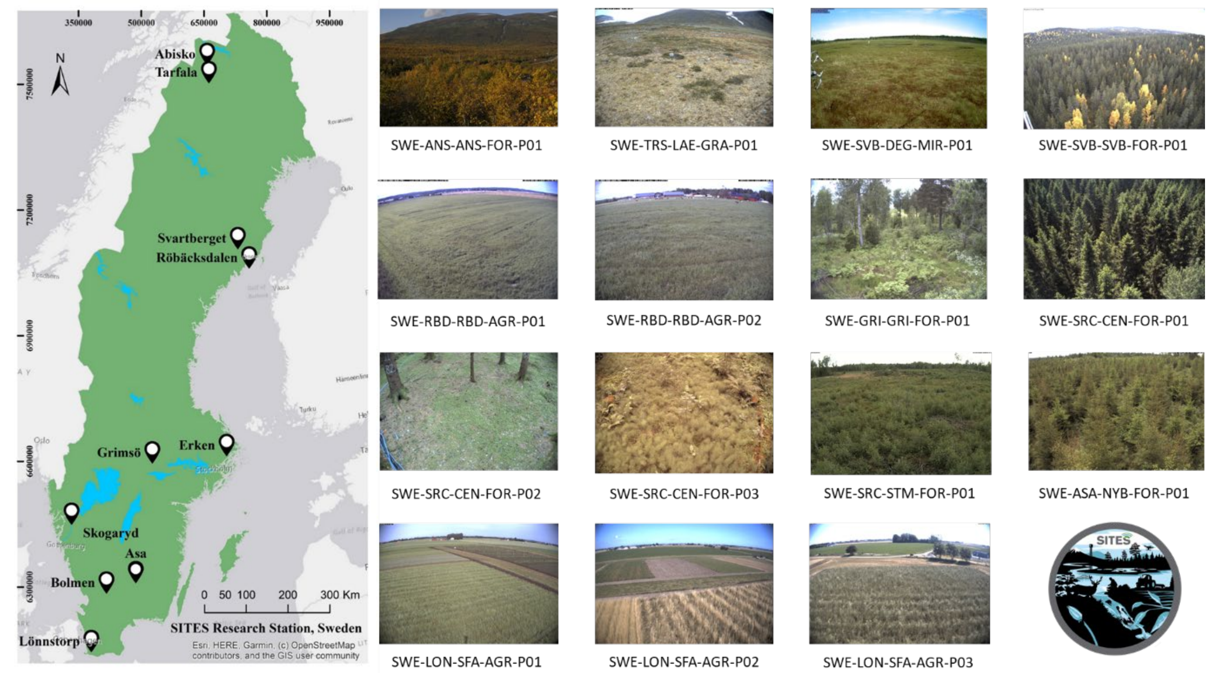


Figure 1. SITES research stations location within Sweden and the sample phenoCam camera view for stations participating in SITES Spectral thematic program.

1.2 PhenoCam Data

PhenoCams record a coloured three layer green (G), red (R), and blue (B) images of the vegetations in an ecosystem the camera is pointed to throughout the season and the seasonal trends of this color information (digital numbers; DN). For example (*Figure 2*), as plants turn green in the spring, the amount of brightness of green in the image increases, and when the plant turns red in the fall, the camera will record that too.

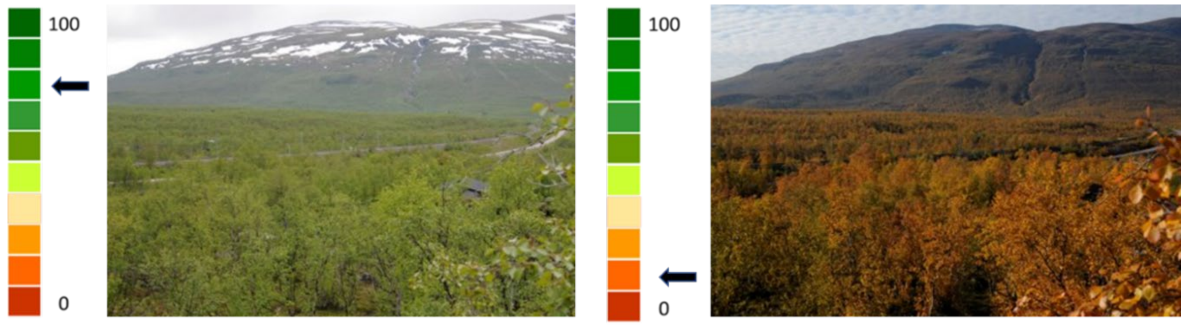


Figure 2. PhenoCam pictures from SITES Abisko Scientific Research Station showing the birch forest over the spring and autumn of 2021. The vertical bar represents how the intensity of green color changes over the seasons.

The images acquired can be splitted into each individual bands R, G, and B, and for a particular region of interest (ROI), quantify the intensity of each color. The relative image brightness for the green channel, normalized against the overall image brightness (i.e., a total of RGB channels together) is called the Green Chromatic Coordinate (GCC) while the same with the red channel is defined as the Red Chromatic Coordinate (RCC) (Gillespie et al. 1987)

$$GCC = \frac{DN_G}{(DN_R+DN_G+DN_B)} \tag{1}$$

$$RCC = \frac{DN_R}{(DN_R+DN_G+DN_B)} \tag{2}$$

where DN_R= mean red DN, DN_G= mean green DN and DN_B= mean blue DN

These two indices (*equation 1 and 2*) have been extensively used to extract information about vegetation phenology (Richardson et al. 2018a; Seyednasrollah et al. 2019; Liu et al. 2020; Thapa et al. 2021).

Plotting the time series of GCC or RCC over the years can indicate when the trees begin to turn green, when they are at their greenest, when the leaves begin to turn red, when they are the reddest, and when the leaves start to fall or the trees die back for the winter. The continuous recording of such data over time makes it possible to compare the dates of plants turning green in the spring and dying back and losing their leaves in the fall, to see if the dates change significantly from year to year.
SITES offers the archived phenoCam data in different levels and format i.e., from the raw level 0 (L0) data product at original half hourly, hourly, or sub-hourly temporal resolution to the analysis ready quality flagged time series level 2 (L2) data. All these different data levels are made available through the SITES data portal (https://data.fieldsites.se/portal/). Table 1 lists all available data levels in the data portal for the phenoCam sensor.

Table 1. List of PhenoCam data levels and their names

Level	**Product names**
L0	Raw images
L1	Quality filtered images
	Daily averaged GCC image
L2	Daily averaged RCC image
	Daily averaged RGB image
L3	Daily averaged GCC-RCC time series

1.2.1 Data level L0 – Raw images

This product is the original phenocamera images captured at a specific time interval and doesn't undergo any sort of image processing. However, at this stage, it is made sure that each the phenoCam image follows the proper image naming convention as per the standards of SITES.

Most of the stations within SITES are programmed to capture images every half hour while this number varies in case of few stations. In addition, the stations operate phenoCams all year round in most of the cases with few exceptions. *Table 2* shows the phenoCams at each location with temporal resolution as well as the time of the year they are in operation.

*Table 2. PhenoCams location at each SITES station with the image frequency and the year when they started observations. Note * for Asa refers that there is no data from 2020 onwards as they are upgrading the mast while * Tarfala suggests that it is no longer part of SITES since 2023.*

Stations	PhenoCams	Temporal resolution	Start year
Abisko	SWE-ANS-ANS-FOR-P01	2 hours	2010
Asa	SWE-ASA-NYB-FOR-P01	1 hour	2015*
Grimsö	SWE-GRI-GRI-FOR-P01	0.5 hour	2020
Lönnstorp	SWE-LON-SFA-AGR-P01	0.5 hour	2018
	SWE-LON-SFA-AGR-P02		2018
	SWE-LON-SFA-AGR-P03		2018
Skogaryd	SWE-SRC-CEN-FOR-P01	1 hour	2018
	SWE-SRC-CEN-FOR-P02		2018
	SWE-SRC-CEN-FOR-P03		2018
	SWE-SRC-STD-FOR-P01		2013
Svartberget	SWE-SVB-DEG-MIR-P01	1 hour	2016
	SWE-SVB-SVB-FOR-P01	0.5 hour	2019
Röbäcksdalen	SWE-RBD-RBD-AGR-P01	0.5 hour	2019
	SWE-RBD-RBD-AGR-P02		2019
Tarfala	SWE-TRS-LAE-GRA-P01	0.5 hour	2019*

1.2.2 Data level L1 – Quality filtered images

Level 1 phenoCam data are quality filtered images. The images taken early in the images and late in the evening are removed from the L0 data by applying various quality filters. The images acquired only between 10:00 AM – 14:00 PM are extracted by applying the time filter to avoid low solar elevation angle. After this step, an effort is made to remove the poor quality pictures from the collection. Images that appear too bright (solar glare), too dark, blurry, bad weather (foggy), disorientated and images with obstacles (e.g. birds) are considered to be the bad quality and hence removed manually from L1 collections. *Figure 3* shows few examples of such bad quality images that are removed at this stage.

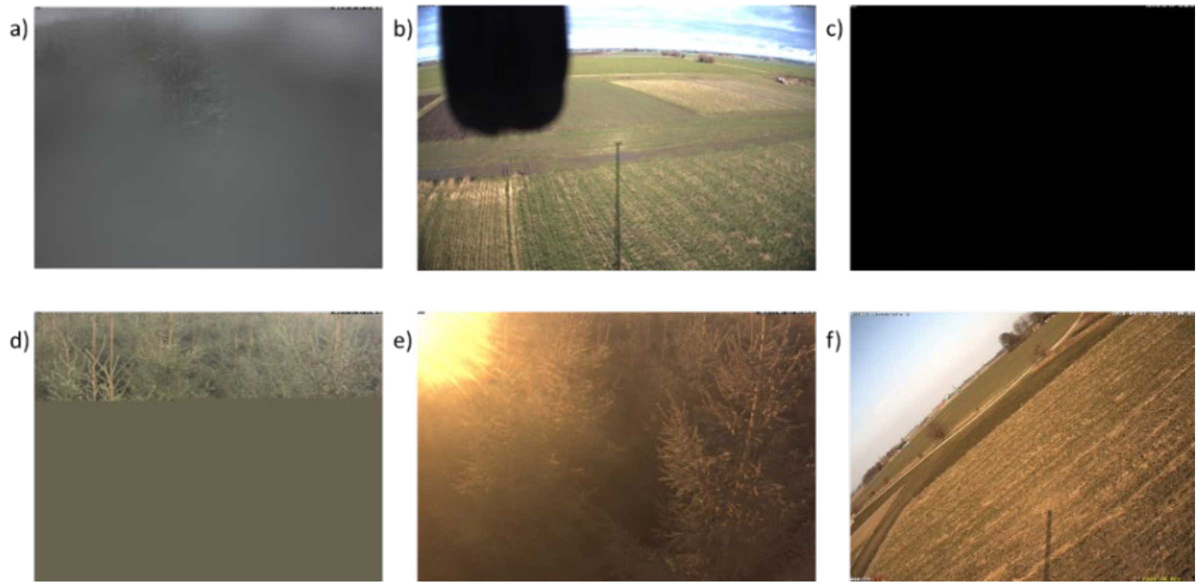


Figure 3. Bad quality pictures: a) Foggy, b) Obstacles, c) Dark, d) Stripes, e) Solar glare, and f) Disoriented images. All such poor quality images are removed at this stage.

1.2.3 Data level L2 – Daily Averaged GCC, RCC and RGB images

The quality filtered L1 images available daily are then processed to generate three different level 2 data products. The three L2 data products are the daily averaged GCC, RCC, and RGB images over whole camera field of view (FOV). First, the images are grouped together for each day of the year (DOY) and then the available images for each day are averaged. The daily averaged RGB images matches with the same naming convention as well as the image properties of the original L0 images. Later, these daily averaged RGB images are used to compute the spectral vegetation indices (VIs) called GCC and RCC (*equation 1 and 2* respectively) over whole camera FOV and hence named as the daily averaged GCC and RCC. The daily averaged GCC and RCC are available as 8 bit grayscale images. All L2 data products are exported in *.jpg* format and are available in SITES data portal as a zipped (*.zip*) annual file containing the daily average alongside metadata information in text (*.txt*) file. *Figure 4* shows an example of all L2 data products.

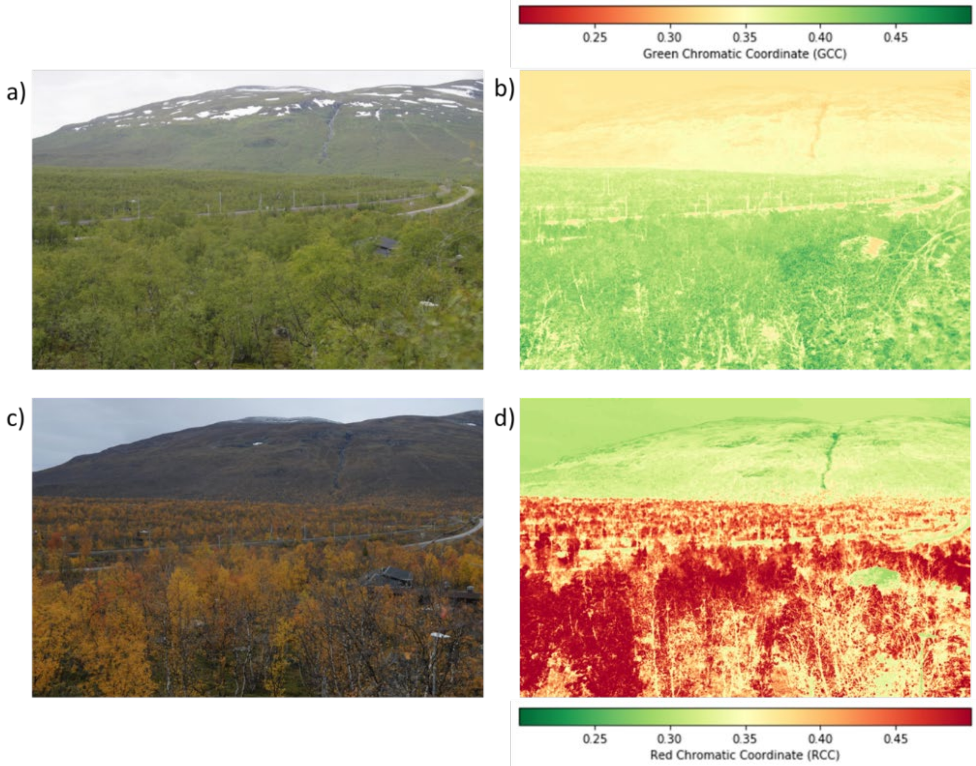


Figure 4. L2 phenoCam data products from SITES Abisko Scientific Research Station. Here, a), b) are the daily averaged RGB images for 181 and 269th DOY in 2021 and (c), (d) are the corresponding daily averaged GCC and RCC respectively for the same data. Note that the daily averaged GCC and RCC presented here are color scaled but these products are available as grayscale images.

1.2.4 Data level L3 – Daily averaged GCC-RCC time series

Level 3 phenoCam data are the time series of daily averaged GCC and RCC computed for a defined polygonal ROI. The L3 time series data are derived using the quality filtered L1 data and are available in the SITES data portal as a comma separated *.csv* file. In addition the L3 data files contain all metadata information as headers. ROIs are defined in a way that best represents the vegetations and area of interests within the camera field of view and the number can vary from one to many. *Figure 5* shows one example of how ROIs are defined and how the numbers can vary across stations.

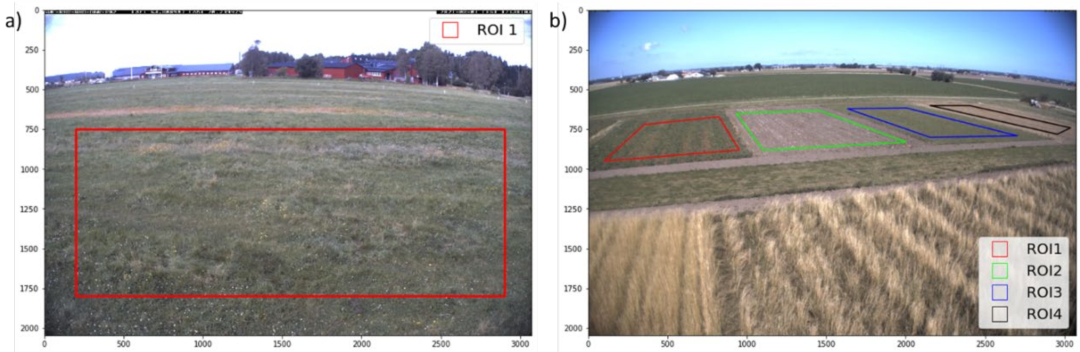


Figure 5. ROI definition to the images (2022) from phenoCams at two agricultural research stations within SITES a) Röbäcksdalen and b) Lönnstorp research station. The number of ROIs vary from station to station based on the vegetations and area of interests within the camera FOV.

The ROI is constant for all the images of each phenoCams over the years. Only the RGB color information within the ROIs are used for extracting the time series of RGB DN triplets. At this stage, it is very important to make sure that the time series data are not affected by the camera movements as this would alter the camera FOV and thus the ROI pointing to a different area and/or vegetations. In case of camera movements detection, for example the one in *Figure 6*, such images are processed separately by using a revised ROI that fits the same area.

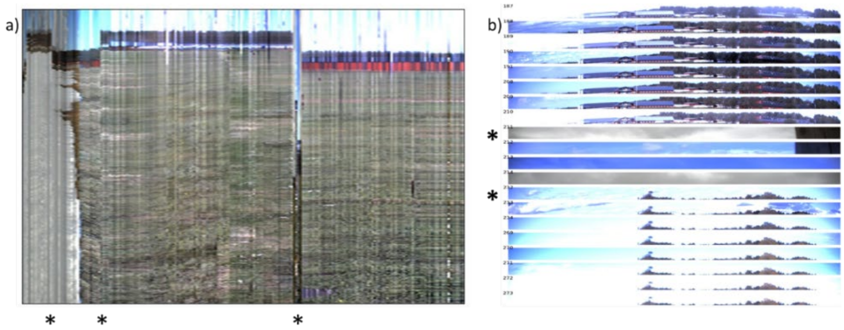


*Figure 6. Camera movements at SITES Röbäcksdalen Field Research station in 2019 showing the a) vertical and b) horizontal shifts in FOV. Vertical shifts are detected when plotting the mid column of annual midday images while horizontal shifts are detected by tracking a certain portion of image covering the horizon. Note that * represents the day when camera movements were detected.*

These average RGB DN triplets for defined ROI from each of the images are extracted using the OpenCV and numpy Python packages followed by the computation of the time series of daily averaged GCC and RCC. Alongside, the time series data of daily averaged GCC and RCC in the L3 data file, other informations are also provided such as timestamp, DOY, daily average of RGB channels separately for each ROIs defined, standard deviation of the computed daily GCC and RCC values, number of images used for computing average, maximum solar elevation for each day, and most importantly the three digit quality code value. All columns in the L3 data file except for the timestamp and DOY, the days when there are no images available, are defined as NaN. *Figure 7* shows one instance of L3 daily averaged GCC and RCC time series.

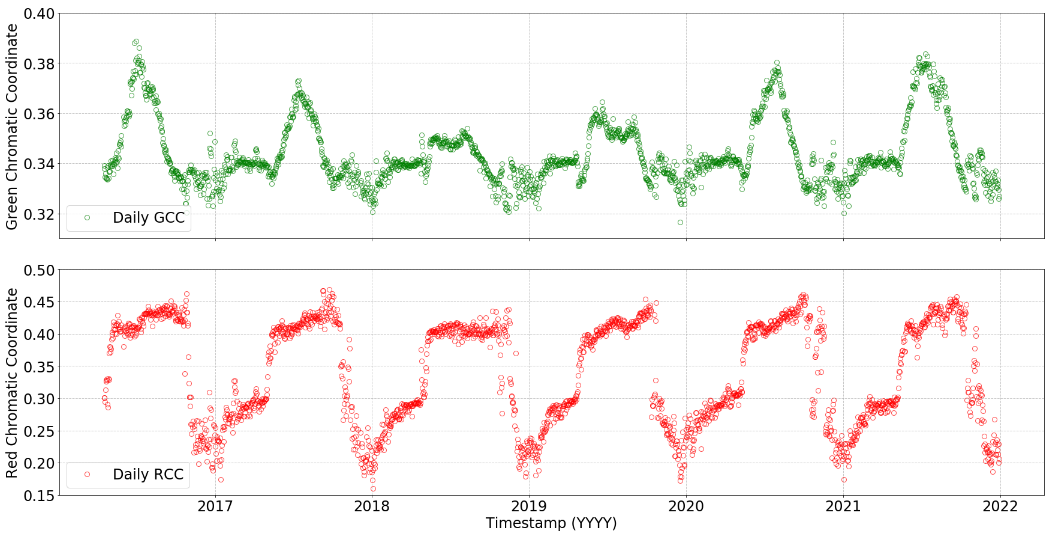


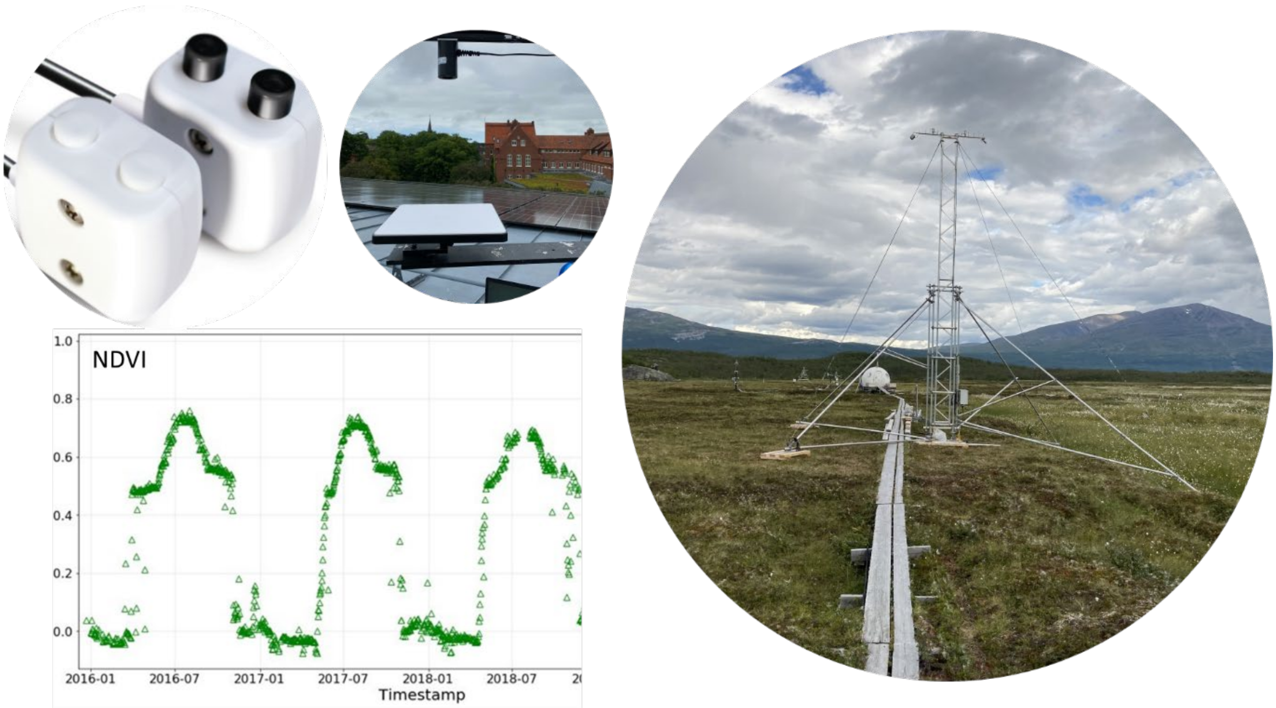
Figure 7. L3 daily averaged time series of GCC (top) and RCC (bottom) during the year 2016 - 2021 derived from quality filtered L1 images collected by phenoCam at SITES Svartberget research station, Degerö.

The plot above shows 6 years of daily average GCC and RCC data from phenoCam located at Degerö, Svartberget research station. The phenoCam is pointed to a nutrient poor minerogenic mire dominated by flat mire lawn plant communities with bog mosses. GCC is low in the winter, increases sharply in the spring, drops off gradually over the summer followed by sharp drop in the autumn. RCC, on the other hand is low in the winter, increases gradually during the spring, remains stable throughout the summer, peaks in the autumn and falls sharply eventually. And these phenology patterns repeat year after year, with low to high degree of variation in the timing of green-up and green-down phenophase events, depending on how the weather progresses in that year, generally advanced spring green-up and delayed green-down in case of warmer weather (Richardson et al. 2018b).

In the given time series of daily averaged GCC data, especially during the recent 2018 drought year in Sweden, the GCC time series appears very disturbed as the values are very low compared to the year before and the impact seems to be followed the year after with signs of returning to normal. This disturbance in the time series of GCC would ultimately affect the phenophase transition dates estimated from it.

There is no gap filled, smoothed time series of daily averaged GCC and RCC data available in the data portal. There are many time series smoothing algorithms available for reducing noises in the data such as spline smoothing, Savitzky-Golay fitting, double logistic function, piecewise polynomial function, and gaussian fitting function etc. All these different fitting functions can reduce noise and improve the signal quality but there is no such single functions that best fits all the time series data (Cai et al. 2017).

Chapter 2. Spectral Reflectance Sensors (SRS)



2.1 Spectral Reflectance Sensors (SRS) Network

Light sensor network in the support of optical data sampling in a certain narrow or broad spectral bands provides a complement to monitor vegetation growing status for estimating the biogeochemical processes (Eklundh et al. 2011; Jin and Eklundh 2015).

They are increasingly used for vegetation monitoring by measuring the reflectance in different wavelength bands. Long term spectral measurements from such networks provide reference data for evaluating various satellite derived biophysical parameters (Fensholt et al. 2004) and helps in establishing a link between earth observation satellite based measurements and ground observations (Eklundh et al. 2011). The spectral data collected from such sensor networks have demonstrated the effectiveness of monitoring the vegetation phenology and productivity. Furthermore, the spectral measurements improve the understanding of vegetation dynamics and can very well capture the disturbance impact (Olsson et al. 2017).

SITES Spectral has 15 such SRS sensor set up at 6 SITES stations (*Table 3*), located at different locations across Sweden recording the incident and reflected radiations in different spectral bands, every 10 seconds and averaged every 10 minutes, and all year long of different ecosystems. Most participating stations use a two channel (R-NIR) sensors either Skye or Decagon model. The sensors at forest locations within the research stations have a 4 channel (R, NIR, 2 REG bands) sensors. And lastly, the mire locations within the stations have a 4 channel including SWIR band apart from R, NIR and REG. The sensor readings collected are uploaded to the Lund University server for archiving, processing, and dissemination of different data products to the SITES data portal.

*Table 3. Fixed Sensor location at each SITES station with the sensor models, number of sensor pairs, wavelength specification and the year when the sensors started observations. Note * for Asa refers that there is no data from 2020 onwards as they are upgrading the mast * for Tarfala suggests that it is no longer part of SITES since 2023.*

Stations	**Fixed multispectral sensor**	**Model – No. of Pairs**	**Wavelength (nm)**	**Start year**
Abisko	SWE-ANS-MJA-HEA-F01	Decagon – 2	650, 810	2016
	SWE-ANS-SSE-MIR-F02	Skye – 1	704, 740, 858, 1640	2016
	SWE-ANS-SSW-MIR-F01	Skye – 1	530, 570, 650, 810	2015
Asa	SWE-ASA-NYB-FOR-F01	Decagon – 2	650, 810	2016
Lönnstorp	SWE-LON-SFA-AGR-F01*	Decagon – 2	630, 800	2016
	SWE-LON-SFA-AGR-F02*	Decagon – 1	630, 800	2016
	SWE-LON-SFAB-AGR-F01	Skye – 3		2022
Skogaryd	SWE-SRC-CEN-FOR-F01	Skye – 2	469, 552, 644, 856	2015
			704, 740, 860, 1636	

	SWE-SRC-CEN-FOR-F02	Decagon – 1	650, 810	2019
	SWE-SRC-CEN-FOR-F03	Decagon – 1	650, 810	2019
	SWE-SRC-FOL-WET-F01	Decagon – 2	650, 810	2017
	SWE-SRC-MYC-MIR-F01	Skye – 2	704, 740, 860, 1636	2018
		Decagon – 2	531, 570, 650, 810	2017
	SWE-SRC-STD-FOR-F01	Decagon – 1	532, 570, 650, 810	2017
Svartberget	SWE-SVB-DEG-MIR-F01	Skye – 2	704, 740, 858, 1640	2018
			532, 570, 645, 858	
		Decagon – 1	532, 570, 650, 860	
	SWE-SVB-SVB-FOR-F01	Decagon – 2	650, 860	2019
		Skye – 2	531, 570, 640, 860	
Röbäcksdalen	SWE-RBD-RBD-AGR-F01	Decagon – 2	650, 810	2016
	SWE-RBD-RBD-AGR-F02	Decagon – 1	630, 810	2016
		Skye – 1		2023
Tarfala	SWE-TRS-LAE-GRA-F01	Decagon – 4	630, 800	2016*

2.2 Spectral Reflectance Sensors data

The spectral reflectance sensors (SRS), often referred to as fixed multispectral sensors are the multiple band radiometers designed to measure the incident and reflected radiation in various spectral bands in visible electromagnetic spectral range and beyond. The visible bands SRS used within SITES stations include R, G, and B while there are also infrared bands such as near-infrared (NIR) and shortwave infrared (SWIR) bands. Some of the sensors within the network also has the red edge (REG) bands that matches the wavelength configuration with some of the widely used satellite systems such as MODIS and Sentinel 2. The choice of the bands in sensor configuration varies from station to stations based on the characteristics of the stations.

The SRS pair measures the incident and reflected radiation in spectral wavelenghts band suitable for the computation of one of the commonly used vegetation indices such as the Normalized Difference Vegetation Index (NDVI) or Enhanced Vegetation Index (EVI). These sensor pairs are manufactured differently such as the hemispherical and field-stop versions. The hemispherical versions (*Figure 8, a*) are built with diffusers for making cosine-corrected measurements, and are primarily designed for up-looking measurements of incident radiation.

The field stop versions (*Figure 8, b*) have a narrow/restricted FOV and are designed for pointing downward to measure canopy reflected radiation in the sensor configured wavelength bands.

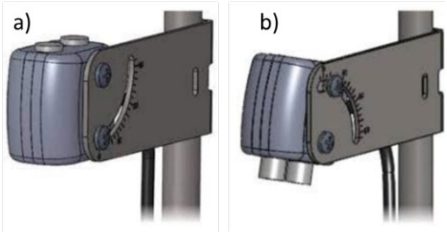


Figure 8. One instance of spectral reflectance sensor for measuring incoming and reflected radiation using the cosine corrected hemispherical sensor (a) and field-stop lens sensor (b). Source: http://www.metergroup.com

To calculate NDVI or any other vegetation indices, it is required knowing both the incoming and reflected radiation. Unlike the reflected radiation, the incoming radiation is spatially uniform over an area or above the canopy. So, one up-looking radiometer can be used as a pair to many down-looking sensors to compute the VIs (*Figure 9*).

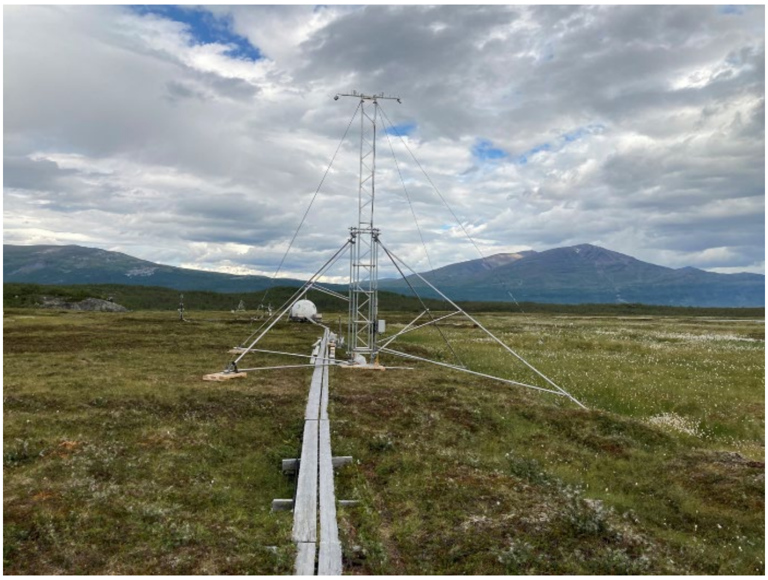


Figure 9. Photograph of the Stordalen mire site located at SITES Abisko Scientific Reseaarch Station on August 6, 2021, showing the spectral mast along with the up and down-looking sensors. The sensors are viewing the vegetations on the dry and wet mires.

All the SRS sensor pairs within SITES stations are co-located (mostly) on the same mast together with phenoCam and have overlapping ROIs with ground projected footprint of the phenoCam sensors. The length of these masts varies between a couple of meters to hundreds of meters. It is ensured that the up-looking radiometers are levelled and they have a hemispherical field of view while the down-looking radiometers are inclined at an oblique angle of off-nadir 45° at all reserach stations within SITES. The up-looking sensor has wide FOV (180°) while the down-looking sensor has confined FOV and varies between sensor types (for example, 36° for decagon sensors). Down-welling and up-welling sensors are programmed to measure the incoming and reflected radiation simultanteouly every 10 seconds, at a wavelength band specified and are aggregated to every 10 minutes interval. These incoming and reflected radiations are then uploaded to the Lund University server for archiving, processing, and dissemination of different data products to the SITES data portal.

SITES offers the archived data in different levels and format i.e., from the raw level 0 (L0) data product at original every 10 minutes resolution to the analysis ready quality flagged time series level 2 (L2) data. All these different data levels are made available through the SITES data portal (https://data.fieldsites.se/portal/). *Table 4* lists all available data levels in the data portal for the fixed multispectral sensors.

Table 4. List of all fixed multispectral sensor data levels and their names in the data portal

Level	Product names
L0	Sun irradiance and vegetation radiance
L1	Calibrated reflectance and spectral vegetation indices
L2	Daily averaged reflectance and spectral vegetation indices

2.1.1 Data level L0 – Sun irradiance and vegetation radiance

The archived L0 data are the original raw logger data which are the voltage readings (in mV units) or the nominal radiance (in $W * m^{-2} * sr^{-1}$ or $\mu mol * m^{-2} * sr^{-1}$) from the multispectral sensors that are averaged and available at original 10 minutes temporal resolution. The missing data are defined as not-a-number (NaN). The data initially are exported as (*.dat*) file with 4 header rows followed by data organized in columns, one row per 10 minutes average. There can be multiple columns based on the number of channels available within a particular sensor. For example, *Table 5* shows the typical raw data L0 file columns for a particular station.

Table 5. Raw data (L0) file columns showing general information available for a pair of 4 channel fixed multispectral sensor in the data table. Note that the units for channel readings might differ from sensor to sensor.

Name	Contents	Units
TIMESTAMP	Time stamp, local time zone without summer time adjustment.	
RECORD	Record number of observations	
BattV_Min	Minimum battery level	V
PTemp_C_Avg	Logger temperature	degree C
Up_Ch1_Avg	Channel 1 incoming radiation	mV
Up_Ch2_Avg	Channel 2 incoming radiation	mV
Up_Ch3_Avg	Channel 3 incoming radiation	mV
Up_Ch4_Avg	Channel 4 incoming radiation	mV
Up_Temp_Avg	Temperature of sensor for incoming radiation	degree C
Dw_Ch1_Avg	Channel 1 reflected radiation	mV
Dw_Ch2_Avg	Channel 2 reflected radiation	mV
Dw_Ch3_Avg	Channel 3 reflected radiation	mV
Dw_Ch4_Avg	Channel 4 reflected radiation	mV
Dw_Temp_Avg	Temperature of sensor for reflected radiation	degree C

Later, these files are exported as a .csv files for each year and made available in the data portal. No processing steps are involved at this stage. *Figure 10* shows the plot of raw L0 channel readings from a pair of sensors.

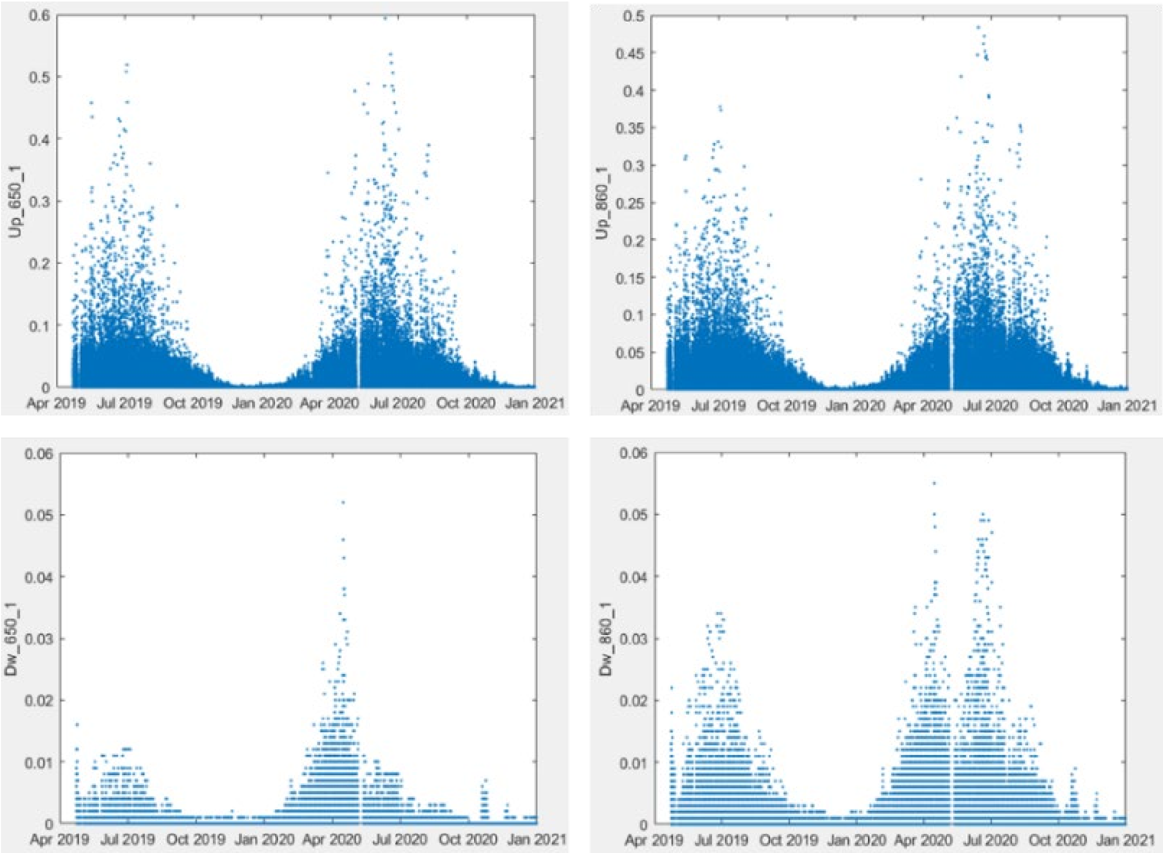


Figure 10. Raw data plot for a pair of up and down-looking fixed multispectral Decagon sensors at SITES Svartberget Research station showing incoming and reflected radiations during 2019-2020. Up_650_1, Up_860_1 are the irradiance data and Dw_650_1, Dw_860_1 are the radiance data. Both data are in mV units and represent two channels: red channel at 650nm and NIR channel at 860nm respectively.

2.1.2 Data level L1 – Calibrated reflectance and spectral vegetation indices

The L1 product is named as calibrated reflectance and spectral vegetation indices in the data portal. At this stage, the raw data L0 are processed and transformed into physical values of reflectance at original timestamp. During this process, the data is treated with calibration factors for each channels available for a particular sensor. The calibrated reflectance data is hence used to calculate different spectral vegetation indices.

The process of creating L1 time series data starts with checking to see if there are any offsets observed in the raw (L0) data. Offsets observed in the original L0 values are subtracted from the values to remove the offsets. The period with missing data are filled with NaN values. The gain values that are obtained from the sensor calibration procedure are applied in the calculation of calibrated reflectance for the time period following the calibration until the next calibration using *equation 3* as shown below:

$$R_{Cal} = \frac{R_L}{k} \cdot \frac{V_2}{V_1} \quad (3)$$

where R_{Cal} is the estimated calibrated channel reflectance, R_L is the standard reflectance of the lambertian reflectance panel used in the calibration, k is the obtained calibration coefficient for the given channel, V_2 is the observed reflected radiation in mV and V_1 is the observed incoming radiation in mV. For a full description of the calibration procedure see Jin and Eklundh (2015).

The non physical reflectance values at this stage i.e., calibrated reflectance values above 1.0 or below 0.0 are replaced with a NaN values.

Finally, the calibrated reflectance values are used to compute different spectral vegetation indices. The computation of spectral vegetation indices are based on the type of the spectral channels available for a given sensor pairs. *Table 6* lists out common VIs computed with the data collected by fixed multispectral sensors at the research stations within SITES.

Table 6. Lists of spectral vegetation indices computed (varies by station depending on available spectral channels on a fixed multispectral sensors).

Index	Equation	Reference
Normalized difference vegetation index	$NDVI = \frac{(NIR - R)}{(NIR + R)}$	(Rouse et al. 1973)
Normalized difference wetness index	$NDWI = \frac{(NIR - SWIR)}{(NIR + SWIR)}$	(Gao 1996)
2-band Enhanced Vegetation Index	$EVI2 = 2.5 * \left(\frac{NIR - R}{NIR + 2.4 * R + 1}\right)$	(Jiang et al. 2008)
Photochemical Reflectance Index	$PRI = \frac{(G_{531} - G_{570})}{(G_{531} + G_{570})}$	(Gamon et al. 1997)
Scaled Photochemical Reflectance Index	$\text{sPRI} = \frac{(1 + \text{PRI})}{2}$	(Rahman et al. 2004)
Difference vegetation index	DVI= $NIR - R$	(Richardson and Wiegand 1977)
Normalized difference red edge index	$NDRE = \frac{(NIR - REG)}{(NIR + REG)}$	(Gitelson and Merzlyak 1994)

The fixed multispectral sensors at SITES stations are calibrated regularly i.e. every 2 years to ensure the quality of the spectral data being collected. This is later on used as one of the ways to flag up the quality code values for the L2 daily averaged reflectance and spectral vegetation indices. Both the reflectance and spectral vegetation indices computed at this stage are unitless. The L1 data are available in the SITES data portal as a comma separated *.csv* file. In addition the L1 data files contain all metadata information as headers.

Figure 11 shows one example of time series of L1 data.

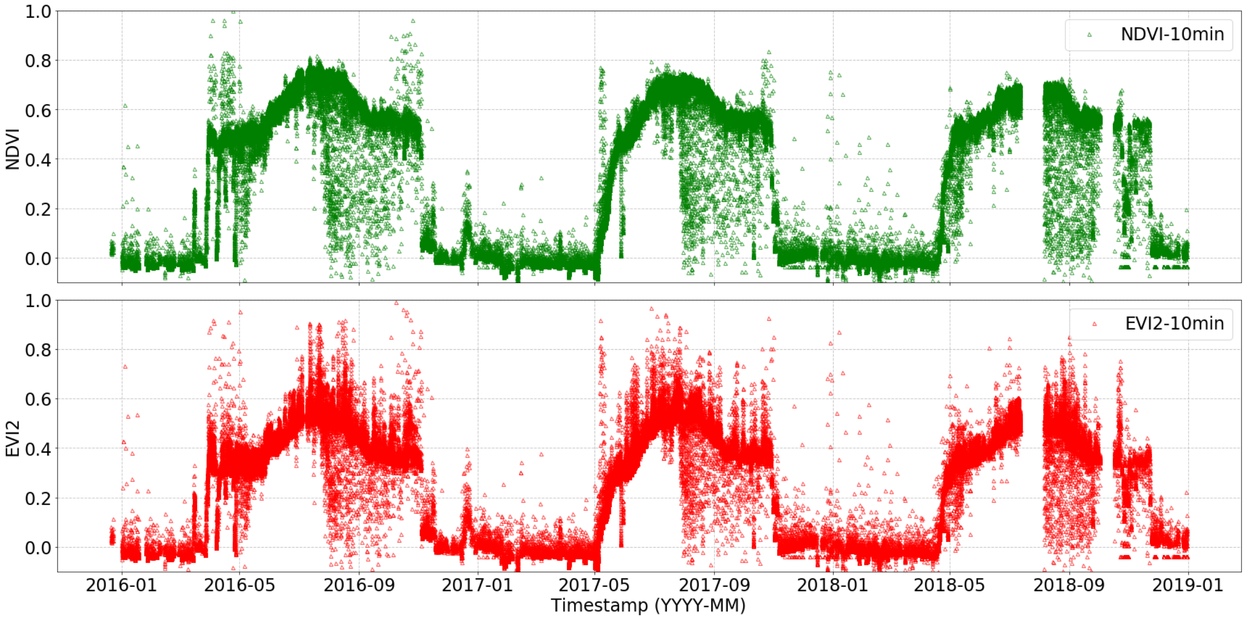


Figure 11. Time series of spectral vegetation indices such as NDVI (top) and EVI2 (bottom), computed using calibrated R and NIR reflectances at original timestamp (every 10 minutes). The data for this plot is collected at Stordalen mire located at SITES Abisko Scientific Research station. The gaps in the data refers to the period with no data due to power supply failure.

2.1.3 Data level L2 – Daily averaged reflectance and spectral vegetation indices

The multispectral data level L2 is derived from the data level L1 by averaging the calibrated reflectances and the spectral vegetation indices based on the type of sensors used on a daily basis. L2 time series data is named as daily reflectance and spectral vegetation indices on the data portal. The data files are made available as .csv files along with the metadata as headers.

As seen on the plot (*Figure 11*), the L1 data still has lots of noises. So, in order to get rid of these noise, first of all the L1 data is treated with filtering technique like the one applied for phenoCam pictures. The time filter is applied to the L1 data and the time window at this stage is the same i.e. all the L1 data within 10:00 - 14:00 is filtered out. These time filtered L1 data are then aggregated on a daily basis to compute daily average reflectances and spectral vegetation indices based on the channels available in the given sensor pairs. The daily average is computed only if there are enough values that are not NaN (if less than 6 values per day, the daily average is defined as NaN). In addtion to this, there will be a separate column with a three digit quality flag value for each daily averaged values. The data frequency, sensor calibration date and the sensor degradation over the years are taken into account when flagging up quality code values (See *Chapter 5* for information on quality flagging). The final daily aggregated reflectances for each channels and spectral VIs are exported to a *.csv* file. Furthermore, the metadata headers are added to the data file before uploading it to the data portal.

An example of daily averaged vegetation indices (e.g. NDVI and EVI2) for one pair of fixed multispectral sensors at SITES Abisko scientific research station is given in *Figure 12*.

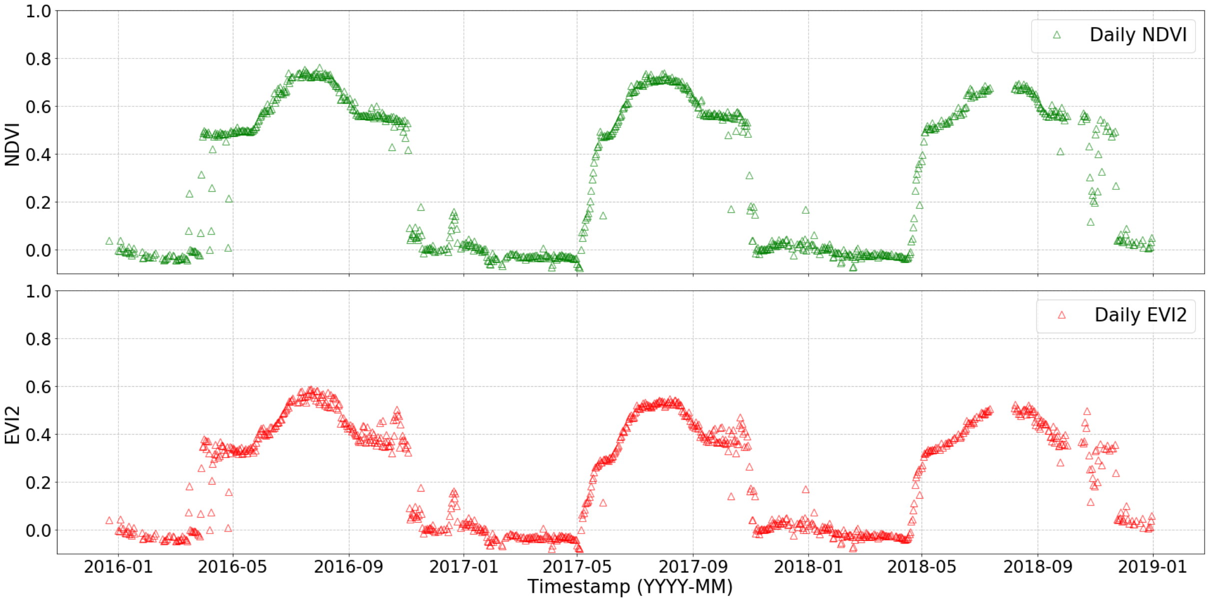
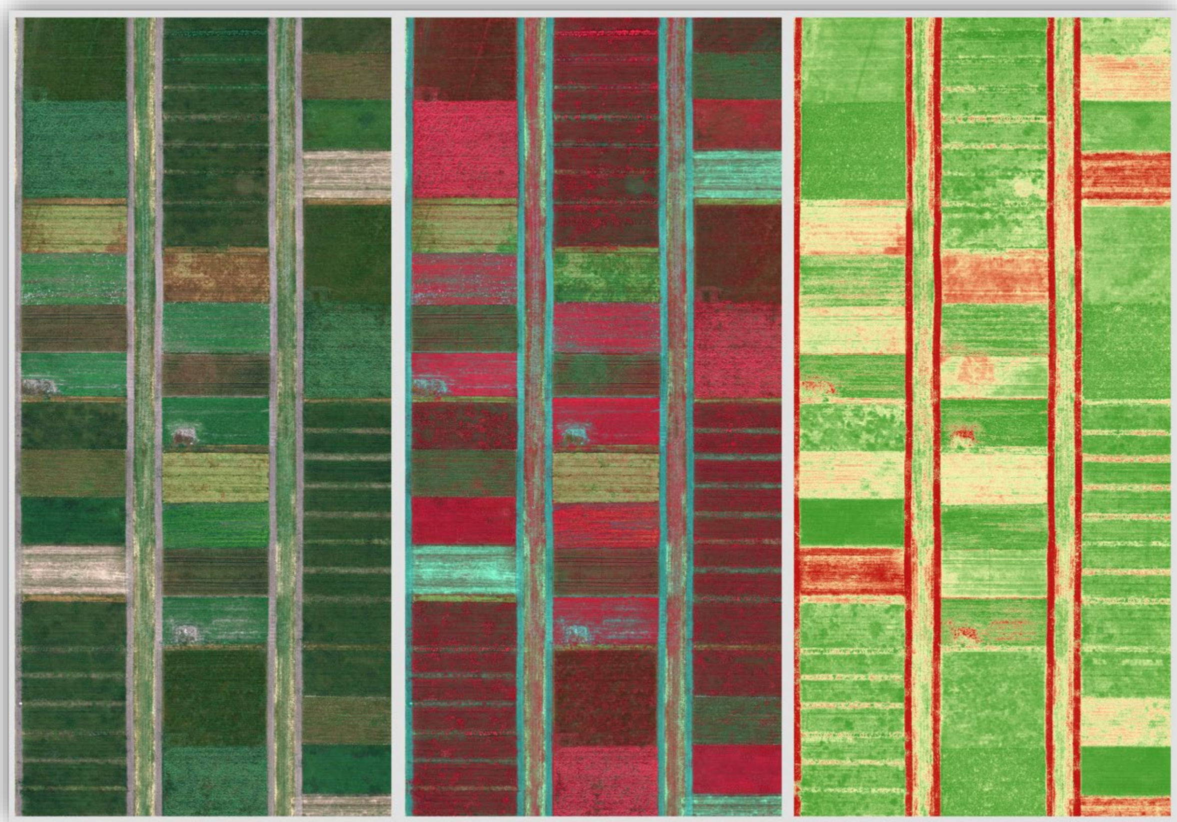


Figure 12. Time series of spectral vegetation indices such as NDVI (top) and EVI2 (bottom), computed using calibrated R and NIR reflectances aggregated on a daily basis. The data for this plot is collected at Stordalen mire located at SITES Abisko Scientific Research station. The gaps in the data refers to the period with no data due to power supply failure.

Chapter 3. Unmanned Aerial Vehicle (UAV)



3.1 UAV Data

SITES research stations carry out periodic UAV flights every year throughout the vegetation growing season using different RGB and multispectral sensors. The frequency of UAV flights varies from station to station based on the need of the station. Most of the stations fly once a year while there are couple of stations that conduct drone flights multiple times throughout the vegetation growing season. Based on the type of sensors onboard the drone, the flight data collected are processed in Agisoft Metashape software (Agisoft, St. Petersburg, Russia) for generating multiple data products.

In the earlier years of SITES, GoPro cameras were used for conducting RGB flights. For multispectral data, Parrot Sequoia and Mica Sense Red Edge sensors served for a couple of years at most of the stations at the beginning. However, the stations started to use DJI P4 multispectral sensors since 2021 because of its robustness and easy to use facilities. Parrot sequoia and mica sense sensors needed to be mounted onboard drones and to do this on a regular basis was time consuming and not very efficient compared to one button take off to mission of DJI P4 sensors. These UAVs are equipped with sunshine sensors which collects amount of incoming light at the time of image capture. This makes it easy to keep track of the light conditions variation throughout the entire flight. These UAVs are also equipped with RTK GPS facilities enabling to achieve high geometric accuracy.

There are different data products that could be derived from UAV acquired images based on the type of sensors being used. *Table 7* lists all the data products defined for multispectral UAV flight data.

Table 7. List of all multispectral UAV data levels and their names in the data portal

Level	Product names
L0	UAV MSP flight datasets
L1	Enhanced MSP images
L2	MSP Point Clouds MSP Orthomosaic NDVI Layer
L3	Digital Surface Model (DSM)

3.1.1 Data level L0 – UAV MSP flight datasets

This data product is simply the raw multispectral images acquired during each UAV flight mission. The number of images for each flight differs from station to station based on the flying height, frontal and side overlap, area coverage and type of sensors used. Parrot Sequoia sensor collects multispectral data in four spectral bands namely G, R, REG, and NIR while Micasense and DJI P4 sensor has also B band in its sensor configuration. All the flight images are available in *.tif* format.

In addition to the flight images, the data product also contains images of the reflectance calibration targets captured both at the start and end of each UAV flight. For geometric correction, a well distributed and sufficient number of ground control point (GCP) coordinates are measured with high geometrical accuracy in SWEREF99_TM projection system. The GCPs are available in a *.txt* file alongside with this data product. For the UAVs enabled with RTK

GPS, no need to establish the GCPs as the camera center coordinates are corrected using the Swedish GPS network SWEPOS.

3.1.2 Data level L1 – Enhanced MSP images

The raw multispectral images L0 from the research stations are copied to a common folder and renamed as per the SITES naming convention. These images are then prepared for applying various corrections. The corrections applied to the images at this stage are: a) Exposure calibration, b) Vignetting correction, and optionally c) irradiance normalization.

First, the exposure calibration and vignetting correction were perfomed on all the individual images of a flight by using the methods described by Parrot (Parrot SEQ AN 02) and DJI P4 (P4 Image Processing Guide). These two corrections are made based on the information derived from the image metadata (EXIF/XMP). These corrections are applied with in-house developed Python scripts.

Vignetting is the radial decrease in pixel values that makes image to appear darker on the edges (Goldman 2010). The vignetting effect is corrected by dividing the exposure-calibrated images with the vignetting polynomial, where the polynomial is derived from the EXIF/XMP image metadata. The vignetting effect differs from sensor to sensor and the correction factor is derived based on the image processing guides provided by the sensor manufacturers.

For irradiance normalization, first of all the irradiance information for each images per band are extracted and saved in a *.xlsx* format. Based on the nature of the raw sunshine sensor based irradiance data plot for each band, it is decided whether the given flight needs or can be corrected for variation in irradiance throughout the UAV flight. The irradiance data pattern varies based on the variation in weather conditions during the UAV flight. Based on when (timestamp) and the weather conditions during the UAV flight, the flight is classified into three categories: a) smooth and consistent irradiance pattern on a completely overcast or sunny condition, b) slight variation in irradiance data and c) complex irradiance pattern during mixed weather conditions. There is no need to apply the concept of irradiance normalization for flights that shows no variation at all during the UAV flight conducted on a complete overcast conditions or for the flight on a sunny condition when the variability correlates with the flight direction. However, if there is slight variation in irradiance data as in *Figure 13*, the data can be normalized. The irradiance data that looks very complex (high variation) throughout the UAV flight is processed further without irradiance normalization. The quality of multispectral UAV flight data is impacted by how well the variation can be normalized or not.

The different multispectral flight data with varying sunshine sensor irradiance data cases and how they impact the quality of the UAV data products are discussed in Chapter 5.

Irradiance normalization is perfomed on the individual images after the exposure calibration and vignetting correction (Olsson et al. 2021). However, the raw irradiance data from the sunshine sensor are noisy and sensitive to sensor orientation (yaw, pitch, roll) since they do not have a cosine corrector. Hence, polynomials or splines are fitted to smooth the raw sunshine sensor data, and the value of the polynomial or spline, from here on called smoothed irradiance, is used to normalize the images instead of the raw data. *Figure 13* shows raw sunshine sensor data and fitted splines in cloudy conditions, i.e., with mainly diffuse light (a, c), and raw sunshine sensor data and a fitted polynomial in mainly sunny conditions with thin clouds, i.e., with mainly direct light (b), where the variability correlates with flight direction. The choice of polynomial or spline fit was made on the basis of a visual inspection by manually fitting both the polynomial and splines to the data, and selecting the best fit thereafter.

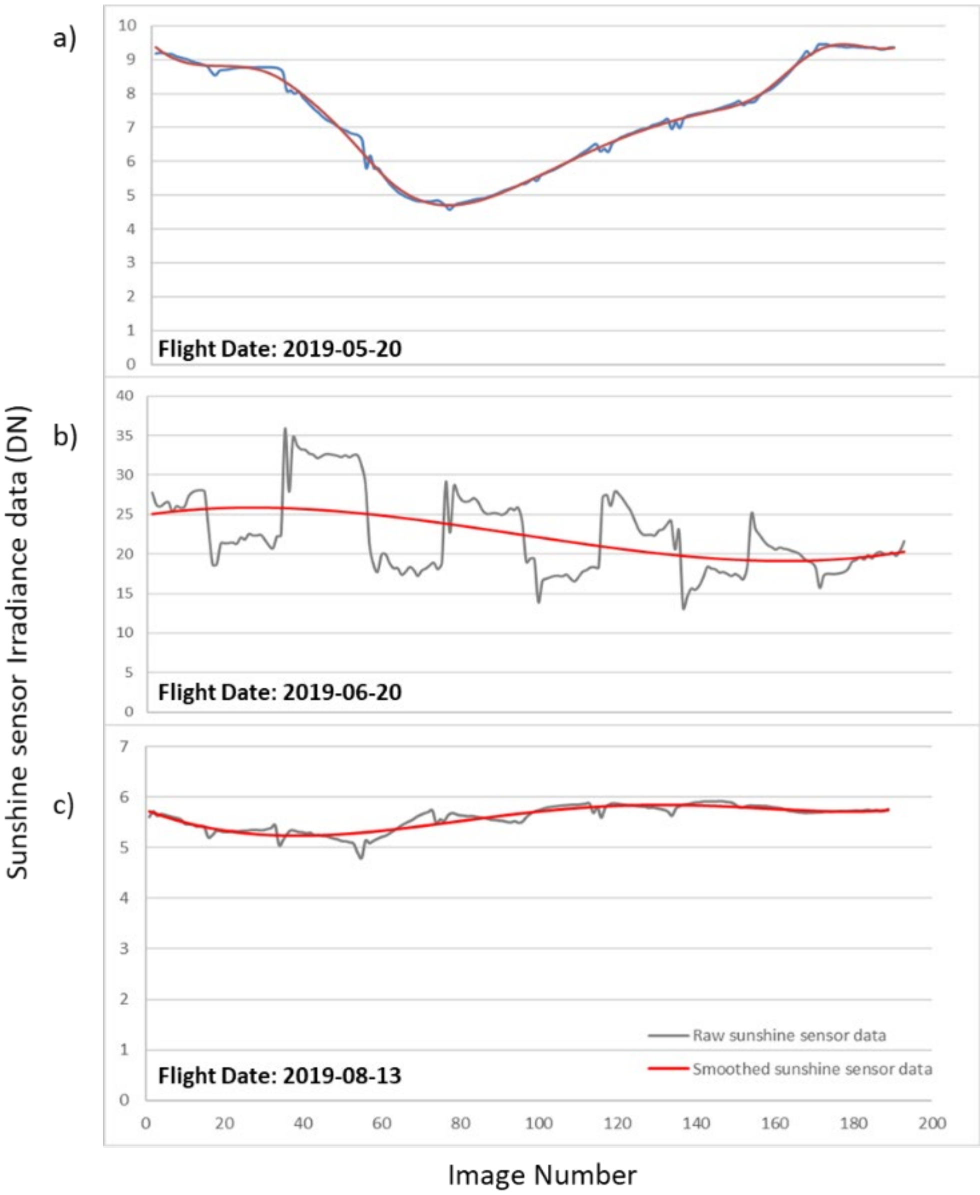


Figure 13. Raw (gray) and smoothed sunshine sensor data (red) for the MSP UAV flights conducted at SITES Lönnstorp Research station on a) 20 May 2019, b) 20 June 2019, and c) 13 August 2019 (Olsson et al. 2021).

The image-specific irradiance normalization factor ($C_{in,i}$ *equation 4*) was then calculated for each image.

$$C_{in,i} = \frac{I\,(\lambda)}{I_i(\lambda)} \tag{4}$$

where $C_{in,i}$ is the irradiance normalization factor for image i, I(λ) is the mean smoothed irradiance for wavelength band λ for all images in a flight, and $I_i(\lambda)$ is smoothed irradiance from the sunshine sensor for image i. The irradiance normalization was then performed by dividing all pixels in an image with the image specific normalization factor ($C_{in,i}$) . To minimize the normalization factor and avoid large modification of images, all images were normalized against the mean smoothed irradiance for the entire flight.

The images after the exposure, vignetting and irradiance normalization (optionally) are exported to the same *.tif* extension followed by copying the original metadata using EXIF tool and they are zipped together along with the GCP files. This *.zip* file is uploaded to the data portal.

A common problem to the images at this stage is the saturation of the pixels. Based on the sensor used the pixels can be found to be saturated in B, G R, and RE wavelength bands. Generally, these saturated pixels are over the bright areas within the flight location such as the gravelled roads, rocks which are not that important from the perspective of vegetation monitoring. So, while performing the irradiance normalization, an effort is made to create a mask file for each image masking out the saturated pixels. This mask file is used later during the orthomosaic generation step to avoid the influence of saturated pixels on the orthomosaics.

3.1.3 Data level L2 – MSP Point clouds, Orthomosaic, NDVI and L3 – DEM

The enhanced multispectral images are imported into Agisoft Metashape for further processing and generating multiple data products. Metashape can be launched in LUNARC Aurora by using the following commands in the Terminal (LUNARC documentation for more details):

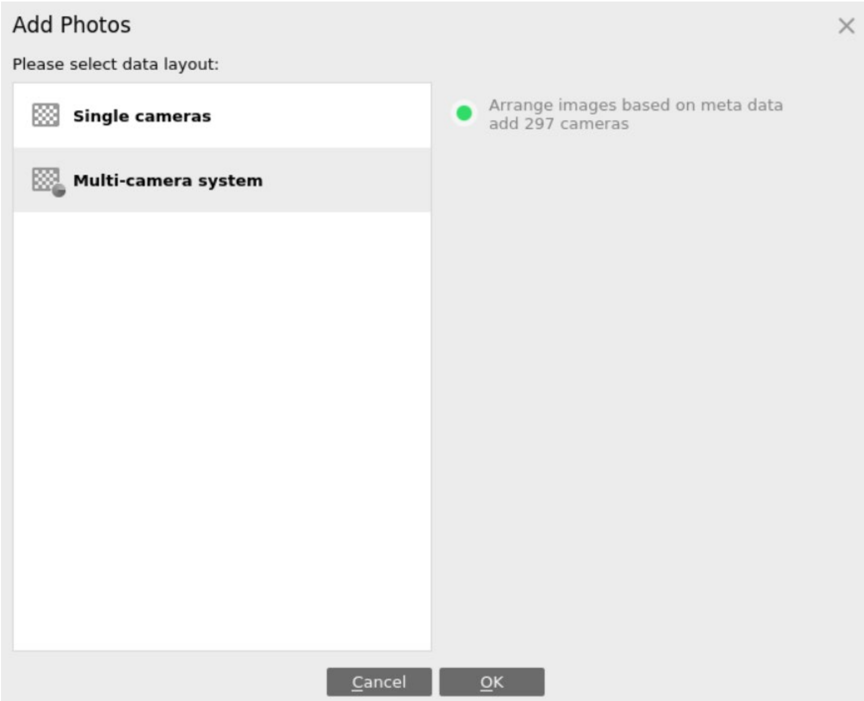
module load metashape
MS-start-GUI

Metashape supports the processing of multispectral images captured with commonly used sensors like DJI P4, Parrot sequoia and Micasense. The step by step image processing methods followed to process and derive different photogrammetric products within SITES are explained below:

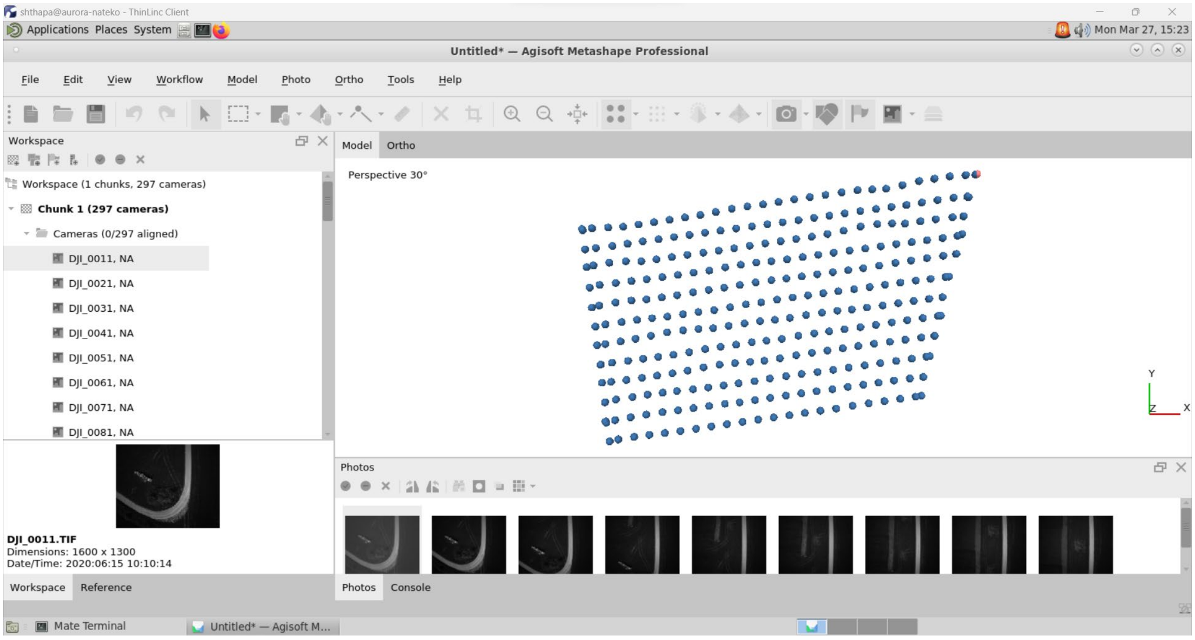
- **a)** Add Photos
- **b)** Calibrate Reflectance
- **c)** Align Photos
- **d)** Camera Calibration
- **e)** Optimize Cameras
- **f)** Build Point Cloud
- **g)** Build DEM
- **h)** Calibrate Colors
- **i)** Build Orthomosaic
- **j)** Raster Transform
- **k)** Export products
- **l)** Radiometric calibration

3.1.3.1 Add Photos

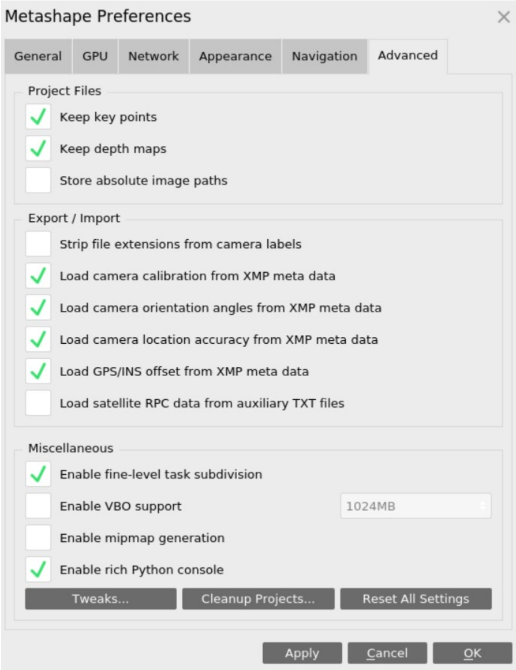
In order to add multispectral images, select *Add Folder...* command from the *Workflow* menu. In the *Add Photos* dialog box browse and select all images from the source folder to be processed. Select *Multi-camera system* option in the data layout section after that and hence Metashape will automatically arrange the bands according to the information available in the image metadata.



As soon as the images are loaded, the number of images imported in the *Workspace* pane are visible and the coordinate information will be automatically loaded from the image metadata to the *Reference* pane in the bottom left corner of the screenshot below.



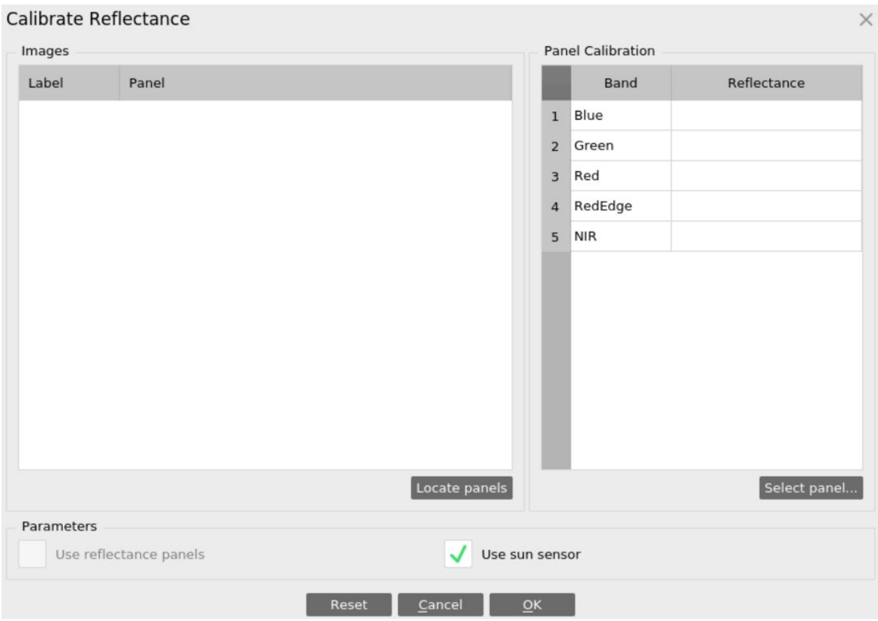
It is suggested that prior to adding images prior to adding images to the project the following options are to be enabled in the *Advanced* tab in the *Preferences* dialog (select *Tools* > *Preferences...*): *Load camera orientation angles from XMP metadata* and *Load location accuracy from XMP data*.



If this data is missing, then set accuracy (m) manually. Select *Set Accuracy...* command from the images context menu on the *Reference pane* and input accuracy data both for a position (i.e. x,y,z coordinates) and orientation (i.e. [yaw, pitch, roll] angles) data. It is possible to select several cameras and apply *Set Accuracy...* command simultaneously to all of them.

3.1.3.2 Calibrate Reflectance

For the flights without ground reflectance calibration targets or the flights that are not normalized for variation in irradiance, use only the *Sun sensor* option to be used for the reflectance calibration. To do this open *Tools* menu and choose the *Calibrate Reflectance...* option. Enable *Sun sensor* for calibration and press *OK* button as shown in the screenshot below:



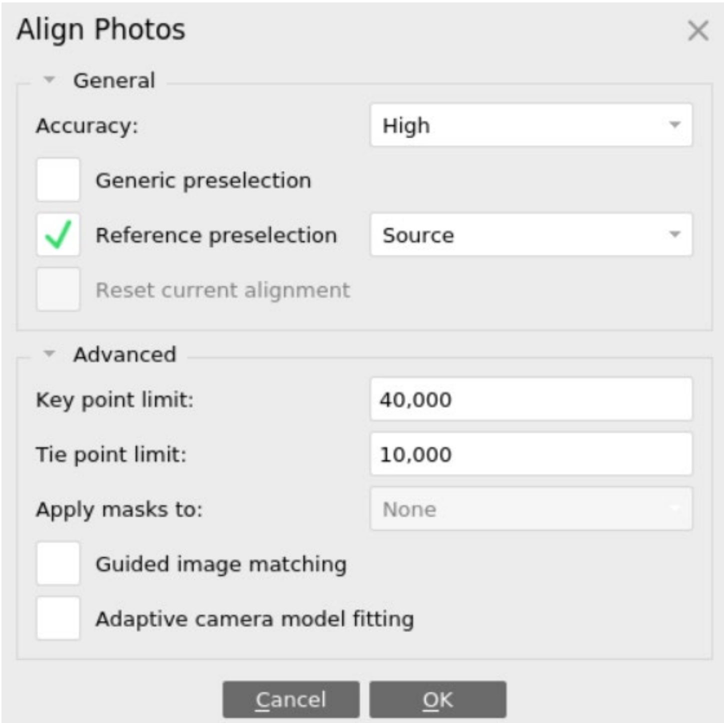
If there is availability of the calibration targets images, they will be used in the radiometric calibration which will be discussed in the radiometric calibration section.

3.1.3.3 Align Photos

Once the images are loaded into Metashape, they need to be aligned where the camera position and orientation for each multi-camera system are estimated. This generates a tie point cloud consisting of the tie points. Additionally, the relative offsets of the sensors in the rig and the interior orientation (calibration) parameters for each sensor are estimated. The image alignment step begins by selecting the *Align Photos...* command from the *Workflow* menu.

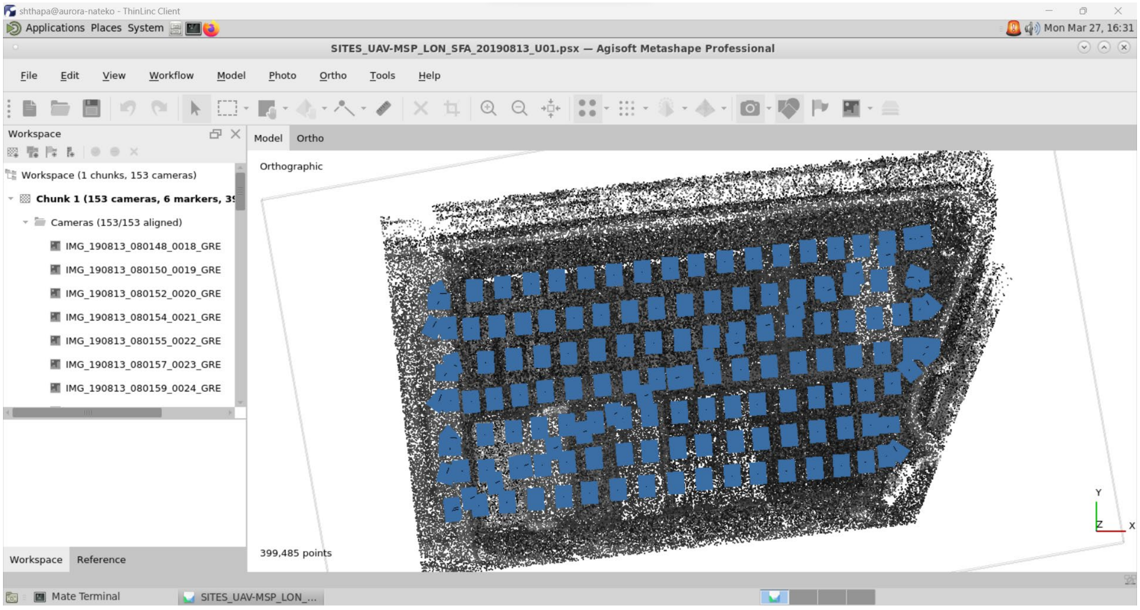
In the *Align Photos...* dialog box select the desired alignment options. The settings presented on the *Align Photos...* dialog box are set based on the particular dataset parameters and can be referred to as example values, the settings may differ depending on the project.

Disabling the *Generic preselection* option may be helpful for the low resolution images of the areas with maximum vegetation cover, such as forests or crop fields.



Under the Advanced drop down button, one can change the key point and tie point limit. For images from homogeneous areas the low tie point limit setting could be changed to a lower values. Click OK button when the parameters are defined. The progress dialog box will appear displaying the current processing status. To cancel the processing click Cancel button.

The result of the align photos operation will be shown in the *Model* view as the estimated camera locations and tie point cloud (representing the matching points between the images):



Each of the blue polygon in the *Model* view represents the camera location where the images are captured. Behind the camera polygons are the tie points generated after the *Align Photos...* step. If necessary, adjust the reconstruction volume (bounding box) in the *Model* view mode using *Rotate Region, Resize Region,* and *Move Region* instruments.

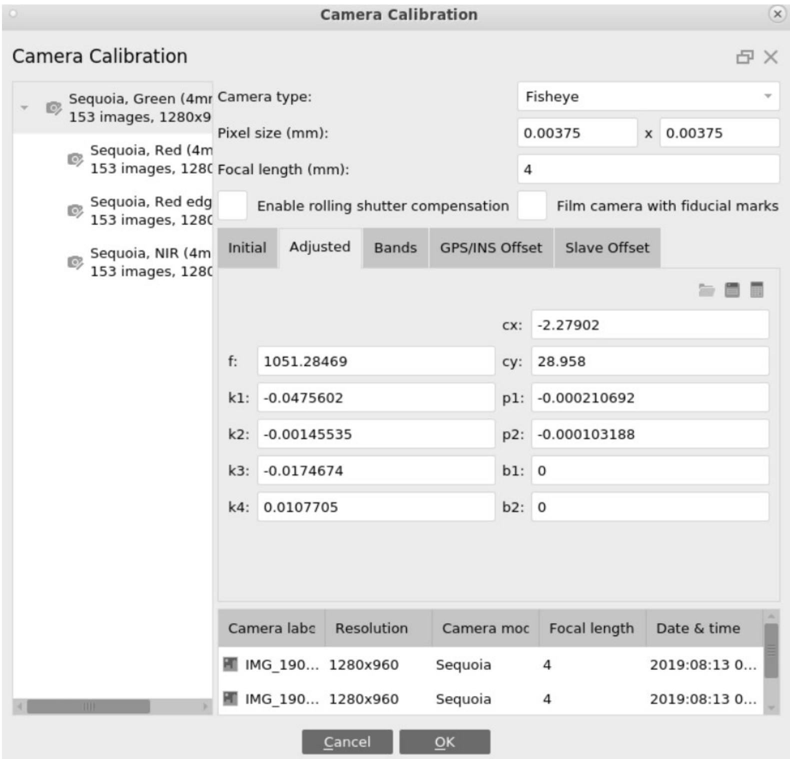
If the flight was conducted with placing well distributed and sufficient number of GCP markers on the flight area, then those points are imported here and geometric correction of the tie points are carried out based on the highly accurate GCP marker coordinates. This should be done before optimization. Under the *Reference* tab, you can choose the *Import Reference* and navigate to choose a comma separated *.csv* or a *.txt* file for GCP coordinates. For a multi-camera system, it will be enough to specify the projection of the marker on only channel (not necessarily on all).
For RTK enabled MSP UAV flights, if prior to adding photos into the Metashape the following options are enabled in the *Advanced* tab in the *Preferences* dialog (select *Tools > Preferences...*): *Load camera orientation angles from XMP metadata* and *Load location accuracy from XMP data*, then no need to import any GCP markers for geometric correction.

3.1.3.4 Camera Calibration

Metashape supports four major types of camera: frame camera, fisheye camera, spherical camera and cylindrical camera. In order to determine the interior orientation parameters of the cameras, including the distortion parameters of the camera lens, the camera calibration is performed. The camera calibration parameters can be input manually, if they have been acquired as a part of precalibration procedure.

The following calibration parameters can be determined:
f – focal length (measured in pixels)
cx, cy – coordinates of principal point (in pixels)
b1, b2 – affinity and non-orthogonality (skew) coefficients (in pixels)
k1, k2, k3, k4 – radial distortion coefficients (dimensionless).
p1, p2 – tangential distortion coefficients (dimensionless).

During the *Align Photos* processing step, initial calibration data will be adjusted. Once *Align Photos* processing step is finished, the adjusted calibration data will be displayed on the *Adjusted tab* of the *Camera Calibration* dialog box.

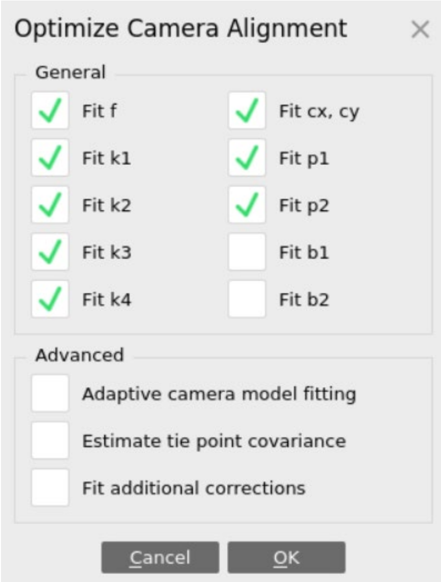


It is often good reviewing the adjusted values for the *cx*, *cy*, *b1* and *b2* parameters in the *Camera Calibration* dialog as it helps to check if the values for these parameters are too large (hundreds or more for *cx*, *cy* and tens or more for *b1*, *b2*). If those values appear to be off the limit as mentioned, then it may be reasonable to re-align the data set having these parameters fixed.
Usually, the values for *cx* and *cy* are around a couple dozens of pixels (or less), as the values are measured from the center of the image. And the values for *b1* and *b2* are usually not more than a few units.
If the values are hundreds or even thousands - it indicates that the estimation of the calibration parameters is probably incorrect. Sometimes it may be related to the flight mission specifics, for example, if the camera is not rotated by 180 degrees on the neighboring flight lines. You can try to fix the mentioned parameters and re-align the data set having "*adaptive camera model fitting*" option disabled in the *Align Photos...* step.

Metashape offers a number of tools to analyze the camera calibration results (for instance: to view the distortion plot). The SITES UAV data processing protocols includes reviewing the camera calibration results and take necessary actions if needed. In addition, at this stage under the *Bands* tab, the *Black level* is set to 0 for all spectral bands as black level is already handled in the L1 data creation stage.

3.1.3.5 Optimize Cameras

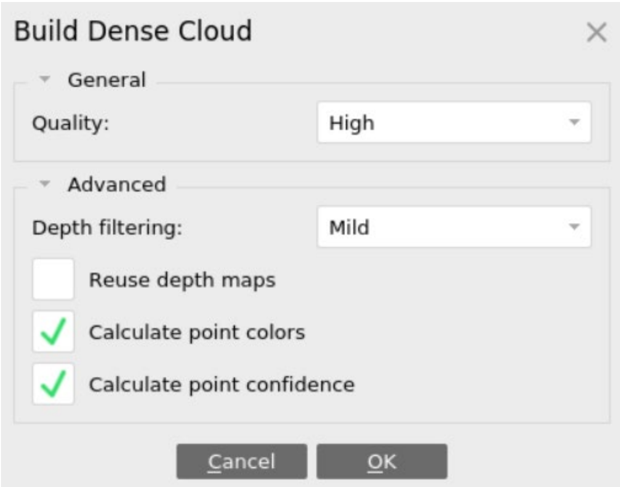
To improve the accuracy of the alignment use, *Optimize Cameras* option from the *Tools* menu and select the parameters for optimization as presented in the image below.



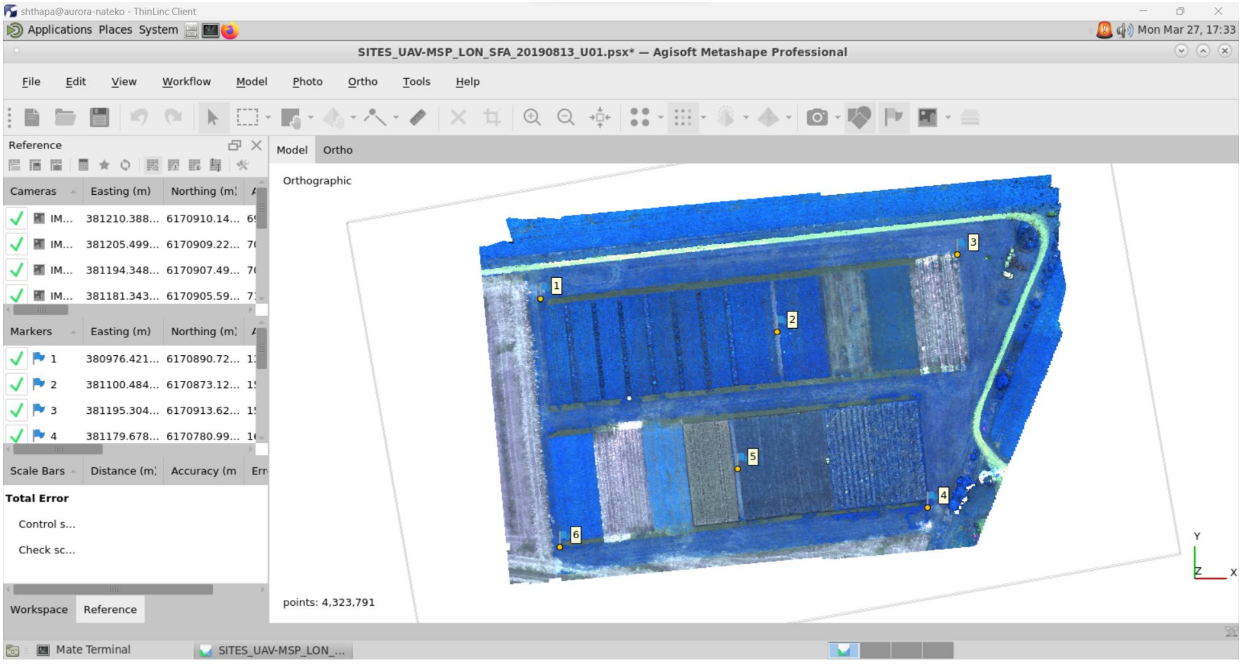
Usually, Fit additional corrections can actually give an improved result for the RTK drone in a project without using GCP. This option allows you to compensate for calibration inaccuracies.

3.1.3.6 Build Point Cloud

After camera optimization process, select the *Build Point Cloud…* command from the *Workflow* menu. A Build Point Cloud dialog box will open where one could select the desired reconstruction parameters. The settings defined on the screenshot below are based on particular dataset and can only be referred to as example values. They can differ from project to project. Once, the settings are defined, click OK to proceed further. The progress bar dialog box appears after that and displays the current processing status. There is also an option to cancel the processing.



The *Quality* (image resolution) parameter can be set to highest setting i.e. *Ultra high* or to the *Lowest*. On the *Advanced* tab, the depth filtering parameter can be set to *mild*, *moderate*, and *aggressive*. The choice of these parameters are project dependent and influences the number of point clouds generated (Tinkham and Swayze 2021).

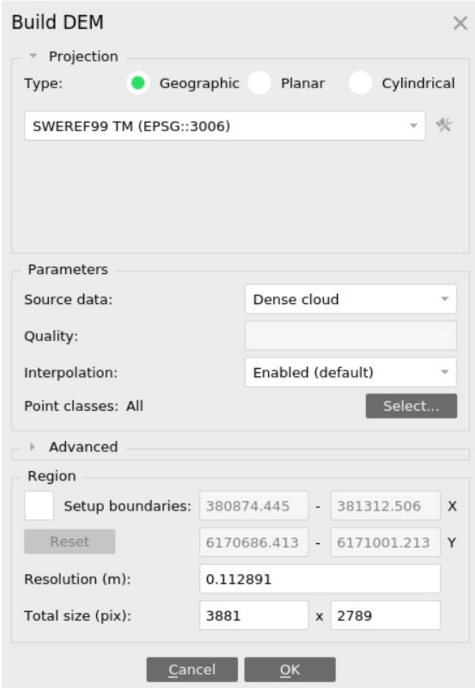


There is an option for setting the image brightness and contrast for a more convenient display. This setting can be achieved by using the *Set Brightness* settings from the *Tools* menu. In addition, the *Reference* tab as well as *Survey statistics* option under the *Tools* menu allows viewing the result and errors. The *Survey statistics* contains information about the camera overlap, camera locations, camera rotations, and ground control points.

Metashape also allows the cleaning of point clouds. However, this process demands high experience and skills required to distinguish between the data and the noise. The noisy points can be deleted by using the different selection tools (*rectangle*, *circle* or *free-form selection*) and can be avoided in further processing steps by adjusting the size of the bounding box region. It is recommended to be very careful while cleaning the point clouds as it might end up deleting the data than the noises.

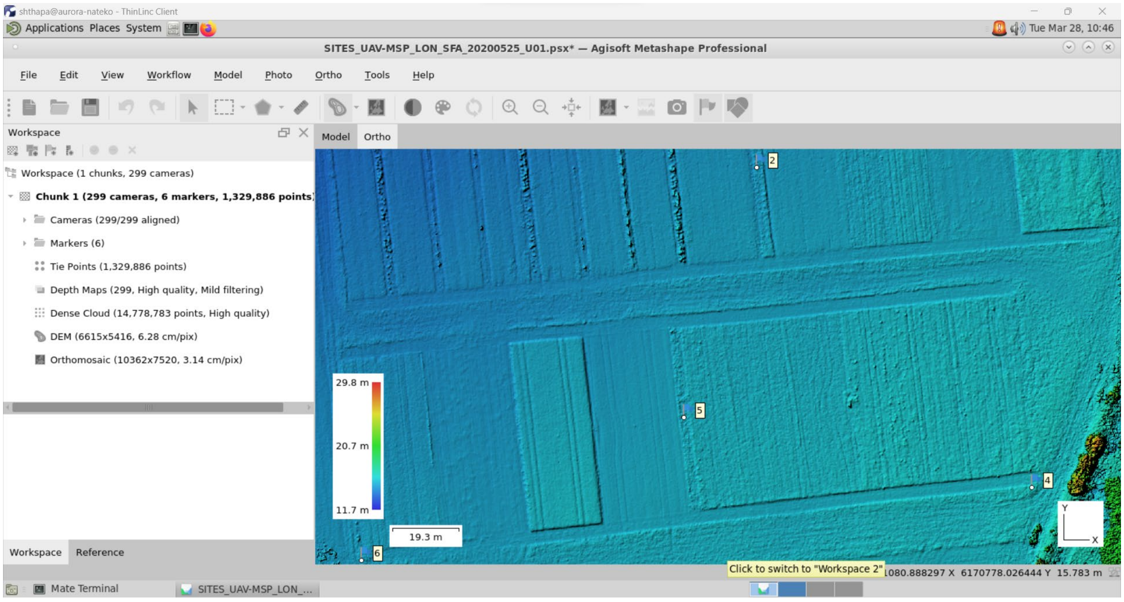
3.1.3.7 Build DEM

After generating the dense point clouds and cleaning the noises if any, the next step is to use these to generate the L3 Digital Elevation Model (DEM). To create DEM, navigate to *Build DEM...* tool available in the *Workflow* menu and define parameters in the dialog box. The projection system is set to Swedish reference system i.e., SWEREF99_TM (EPSG:3006). The source data for the DEM creation is the sparse, dense point clouds or the depth maps. By default, the source data is set to the dense cloud which is what is used in SITES. Under the *Advanced* tab, one can define the X and Y coordinate for *setting boundaries* for the DEM.



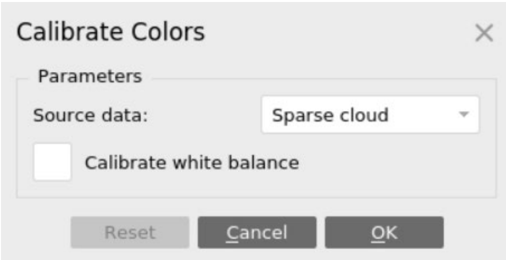
It is also possible to change the resolution of the data however it is recommended to use the default resolution value provided by the system thus achieving the best possible resolution for DEM. Click *OK* and this will popup the processing progress bar dialog box which takes some time depending on the number of images used in the project.

After finishing the processing, it is possible to view the reconstructed DEM surface in the *Ortho view* mode. To display the DEM surface, double-click on the DEM layer instance in the current chunk's contents in the W*orkspace* pane and it will be loaded into the *Ortho view* mode. The DEM surface is draped on top of the hillshade with a color ramp reflecting the variations in the elevations in the area. There is also an option to turn off the hillshade.



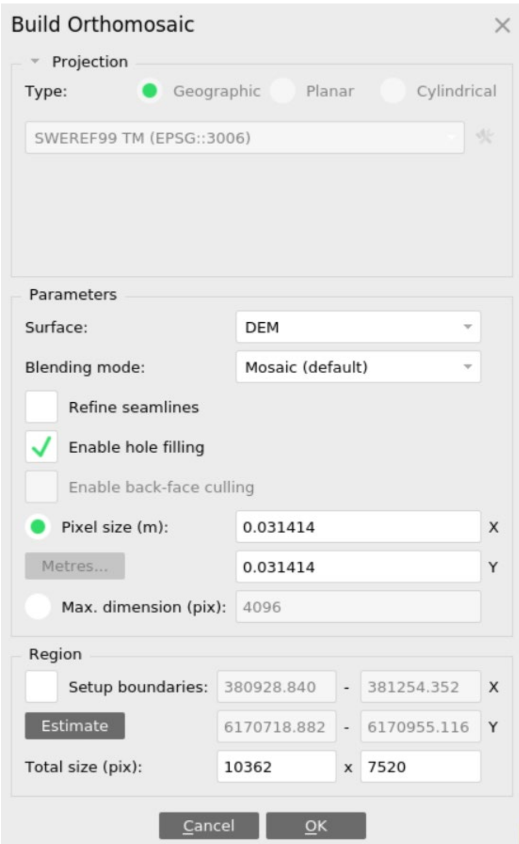
3.1.3.8 Calibrate Colors

The command *Calibrate Colors...* under the *Tools* menu is for vignetting and brightness calibration. This function can be used before the texture/orthomosaic is generated. *White balance calibration* option checked in the *Calibrate Colors...* dialog box will correct each spectral band separately. Conversely, with the option unchecked, all bands will be corrected simultaneously. *Data source* option refers to which surface the overlapping pixels from the different images will be taken into consideration. *Data source* is defined as DEM and the reset button is hit and this avoids applying the vignetting correction twice (remember that the vignetting correction is already applied at L1 data creation stage).



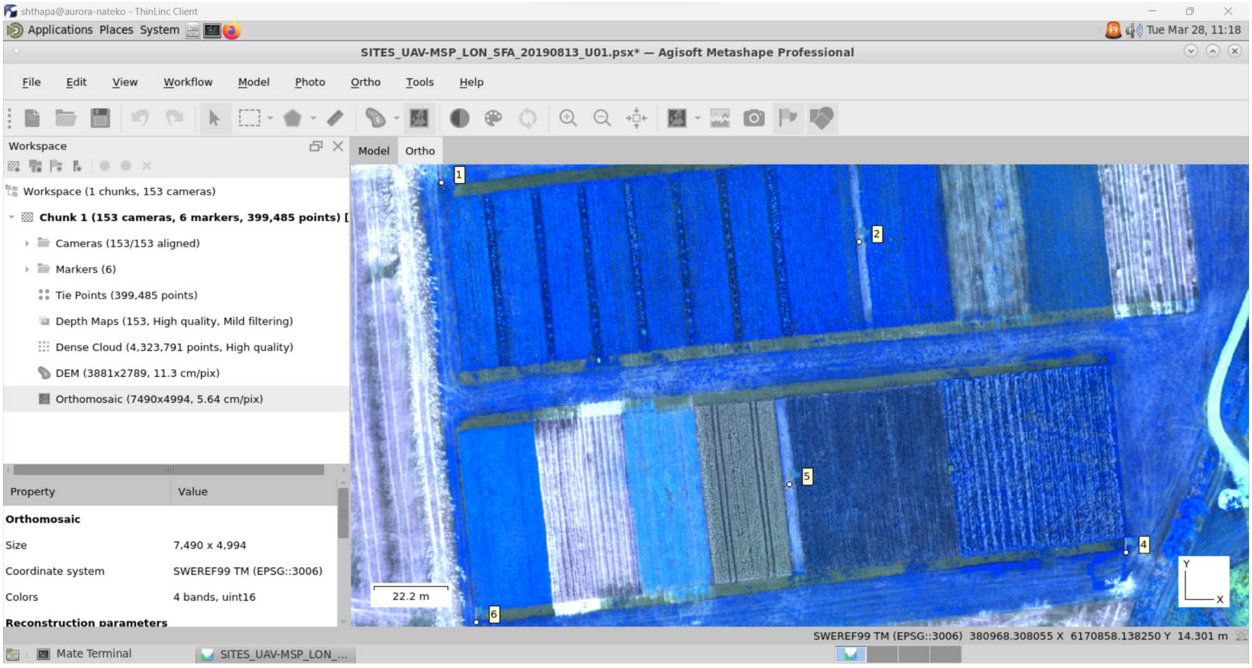
3.1.3.9 Build Orthomosaic

In a similar way, in order to create the orthomosaic, open *Build Orthomosaic...* tool available in the *Workflow* menu and set the parameters as presented in the *Build Orthomosaic...* dialog box.



But before generating the orthomosaic, it is important to import the saturated pixels mask for each images created at L1 stage. This will help to avoid the influence of saturated pixels on the orthomosaic. After that, same coordinate system is defined as of DEM data. The units of measurment correspond to the selected coordinate system. It is recommended to use the same default pixel size values thus achieving the best possible resolution for the orthomosaic.

The *Build Orthomosaic...* processing takes some time depending on the number of images used in the project. After completing the processing, switch back to *Ortho view* mode by double clicking on the orthomosaic layer in the *Workspace* pane to review the generated orthomosaic.

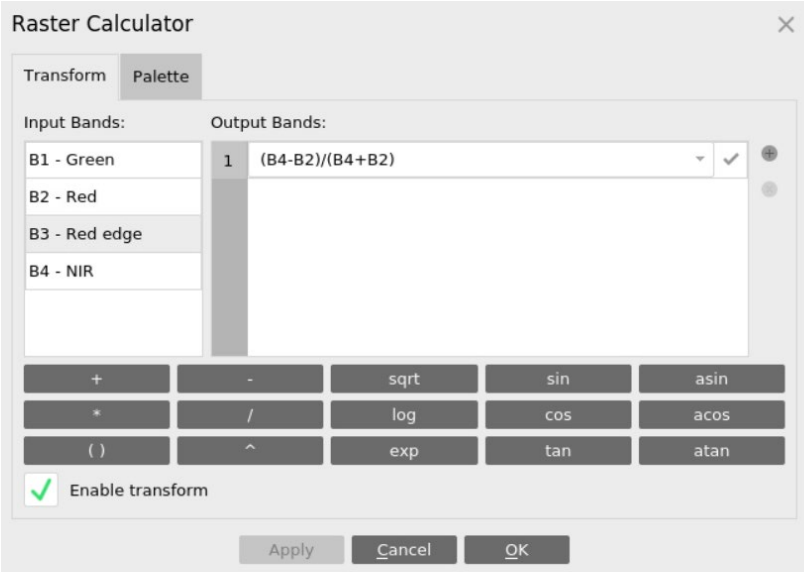


The orthomosaic at this step is radiometrically calibrated based on the sunshine sensor data if the *Sun sensor* option is used for the reflectance calibration at the *Calibrate Reflectance...* step. For SITES data, the orthomosaic generated at this step is not radiometrically calibrated. The images of reflectance calibration targets are used later to calibrate the orthomosaic obtained here. Details about the radiometric calibration is discussed later in the *Radiometric calibration* step.

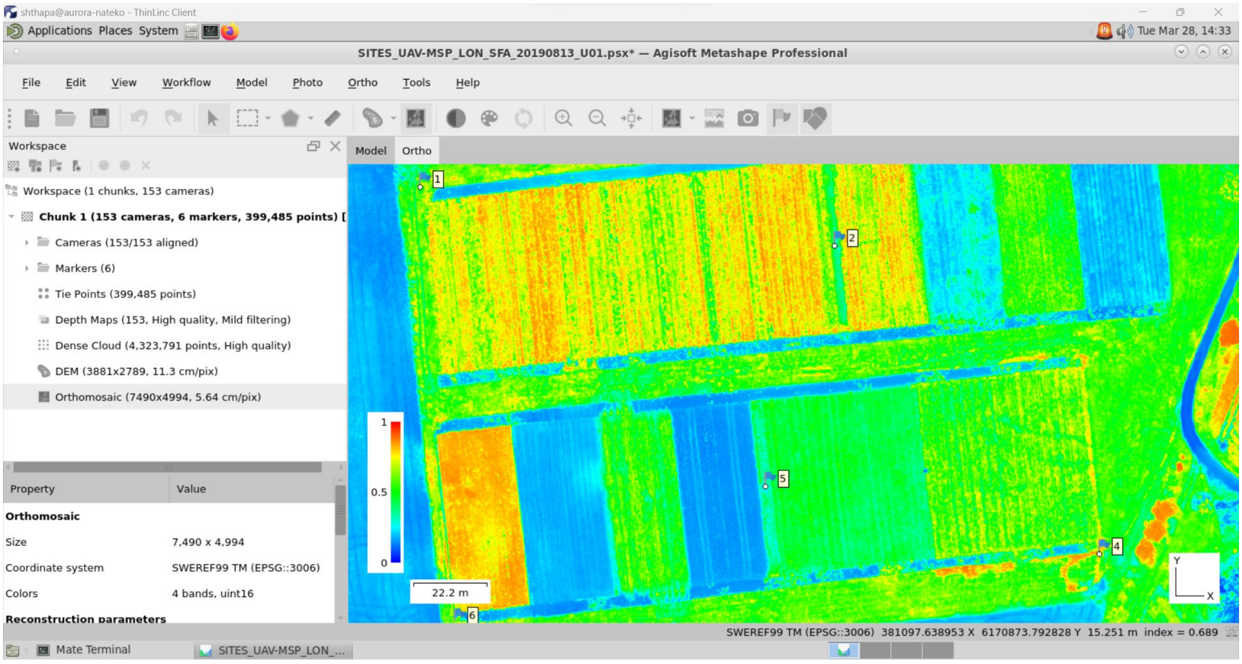
3.1.3.10 Raster Transform

For projects that involves processing of the multispectral images, the spectral bands available in the orthomosaic can be used to compute different spectral vegetation indices such as the commonly used NDVI. The raster transform command can be used by choosing the *Set Raster Transform...* option available under the *Tools* menu. A *Raster Calculator* dialog box pops up on clicking the raster tranform tool.

On the *Transform* tab, one can specify the formula to any spectral index that is of interest and click apply. It is possible to add multiple vegetation indices in the *Output Bands* section to use it for multiple projects.



The formula in the screenshot is for transforming the orthomosaic DN values to NDVI. The color ramp for the NDVI layer can be changed from the *Palette* tab in the top left corner of the *raster calculator* dialog box. Click *apply* to see the changes.



3.1.3.11 Export products

As a final step in the image processing in Metashape, all the photogrammetric products generated are exported to a local drive for preparing to upload to the SITES data portal. The *Export* command under the *File* menu allows to export the L2 data products i.e., the dense point cloud, the orthomosaic, and L3 data product – the DEM data. The geometrically corrected dense point cloud is exported as *.las* dataset and the orthomosaic and DEM as *.tif* format and also allows exporting data in different file formats. *Figure 14* shows one example of orthomosaic (L2) draped on top of the elevation model (L3).

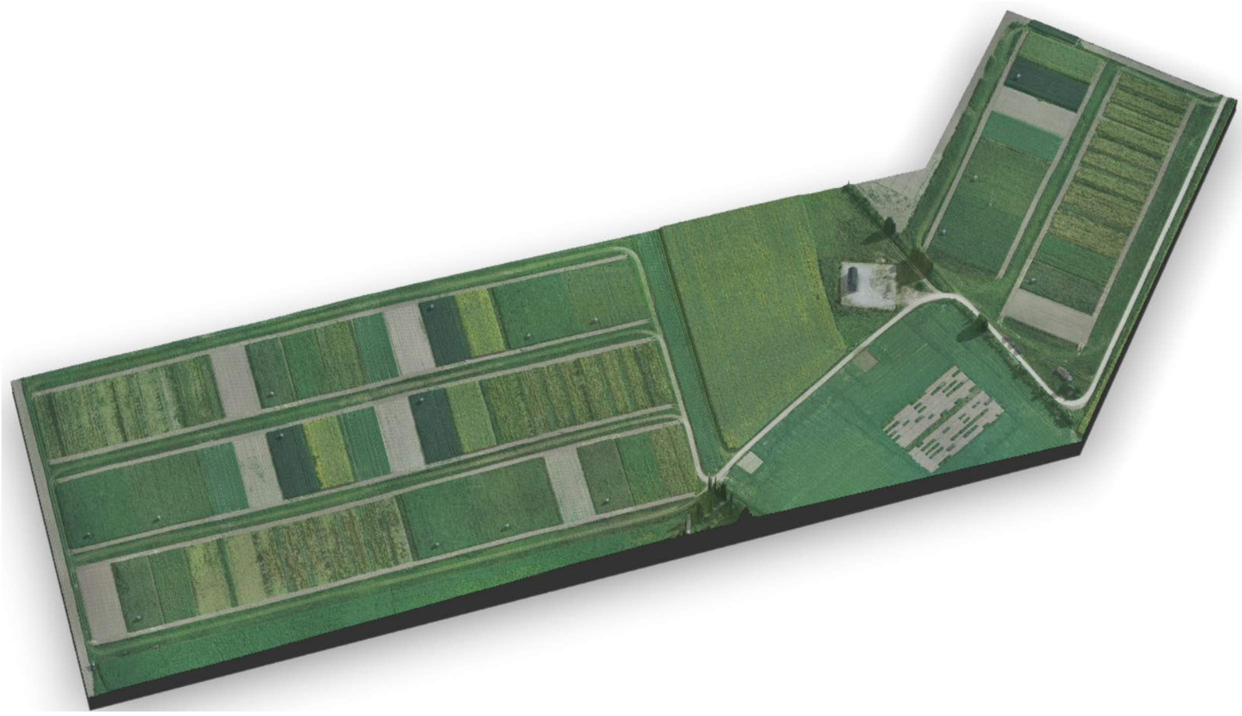


Figure 14. An example showing the UAV L2 data product i.e., the RGB orthomosaic draped on top of the L3 DEM at SITES Lönnstorp station. The data is from 2021.

The filename to these products are defined based on the standards defined within SITES. For instance, the dense point cloud and the orthomosaic are named as:

SITES_UAV-RGB-PTC_SVB_SVB_20220718_L2.las
SITES_UAV-RGB-ORTHO_SVB_SVB_20220718_L2.tif
SITES_UAV-RGB-DEM_SVB_SVB_20220718_L3.tif

In addition to the data products, the data processing report can be generated using *Generate Report...* option in the *Export* command. The processing report contains information about different parameter settings at each steps that is discussed here, the timing of each process, quality of the data and many more. The report is included along with the data products when uploading it to the SITES data portal.

3.1.3.12 Radiometric calibration

For radiometric calibration, the enhanced MSP images of the reflectance panels from L1 stage and the orthomosaic generated at L2 stage are used. The empirical line method (Smith and Milton 1999; Stow et al. 2019) is applied to convert the DN of orthomosaic to surface reflectance according to the *equation (5)*.

$$R = a * DN_{corr} - b, \quad (5)$$

where R is the reflectance, DN_{corr} is the radiometrically corrected DN value after the exposure calibration, vignetting correction, and irradiance normalization while a is the slope of the regression line, and b is the intercept.

The method relies on mean pixel values of three reflectance panels present in the images and their standard reflectance values. The standard reflectance values varies based on the type of panels used. For example (see in *Figure 15* below) black (5%), dark grey (20%) and light grey (50%) for Spectralon panels.

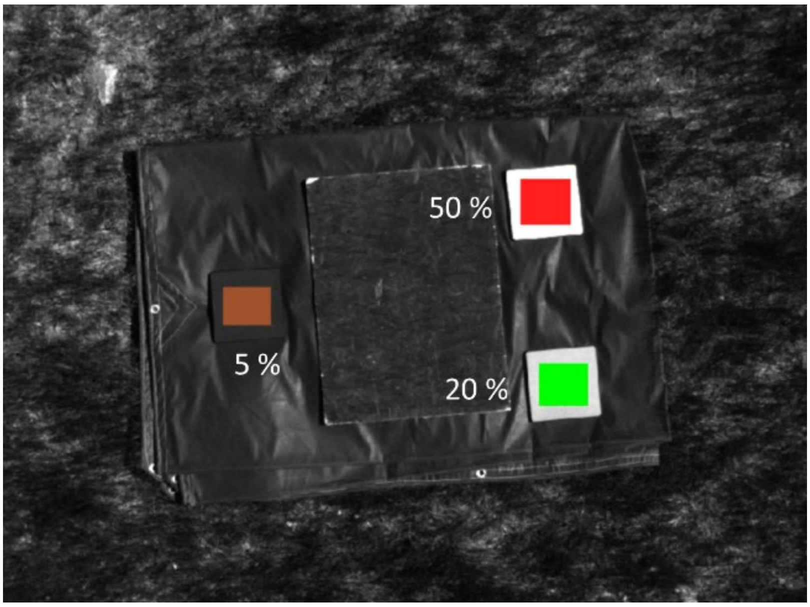


Figure 15. Spectralon reflectance calibration panels with standard reflectance of 5, 20 and 50 % reflectances of size 25 × 25 square cm in images captured for the radiomeric calibration of the orthomosaic.

The equation for the empirical line method is derived using the individual, radiometrically calibrated images (enhanced images at L1 stage) for each band and the standard reflectance values for the calibration panels from the manufacturers. The reflectance panel images for empirical line method should be chosen with care. The images with reflectance panels located centrally are preferred to extract the mean DN values for each band. While extracting the mean DN values, it is ensured that the mean values are entirely inside the reflectance calibration panels and do not include mixed pixels near the edges of the panels that are influenced by the surroundings.

For the spectral bands that has no problem of pixels saturation, all three reflectance panels are used to derive the equation by establishing a linear relationship between the mean DN values and standard reflectance values. But, when there is saturation in pixels of the 50% reflectance calibration panels (which is normally the case), the empirical line is derived based on the mean DN values from two panels only.

Finally, the equation is applied to radiometrically correct the orthomosaic creating a reflectance map which is the basis for creating the NDVI layer (L2).

3.2 UAV RGB Data

The RGB UAV data processing protocol follows almost the similar strategy as MSP UAV data. The only change is that the RGB UAV data are not normalized for the variation in the light conditions throughout the flight. It is rather handled straightforward.

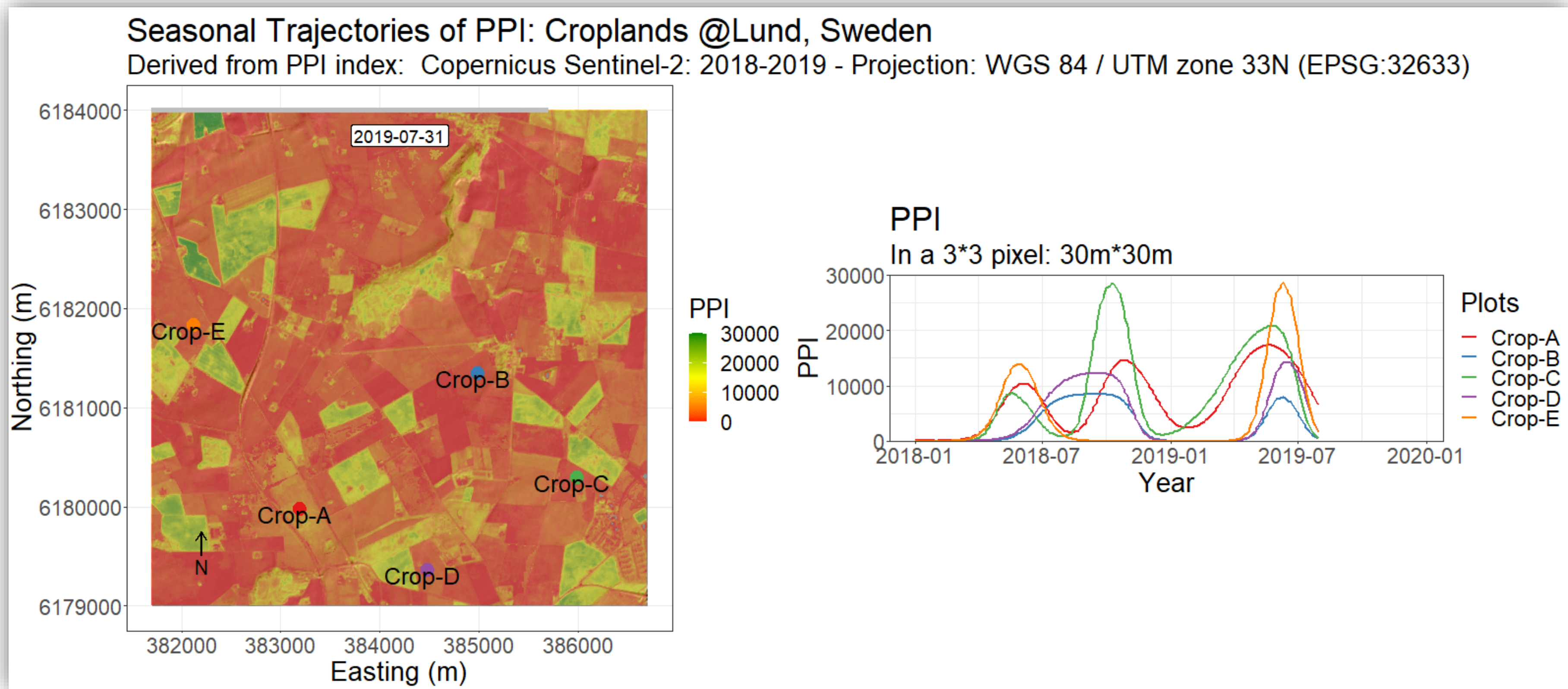
Table 8 lists all the data products defined for the RGB UAV data.

Table 8. List of all RGB UAV data levels and their names in the data portal

Level	Product names
L0	UAV RGB flight datasets
L1	Quality filtered RGB images
L2	RGB Point Clouds RGB Orthomosaic
L3	Digital Surface Model (DSM)

The raw data L0 is the flight datasets which contains RGB images from the flight, GCP (*.txt* or *.csv*) files. The RGB images are available in image *.jpg* file format. The reflectance panel images captured at the start and end of the UAV flights are filtered out and removed. The images are renamed as per the SITES naming convention if needed for data at L1 stage. The low quality images especially the blurry, disorientated, and all the images that are not part of the main mission are removed at this stage. Under L2 data category, there are only two products namely the RGB Point Clouds and RGB Orthomosaic which is the result of the image processing in Metashape as explained in UAV MSP Data section. Finally, the surface model is defined as the L3 data product for RGB UAV data. The file format and naming convention for the UAV RGB data products exactly match with the MSP data.

Both the RGB and MSP orthomosaic (L2) and DEM (L3) data are clipped with a polygon that matches the bounding box of flight area at each station. This is done using QGIS. All the UAV data products are uploaded to the SITES portal as a compressed *.zip* files. When zipping the data, the metadata (*.txt*) file, the processing report is also included. The metadata file contains flight related information such as the date and time of flight, overlap, flight area, wind speed, quality of the data, coordinate system, sensor used, and many more. All these metadata is collected from the UAV pilots at each station through a platform called KOBOToolbox. As soon as the flights are conducted, the flight pilots fill up the survey form available in the toolbox. For the quality flagging of the data, check Chapter 5 where the quality flagging (QFLAG) is explained in more details.


Seasonal Trajectories of PPI: Croplands @Lund, Sweden
Derived from PPI index: Copernicus Sentinel-2: 2018-2019 - Projection: WGS 84 / UTM zone 33N (EPSG:32633)
2019-07-31
Crop-A
Crop-B
Crop-C
Crop-D
Crop-E
N
Northing (m)
6184000
6183000
6182000
6181000
6180000
6179000
Easting (m)
382000
383000
384000
385000
386000
PPI
30000
20000
10000
0
PPI
In a 3*3 pixel: 30m*30m
PPI
30000
20000
10000
0
2018-01
2018-07
2019-01
2019-07
2020-01
Year
Plots
Crop-A
Crop-B
Crop-C
Crop-D
Crop-E

4.1 Satellite

High-Resolution Vegetation Phenology and Productivity (HR-VPP) product (10m * 10m) at high spatial and temporal resolution produced by Copernicus Land Monitoring Service (CLMS) is included in the SITES Spectral thematic program to offer complete set of spectral data from near-surface to satellite sensors orbiting the Earth. These data are available from January 1, 2017 and onwards with a temporal resolution of daily, 10-daily and yearly.

HR-VPP product suite contains 3 different product groups: 1) Raw vegetation indices (VIs), 2) Seasonal Trajectories (STs), and 3) Vegetation Phenology Parameters (VPPs). Out of these 3 product groups, only the seasonal trajectories (STs) and vegetation phenology parameters (VPPs) are used within SITES Spectral.

The seasonal Trajectories (STs) products are derived as a time series of every 10 days on a regular basis by fitting a smoothing and gap filling function to the raw Plant Phenology Index (PPI), generated in the product group VIs (1). The Vegetation Phenology Parameters (VPPs) product bundle are derived from the STs of the PPI index and are provided yearly after the end of the growing season. VPP metrics are provided for up to two growing seasons (s1 and s2). Only the parameters such as start of the season (SOS), end of the season (EOS), length of the season (LOS), amplitute (AMPL), and the seasonal productivity (SPROD) are used within SITES Spectral. For more information about the HR-VPP product (see the documentation).

4.1.1 Satellite Data downloading and Preparation

HRVPP datasets are hosted by WEkEO which is the European commission launched DIAS that offers a single access point to all the Copernicus data and information, alongside the processing resources, tools, and other relevant data. The harmonised data access (HDA) API allows uniform access to the complete WEkEO data catalogue. The API includes data subsetting and batch downloading functionalities.

There are multiple ways the API allows data downloading. As an alternative to using the HDA API directly, SITES use the Python client library ***hda***, which abstracts away the details of the API. In order to use this library, it needs to be first installed in the environment. The step by step process to install and use this Python library is:

a) Open Terminal and install the library by typing: *pip install hda*

b) In the next step, a configuration file needs to be created. The file content should be the data-broker url, and the WEkEO credentials. Users must have already registered for an account at WEkEO platform at this stage.

c) Create a file in the root location of the current user and fill up the file with the following information as mentioned before:

   url: https://wekeo-broker.apps.mercator.dpi.wekeo.eu/databroker
   user: sites (WEkEO username)
   password: ********** (WEkEO password)

d) Save the file as *.hdarc* name and move it to /home/sh3208th

The details of the installation procedure are explained here (https://www.wekeo.eu/docs/hda-python-lib) along with an example script to query and download data. All the available data products can be viewed here https://www.wekeo.eu/data?view=viewer.

A tailor-made Python script for downloading satellite data is embedded into a bash script which can be run in a terminal. In the script, the approximate bounding box information and the satellite tile ID information is predefined for each of the SITES station. In addition, the script creates and uses the copy of a ***.json*** file available within the same file path as the script which contains the data descriptor. The ***.json*** file keeps on updating with each unique data product being downloaded. The script once run asks from user to choose the SITES station and define which year of data to download. All the downloaded data are arranged as a separate folder defined after the product name within a parent folder named after the chosen SITES station.

A second set of Python script is embedded into a separate bash script which prepares the downloaded satellite data. The satellite data preparation task includes two raster operations: 1) reprojection to SWEREF99 projection system which is the standard projection system used within SITES and 2) clipping the reprojected layer to match spatial extent defined for each SITES station (*Table 9*). The end result of the script is the clipped satellite data in SWEREF99 projection system including the quality flag (QFLAG) for the chosen station.

Table 9. Information on satellite tile ID, spatial extent (in SWEREF99_TM) for each SITES station used in order to clip the HRVPP data product and their ground coverage in square kilometers.

Station	Tile ID	Extent (xmin ymin xmax ymax in meters)	Coverage (sq.km)
Abisko	33WXR	639000 7564500 674000 7599500	35*35
Asa	33VVD	478000 6324000 498000 6344000	20*20
Bolmen	33VVD	401000 6291000 441000 6331000	40*40
Erken	34VCM	690000 6630000 710000 6650000	20*20
Grimsö	33VWG	514000 6610000 534000 6630000	20*20
Lönnstorp	33UUB	375000 6162000 395000 6182000	20*20
Röbäcksdalen	34VDR	746500 7074500 766500 7094500	20*20
Skogaryd	33VUE	323000 6463000 343000 6483000	20*20
Svartberget	34WDS	716500 7118000 736500 7138000	20*20
Tarfala	33WXR	650000 7527000 680000 7557000	30*30

There is also a new python package (*PyVPP*) developed for accessing and downloading the HRVPP data products. PyVPP package is especially developed to support phenological data downloading from WEkEO. More information about this package can be found here: https://pypi.org/project/pyvpp/.

SITES spectral uses these satellite derived HRVPP products to give researchers a broad spatial view of vegetation conditions at SITES stations. This work generates data layers describing vegetation productivity and phenology for geographical extent mentioned in *Table 9*. Layer data ready for inclusion in GIS databases will be distributed in analysis-ready format, in the Swedish reference system. The data show the seasonal development of green vegetation and can be useful for monitoring vegetation’s response to variations in weather, human influence, and other factors.

An example of how a satellite derived VPP product i.e., the total productivity shows the trends in vegetation productivity in the agriculutral areas over the years (*Figure 16*).

Figure 16. An example of an agricultural area at SITES Lönnstorp station, showing annual variations in vegetation productivity between 2017 and 2020. Note the strong decrease in productivity in 2018 in response to the drought.

Chapter 5. Data Quality Flagging (QFLAG)

The data quality of all SITES Spectral data are provided along with all the collected data. A three-digit numerical value is provided, which defines the quality flag (QFLAG) value for the data levels derived for all the measurement types that are part of SITES Spectral. Each digit represents the quality of one of the three parameters considered for defining whether data are of low, medium, or high quality. These parameters are different for different spectral sensors and are defined based on the understanding of how each of them can affect the quality of the data. There are separate sections, each explaining the parameters and quality flagging system used within SITES Spectral for each of the data types.

5.1 QFLAG system for PhenoCam Imagery

Out of the four levels of images recorded with PhenoCams: L0, L1, L2, and L3, the QFLAG values are available for the L3 data product only, which includes time series of the daily average of RGB-based vegetation indices (VIs), i.e., Green Chromatic Coordinate (GCC) and Red Chromatic Coordinate (RCC). For each daily averaged VI value, there is a separate column in the data file explaining the quality of that data value. Three parameters are considered to assign a QFLAG value for each daily mean VI value:

1) Presence/absence of snow
2) Number of images used in computing daily average
3) Solar elevation angle

The time series of GCC and RCC are extracted for specific regions of interest (ROIs) within an image using the method as explained in Thapa et al. (2021). The number of ROIs can vary from station to station. So, when computing GCC and RCC for a given ROI, the presence or absence of snow is considered, coded as quality value 100 (snow) or 200 (snow-free). If the chosen ROI for a PhenoCam image contains snow patches, it will be coded as 100 (ROI 3 in *Figure 17*), otherwise as 200, if the ROI is snow-free (ROI 1 and ROI 2 in *Figure 17*).

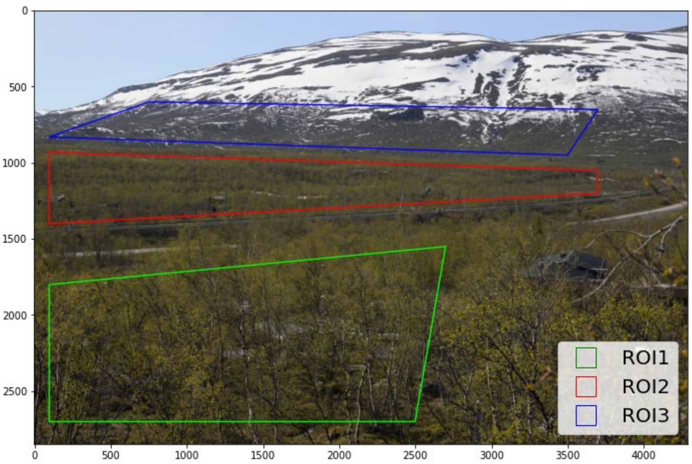


Figure 17. Regions of interest (ROI), i.e., ROI1 (Green), ROI2 (Red), and ROI3 (Blue), used for extracting time series of vegetation indices (VIs) at SITES Abisko Scientific Research Station. Each of these ROI based VIs are quality coded based on the presence or absence of snow within the ROI. Here, ROI3 is coded 100 i.e., presence of snow, while ROI1 and ROI2 are coded 200 i.e., absence of snow.

To compute daily averages, the number of images between 10:00 – 14:00 is considered, after removing poor images (too bright or dark, blurry, foggy, solar glared, disoriented, etc.). Here, if less than 1/3 (i.e., < 2) of total images are used to compute the daily average, it is considered low quality and coded as 10. If between 1/3 – 2/3 (i.e., 2 – 4 images) are used to compute the daily average, it is considered medium quality and coded as 20. Lastly, if the daily average is the result of more than 2/3 of the images (i.e., > 4), such daily averages are considered as high-quality data and given a code value of 30.

Finally, for solar elevation angle, the daily images are classified into three classes based on sun elevation angles: Class 1 (0° - 20°), Class 2 (20° - 30°), and Class 3 (>30°). This scheme was adopted from the webcam network and image database for studying phenological changes in vegetation in Finland (Peltoniemi et al. 2018). The above class categories are coded as 1, 2, and 3 respectively.

A final 3-digit QFLAG value is obtained by adding up the quality code values for each of the three parameters (*Table 10*). For a day when no images exist, the QFLAG is defined as not a number (NaN). In the table, the lowest quality value is 100 while 233 is the highest possible data quality that can be expected in the L3 PhenoCam data.

Table 10: QFLAG values for PhenoCam L3 data. All 3 parameters (snow, image count, solar elevation) are first coded with quality values referring to the low, medium, and high quality. The individual quality values are then summed, and the result is the QFLAG value.

Parameters			
Snow	**Image Count**	**Solar Elevation Angle**	**QFLAG**
-	No Image	-	NaN
100	-	-	100
200	10	1	211
200		2	212
200		3	213
200	20	1	221
200		2	222
200		3	223
200	30	1	231
200		2	232
200		3	233

An example of a L3 PhenoCam time series with QFLAG information is shown in the plot below (*Figure 18*). For the given time series, the data have 8 out of the 10 possible QFLAG values. Each colored symbol represents a unique QFLAG value in the data. The plot shows the importance of having QFLAG information in the data. The spline fit in the first plot is the result of using all the time series data without considering the QFLAG values while the second plot shows the fit when using only high-quality data i.e., 231, 232, and 233. The benefit of having QFLAG information in the data will allow the use of selective QFLAG values for fitting the data.

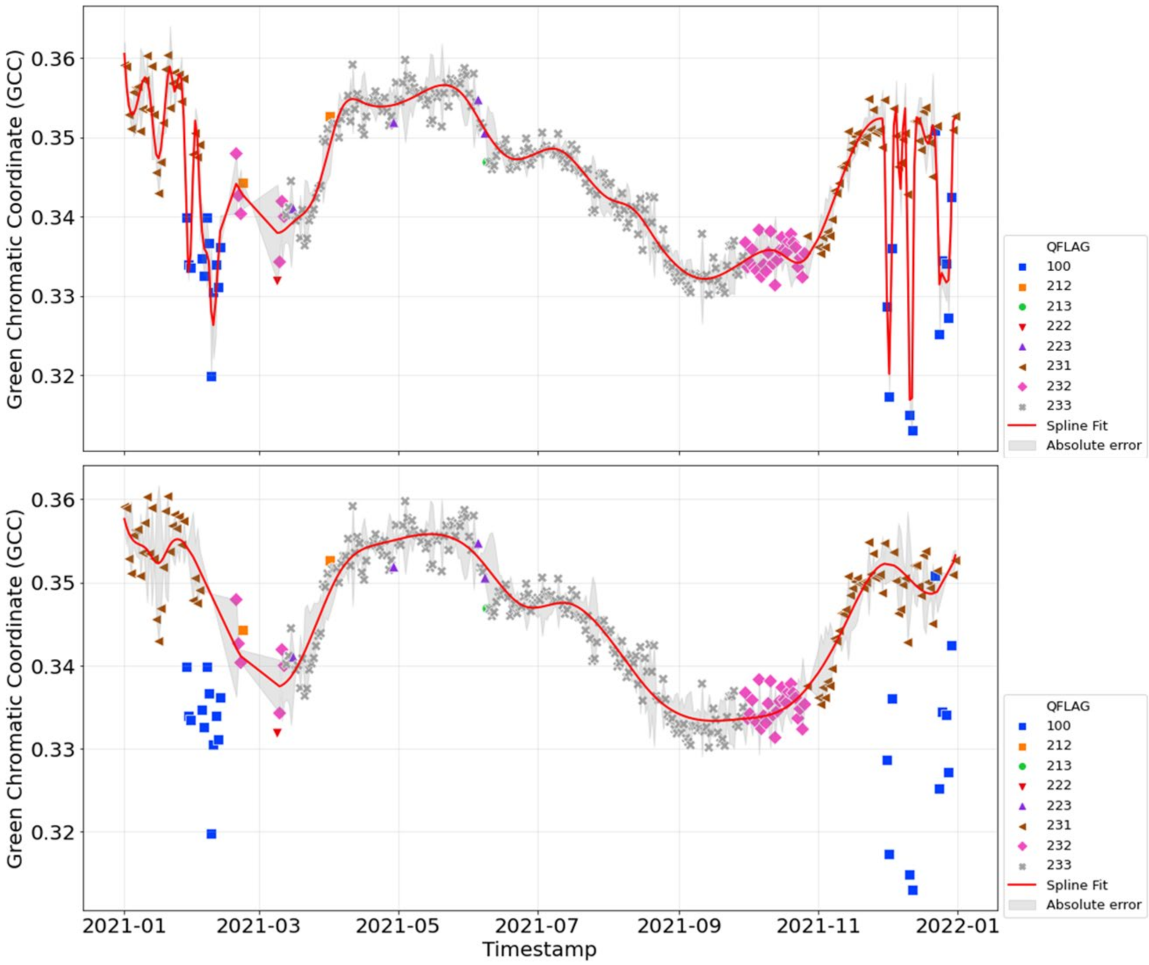


Figure 18: L3 PhenoCam time series of daily averaged Green Chromatic Coordinate (GCC) from SITES Lönnstorp Research Station along with the QFLAG values. Each colored symbol represents a unique QFLAG value in the data. QFLAG value 100 (presence of snow within the ROI) is the lowest quality data while 233 is the highest quality data. The time-series data is fitted with a spline fit (red line). The upper plot shows the fit using the complete time series without considering the QFLAG value, while the bottom plot shows the fit using only the high-quality flagged data i.e., 231, 232, and 233. The grey fill in the background represents the absolute error of the fit.

5.2 QFLAG system for Spectral Reflectance Sensors Data

Out of the three data levels for spectral reflectance sensors: L0, L1, and L2, the QFLAG values are available for the L2 Spectral Reflectance Sensor Data products only, which is the time series of the daily average of vegetation indices (VIs) i.e., Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI). For each daily averaged VI value, there is a separate column in the data file that states the quality of that data. Three parameters are considered to assign a QFLAG value for each daily mean VI value:

1) Data Frequency
2) Time elapsed since sensor calibration
3) Sensor degradation

Data frequency considers the number of valid measurements available for a daily average. The multispectral sensors are programmed to provide average raw incoming and reflected radiation every 10 minutes. However, only the data collected between 10:00 - 14:00 (as is the case for PhenoCam) is used for the computation of reflectance. That means that the maximum number of records within the time window is 30. First, the reflectance is computed for all the data collected within the time window. Only the reflectance values between 0 and +1 are valid measurements. Based on the number of valid reflectance data, the computed daily mean reflectance is coded with quality values as 100, 200, and 300. Daily mean reflectance computed using less than one-third of the maximum possible data points in a day (i.e., < 10) is defined as low quality and coded as 100. Medium quality code value 200 is assigned to the daily mean reflectance which is calculated using 1/3 – 2/3 (i.e., between 10 – 20) of data. Lastly, if the daily average is the result of more than 2/3 (i.e., > 20) of the possible measurements, the daily mean reflectance falls into the high-quality category and hence assigned 300 as the quality code value.

The spectral sensors are mounted on masts and measure and record the incoming and reflected radiation continuously throughout the year for most stations. During a year, there is a high risk that the sensor measurements are influenced by various factors. Figure 3 depicts a few of these factors that may be encountered.

Figure 19: Possible sensor deterioration caused by various factors: A) covered by insects, B) dirt and drops of blood likely from birds killing their prey on the sensor.

Such possible sensor problems and long-term deterioration create uncertain measurements and thereby demands the need for periodic maintenance and sensor calibration. With this in mind, spectral sensors at SITES are calibrated each year following the calibration procedure discussed in Jin and Eklundh (2015). However, it is sometimes not possible to calibrate the sensors frequently at the station as they are located high on the mast. To dismount and calibrate them in good weather conditions, and finally mount them back might be challenging to do on an annual basis. There is also a risk of creating a data gap in the time series during the productive part of the season. So, the time from the latest sensor calibration is defined as one of the data quality factors. The derived daily averaged VI values from the raw incoming and reflected radiation gets the quality flag based on the elapsed time from the latest sensor calibration. If the data values were collected using sensors calibrated less than one year ago, following the standard calibration protocol discussed in Jin and Eklundh (2015), a high-quality flag value of 30 is assigned to the processed time series data. A medium quality flag value of 20 is assigned when the recent sensor calibration was carried out 1-2 years ago. A low-quality flag value 10 is assigned to those data where the last sensor calibration was carried out more than 2 years ago.

Finally, the sensor degradation is calculated after each time the sensors are calibrated. Sensor degradation is calculated by subtracting the latest two calibration coefficients for each channel available. If for one of the available channels: Red (R), Green (G), Blue (B), Red-edge (REG), Near-Infrared (NIR), Shortwave Infrared (SWIR), the degradation percentage is higher than 30%, the data is classified as low quality and is assigned a quality flag value of 1. For degradation less than 30%, the data is classified as high quality and is assigned a quality flag value of 2. The information about the sensor degradation is available in the calibration certificate issued after each calibration.

Finally, the quality code values for the three parameters are summed to generate a 3-digit QFLAG value (*Table 11*). For days when no data exist due to e.g., power failure, the QFLAG is defined as not a number (NaN). In the table, the lowest quality is 111 while 332 is the highest possible data quality for the L2 daily average VI data.

Table 11: QFLAG values for Spectral Reflectance Sensors data. All 3 parameters (data frequency, calibration year, degradation) are first coded with quality flag values referring to the low, medium, and high quality, except for the sensor degradation which is only low or high quality. These individual quality values are then added, and the result is the QFLAG value.

Parameters			
Data Frequency	**Calibration Year**	**Degradation (%)**	**QFLAG**
-	-	-	NaN
100	10	1	111
		2	112
	20	1	121
		2	122
	30	1	131
		2	132
200	10	1	211
		2	212
	20	1	221
		2	222
	30	1	231
		2	232
300	10	1	311
		2	312
	20	1	321
		2	322
	30	1	331
		2	332

An example of L2 spectral reflectance sensor time series data with QFLAG information is shown below (*Figure 20*). For the given time series, the data have 3 out of the 18 possible QFLAG values. Each colored symbol represents a unique QFLAG value in the data. The spline fit in the first plot is the result of using all the time series data without considering the QFLAG values, while the second plot shows a smooth fit due to the removal of low-quality data i.e., 132 in this case. The benefit of having QFLAG information in the data allows the use of selective QFLAG values for fitting the data.

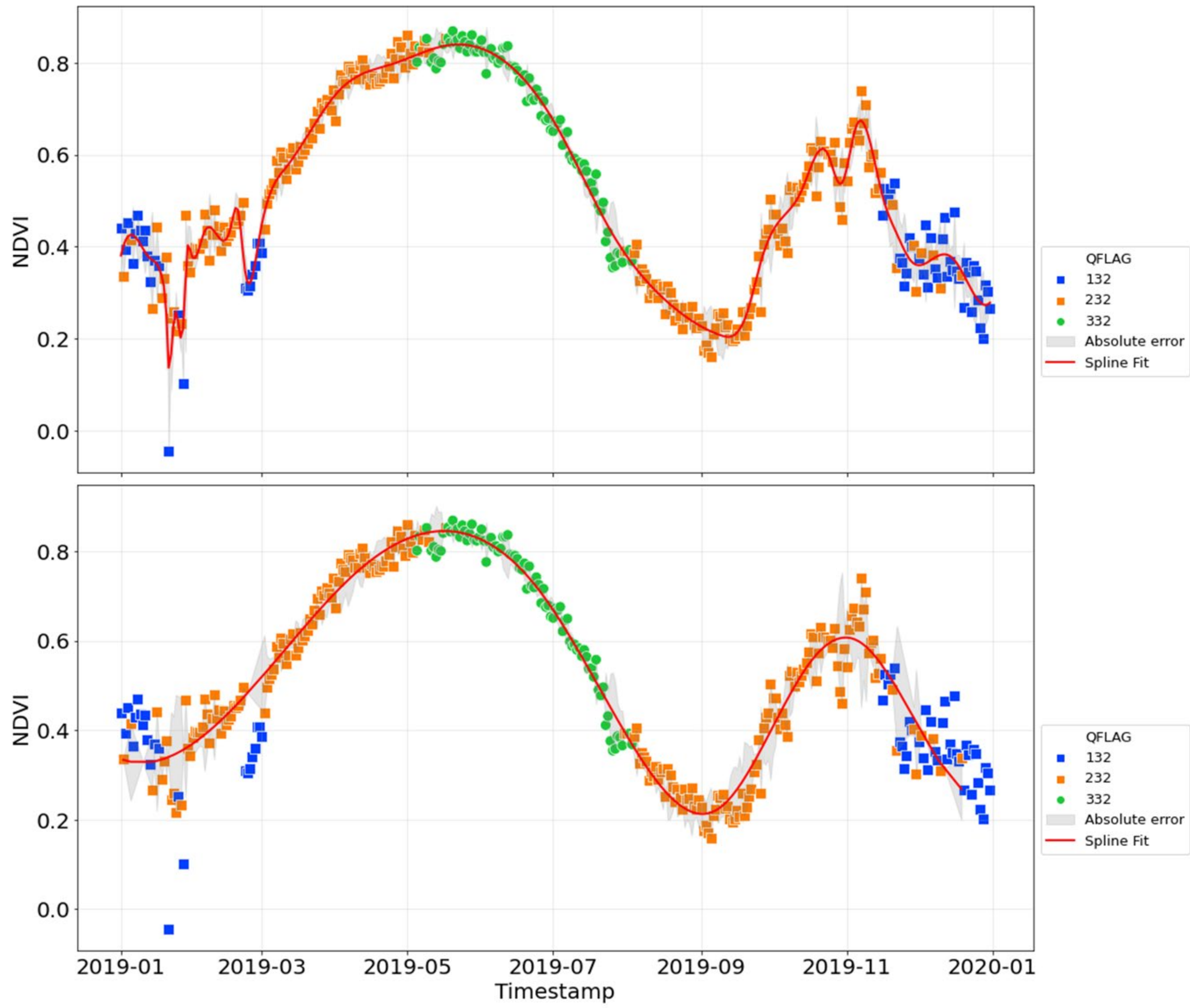


Figure 20: L2 spectral reflectance sensor time series of daily averaged Normalized Difference Vegetation Index (NDVI) data from SAFE-A, SITES Lönnstorp Research Station along with the QFLAG values. Each colored symbol represents a unique QFLAG value in the data. QFLAG value 132 (refers to low data frequency while having high quality for the other two parameters) is the lowest quality data while 332 is the highest quality data. The time-series data is fitted with a spline fit (red line). The first plot shows the fit using complete time series data without considering the QFLAG while the second plot shows the fit removing the low-quality flagged data i.e., 132. The grey fill in the background represents the absolute error of the fit.

5.3 QFLAG system for UAV Data

The QFLAG definition for the unmanned aerial vehicle (UAV) image data is slightly different from the other two sensor types. The UAV data quality flagging is done differently for the type of sensor used: RGB (red, green, and blue) or multispectral (MSP) sensors, and explained in separate sections.

5.3.1 Multispectral UAV Data

For defining the quality of the MSP UAV data, three parameters are considered:

1) Radiometric accuracy
2) Geometric accuracy
3) Visual artifacts

The different products created with UAV data are typically orthomosaics with surface reflectance or vegetation indices, such as the commonly used NDVI. For time series of orthomosaics, to enable comparisons between reflectance and NDVI collected with different sensors and at different times, e.g., for vegetation phenological studies, it is crucial that the data are radiometrically consistent and the surface reflectance estimates from such data are reliable (Olsson et al. 2021; Thapa et al. 2021). So, radiometric accuracy is considered as one of the important parameters.

Firstly, if the MSP flight data matches with one of the following three cases as listed below, it is defined as low-quality data from a radiometric perspective and such data are assigned a quality code value of 100:

Case 1: Irradiance data from the sunshine sensor onboard are missing and no reflectance panel images were captured during the flight

Case 2: Irradiance data from the sunshine sensor onboard are missing but reflectance panel images were captured both at the ascend or descend of the flight

Case 3: Irradiance data from the sunshine sensor onboard are available but have very complex variation throughout the flight

The irradiance data from the sunshine sensor onboard the UAV helps in understanding the stability of incoming light conditions during the flight. For radiometric calibration of MSP UAV data, the images of reflectance calibration panels (nominal reflectance: 5%, 20%, and 50%), captured when ascending and descending vertically over the calibration panels to reduce the effect of viewing angles, are used. If both irradiance data as well as the reflectance panel images are missing (Case 1), the radiometric quality of the data is uncertain and hence such data are categorized as low-quality data.

For the MSP flight that matches with Case 2, only the stability of incoming light while capturing the reflectance panels is known. However, the light conditions during the UAV flight are still uncertain. The mean digital number (DN) values of the reflectance panels that are extracted for radiometric calibration are checked. If there is a high variation in the mean DN values of the reflectance panel images captured both at the ascend or descend of the flight, such a flight is classified as low-quality. The high variation can be linked to different light conditions when the panel images are captured. It could also be the case of shadow falling on the reflectance panel images or the influence of the base material used for the reflectance panel (*Figure 21*).

Figure 21: Example of low-quality reflectance panel images captured during the flight. The first (A) and second (B) image have shadow on the reflectance panels, while the third image (C) shows that the white base material used for holding the panels seems to influence the panel resulting in saturation of the pixel values.

Lastly, if the irradiance data from the sunshine sensor onboard the UAV for a flight represents a complex variation during the flight, the variation is not handled by the irradiance normalization technique as explained in Olsson et al. (2021) and Thapa et al. (2021), the flight is considered to have low quality without considering the presence or absence of the reflectance calibration panel images. An example of such a case is depicted in *Figure 22*.

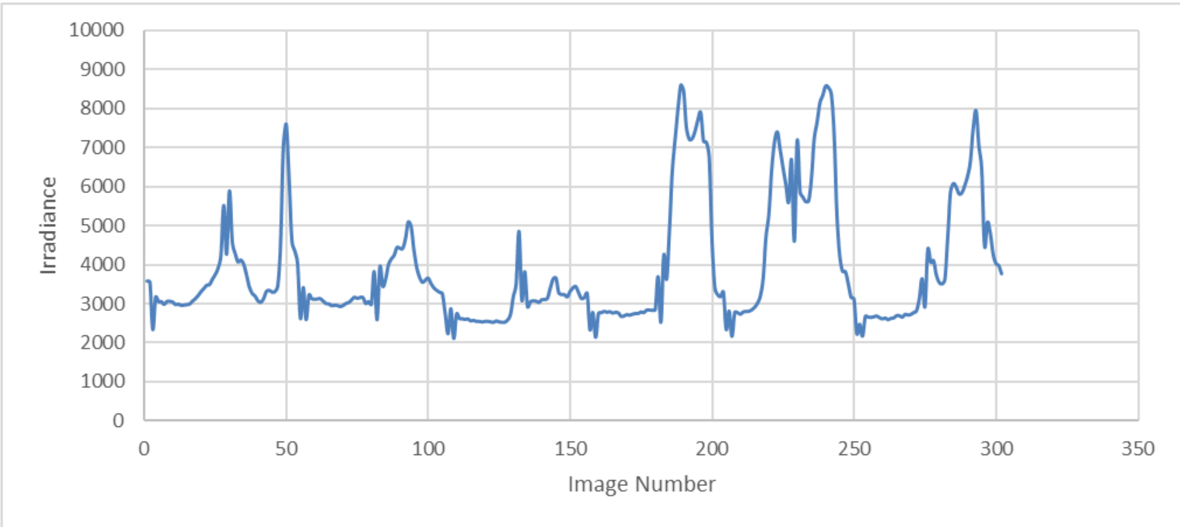


Figure 22: Irradiance data from the sunshine sensor onboard the UAV for a MSP UAV flight conducted at SAFE-A, SITES Lönnstorp Research Station on June 3, 2020. The irradiance data showed the complex light conditions during the flight.

The complex variation in irradiance condition may be reflected in the orthomosaic and NDVI layer. An example of such an orthomosaic and NDVI layer created with data having a complex irradiance pattern throughout the flight is depicted in *Figure 23*. The false-color composite orthomosaic clearly shows dark and bright regions with stronger variation in the lower part (south) of the orthomosaic. And this affects the quality of the NDVI layer (*right, Figure 23*).

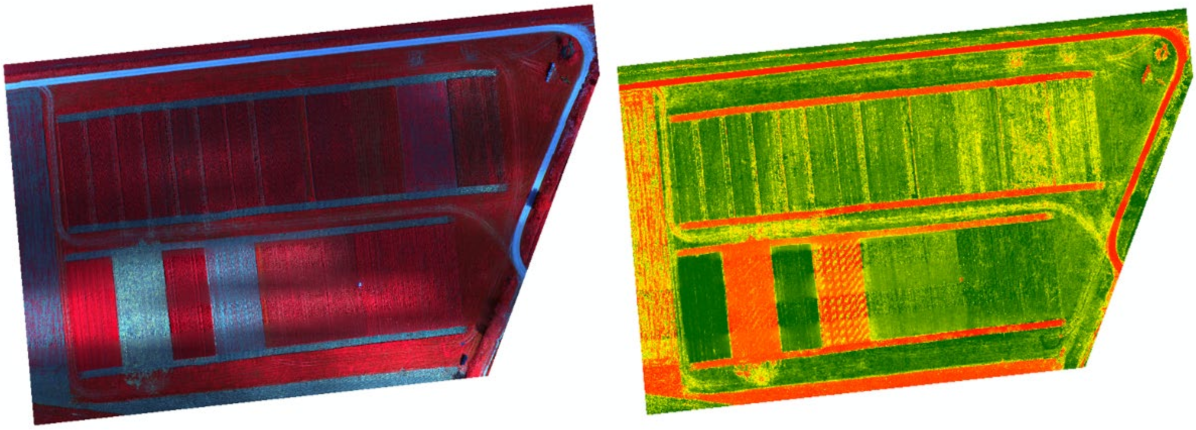


Figure 23: Example of how the complex variation of irradiance pattern can impact the quality of the level 2 data derived orthomosaic (left) and NDVI layer (right). The false-color composite orthomosaic clearly shows dark and bright regions with strong variation in the lower part (south) of the orthomosaic. The same is the case for the NDVI layer.

Secondly, if the MSP flight data match with one of the following four cases as listed below, the flight data are defined as medium-quality from a radiometric perspective and such data are assigned the quality code value of 200:

Case 4: Irradiance data from the sunshine sensor onboard are missing but reflectance panel images were captured both at the ascend or descend of the flight

Case 5: Irradiance data from the sunshine sensor are available but reflectance panel images are missing

Case 6: Irradiance data from the sunshine sensor are available but only reflectance panel images captured either at the ascend or descend are available

Case 7: Significant variation in irradiance data from sunshine sensor for flights conducted with low solar elevation angles

The mean DN values of the reflectance panel images that are extracted for the radiometric calibration of MSP UAV orthomosaics could explain how consistent the light conditions were right at the start and end of the flight. If the extracted mean DN values are consistent in the reflectance panel images, despite the uncertainty of irradiance condition during the flight (Case 4), such flight data are classified under this medium-quality category.

For flight data matching Case 5, the irradiance data from the sunshine sensor onboard UAV are checked. If the irradiance data during a flight appears to be very consistent throughout the flight, it is considered to have medium-quality even though the radiometric calibration cannot be performed due to the missing reflectance panel images.

The quality of the radiometric calibration depends on the number of reflectance panel images. If only a single reflectance panel image or a significant number of images were captured either at the start or the end of the flight irrespective of the irradiance data availability for flight data (Case 6), the mean DN values from the available reflectance panel images can still be used to radiometrically calibrate the orthomosaics, but with lower accuracy and hence considered as medium quality.

Flights conducted on a sunny day with low solar elevation angle (Case 7), especially during the winter, will have significant variation in the irradiance data. Despite the variation, the irradiance pattern (blue line) can still appear stable with a clear regular pattern as shown in *Figure 24*.

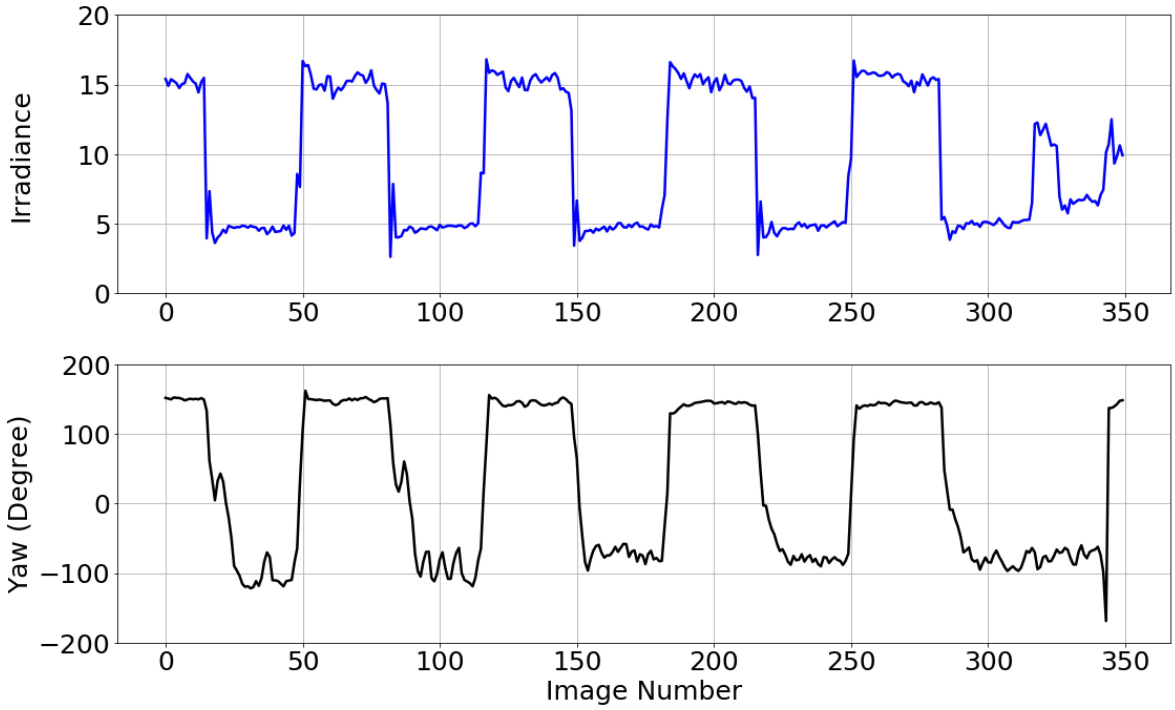


Figure 24. Stable sunshine sensor irradiance data (blue line) for multispectral UAV flight conducted at SAFE-B, SITES Lönnstorp Research Station on November 6, 2018. The plot shows the significant variation with each turn of the UAV as shown by the yaw angle (black line) extracted from the image metadata.

Here, in Figure 8, the pattern with increasing and decreasing irradiance data is not actually due to the change in incoming light but rather to the change in the yaw angle (i.e., the change in the flight direction of the UAV). So, for such a flight, with sun low on the horizon, significant variation in the irradiance data (increasing and decreasing when the UAV is moving towards or away from the Sun respectively), there is a high risk of Bidirectional Reflectance Distribution Function (BRDF) effect on the data as the reflectance properties of vegetation vary with view angle and sun position. Despite having the reflectance panel images captured both at the start and end of the flight, irradiance data from the sunshine sensor for the entire flight, these types of flight data are defined as medium quality.

Finally, multispectral UAV data are classified as high-quality data from a radiometric perspective with quality code value of 300 if they match with one of the following two cases:

Case 8: Consistent irradiance data from the sunshine sensor during the flight with a significant number of reflectance calibration panel images captured both at the ascend and descend of the flight

Case 9: Slight variation in irradiance data from the sunshine sensor during the flight that can be handled with irradiance normalization technique

As mentioned earlier, the number of reflectance calibration panel images impacts the quality of the radiometric calibration. If a MSP flight data matches with the characteristics of Case 8, the flight results in a high-quality flag value. The significant number of reflectance calibration means, there should be at least three images at the start and three images at the end of the flight. The consistent irradiance pattern can be expected for a flight carried out either under complete cloud cover (A) or in completely sunny condition (B), an example of such is presented in *Figure 25*.

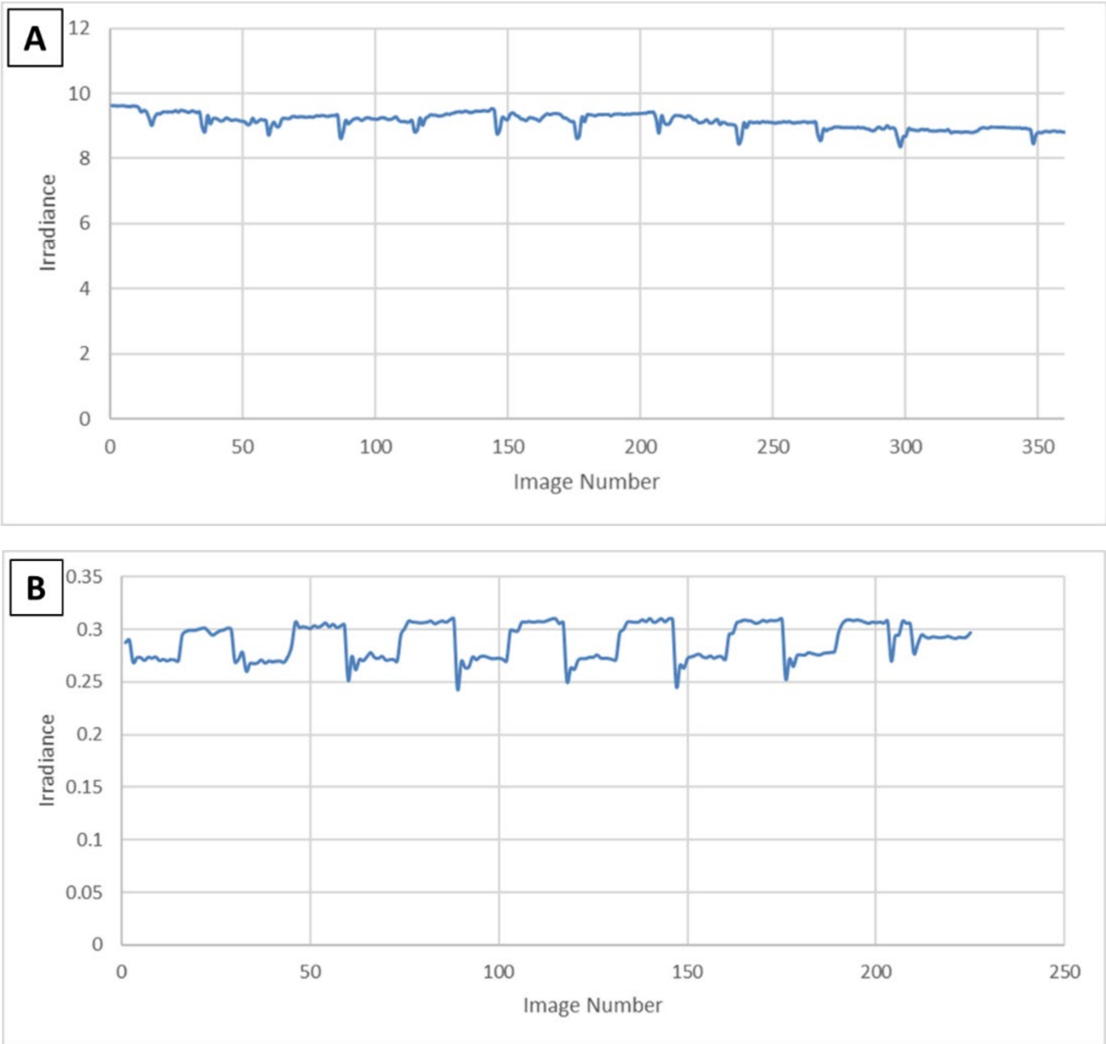


Figure 25: Consistent sunshine sensor irradiance data for multispectral UAV flight conducted under complete cloud cover (A) and in sunny conditions (B). Flight (A) was conducted at SITES Svartberget Research Station on July 10, 2019, while (B) was conducted at SITES Röbäcksdalen Field Research Station on August 28, 2020. Again, the flight pattern (B) refers to each turn of the UAV during the flight. However, such pattern is insignificant on the flight conducted on cloudy condition (A).

If the flight is conducted in a mixed weather condition (partly cloudy) or if there is a gradual change in weather conditions during the flight, there can be a slight variation in the irradiance pattern and if the flight also includes acquisition of significant number of reflectance panel images as mentioned before. Such gradual variation in the irradiance can be handled by the irradiance normalization technique explained in Olsson et al. (2021) and Thapa et al. (2021). The irradiance normalization technique uses either a polynomial fit or a spline fit depending upon the type of the data. The fit is carried out with emphasis on how well the chosen polynomial or spline fit appears at the start and end of the flight. This is because the panel images used for the radiometric calibration are captured at that stage of the flight. The flights that are normalized this way are assigned a high-quality flag value. An example of a slight variation in irradiance and the normalized irradiance is shown in *Figure 26*.

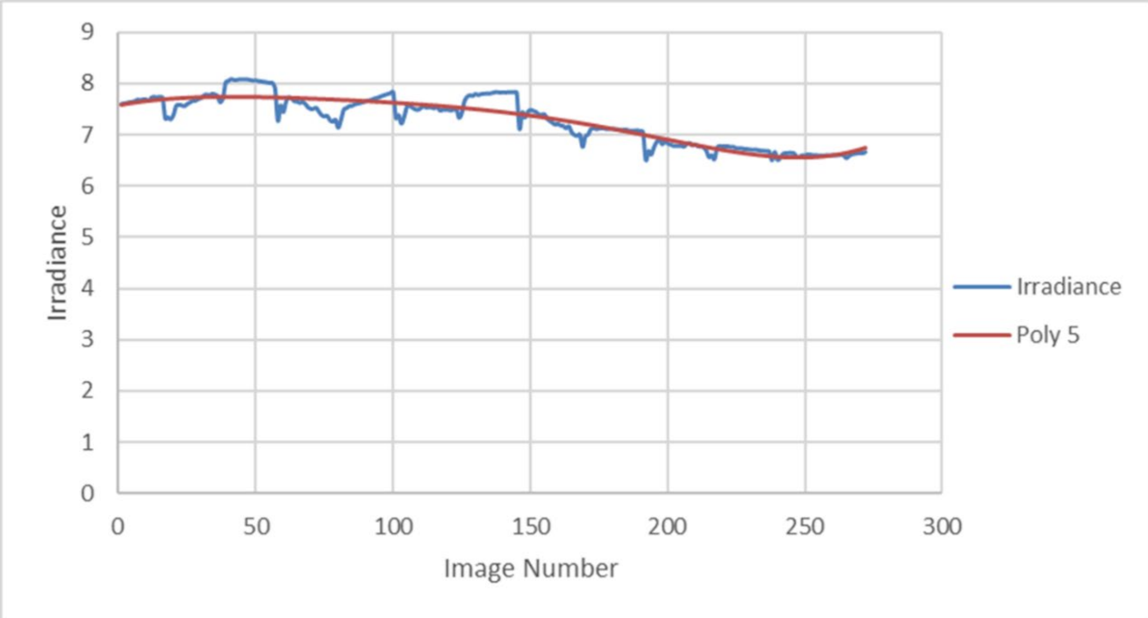


Figure 26: Raw irradiance data from the sunshine sensor onboard UAV (blue line) for a flight (March 28, 2019) conducted at SITES Lönnstorp Research Station. The irradiance data showed slight variation throughout the flight. So, a fifth-order polynomial fit (red line) is used to normalize the variation. Note the fit at the start and end of the flight as it follows the pattern quite nicely there including the rest of the flight irradiance data.

After the radiometric accuracy has been analyzed, the data are also checked for geometric accuracy. *Table 12* shows different cases that affect the geometric accuracy:

Table 12: Different cases that affect the geometric accuracy of a particular UAV flight. Each of the cases is coded with quality flag values as 10, 20, and 30 referring to low, medium, and high quality respectively.

Cases	**Accuracy**	**Code**	**Quality**
No geolocation information, No GCPs but geolocation (UAV-GPS)	Several Meters	10	Low
No. of GCPs (>=4) Root Mean Square Error (RMSE)*	Decimeters	20	Medium
RTK Enabled UAV flights	Centimeters	30	High
*Note: Depends heavily on GCP distribution (Awasthi et al. 2019)			

Lastly, the different levels of derived UAV products, such as orthomosaics (L2), dense point clouds (L2), Digital Elevation Model (DEM, L3), and an NDVI layer (L2) are checked to see if they have any visual artifacts. The visual artifacts are warped vegetations/objects, blurry patches, holes/no data within the orthomosaics, and so on. If the data have any visual artifacts, the data are tagged with quality code value of 1 or 2, referring to low and high quality respectively.

Finally, all these quality code values for each of the three parameters are summed up to generate a 3-digit QFLAG value (*Table 13*). In the table, the lowest quality is 111 while 332 is the highest possible data quality that can be expected for MSP UAV data.

Table 13: QFLAG values for multispectral UAV data. All 3 parameters (radiometric accuracy, geometric accuracy, visual artifacts) are assigned quality code values of 100, 200 or 300; 10, 20, or 30 for radiometric and geometric accuracy referring to the low, medium, and high quality while for the visual artifacts, there are only two quality code 1 or 2, referring to low and high, respectively. The quality code values for individual parameters are summed and the result is the final QFLAG value.

Parameters			
Radiometric Accuracy	**Geometric Accuracy**	**Visual Artifacts**	**QFLAG**
100	10	1	111
		2	112
	20	1	121
		2	122
	30	1	131
		2	132
200	10	1	211
		2	212
	20	1	221
		2	222
	30	1	231
		2	232
300	10	1	311
		2	312
	20	1	321
		2	322
	30	1	331
		2	332

5.3.2 RGB UAV Data

For the RGB UAV data, two parameters are considered:

1) Geometric accuracy
2) Visual artifacts

Geometric accuracy is handled in the same way as the MSP flight (*Table 13*). Apart from this, the RGB data are checked for visual artifacts and the method follows a similar strategy for quality coding as the MSP data. Radiometry is not considered for the quality flagging of RGB data. However, the parameter is kept as a placeholder for future modification and hence has the quality code value of 100. The data that have high-quality geometric accuracy and no signs of visual artifacts are assigned high QFLAG while the data having low-quality geometric accuracy and containing visual artifacts get a low QFLAG value. The data that have mixed geometric accuracy and the presence or absence of visual artifacts are assigned medium quality. The visual artifacts on the orthomosaics and dense point cloud could be the combination of various factors such as lower frontal and side overlap set during the flight, poor geometric accuracy of the ground control points (GCP) used for georeferencing the UAV data products, GCP numbers and their distribution (Awasthi et al. 2019), and the different image processing parameters adopted in image processing software (Tinkham, 2021). Figure 27 and 28 show examples of low-quality UAV products (orthomosaics and dense point clouds).

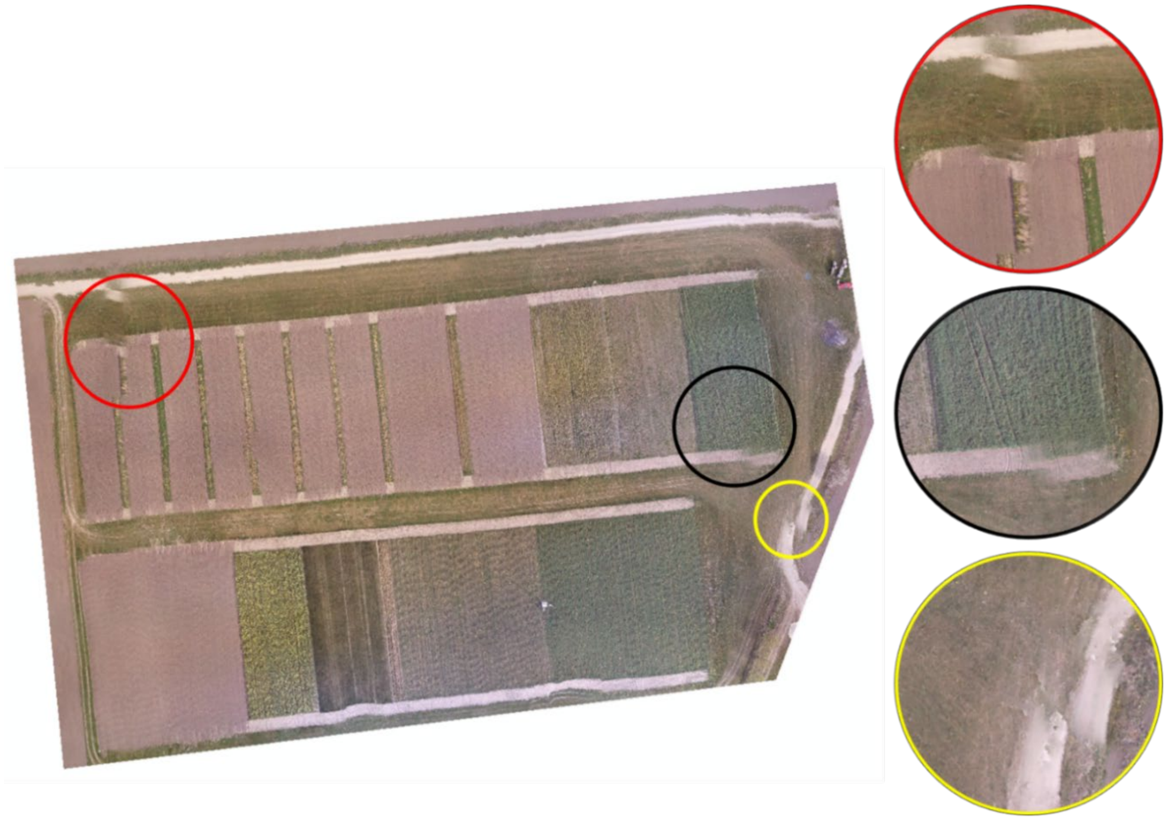


Figure 27: Visual artifacts in the orthomosaics. Three different colored circles on the orthomosaic refer to the artifacts that are zoomed in to show the mismatches and shifting of the features and blurry vegetations. [Visualization: QGIS]

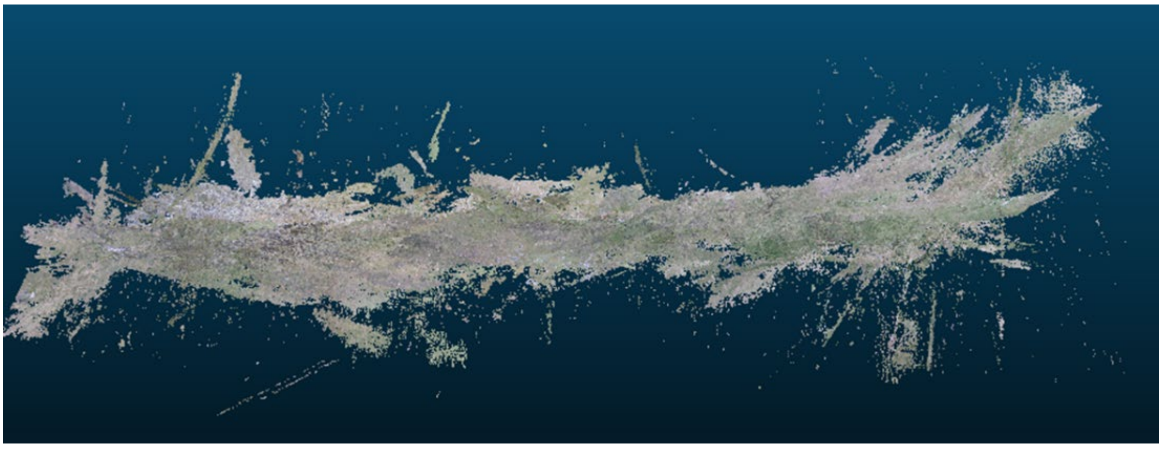


Figure 28: Visual artifacts in the dense point clouds. The profile view of the dense point clouds (L2) derived from UAV-RGB data captured by Parrot Sequoia sensor shows a lot of noise in

the data, especially due to the rolling shutter effect of Sequoia sensor Such a low-quality dense point cloud will result in low-quality DEM data [Visualization: Cloud Compare].

After considering all the parameters, the QFLAG values are assigned to the RGB UAV flight data depending on the quality codes for each parameter. The lowest quality is 111 while 132 is the highest possible data quality for RGB UAV data (*Table 14*).

Table 14: QFLAG values for RGB UAV data. Radiometry is not considered but is assigned a quality code of 100 to be consistent with the 3-digit quality flag values. The geometric accuracy is assigned with quality codes 10, 20, or 30 referring to low, medium, and high-quality data and visual artifacts are assigned quality codes 1 or 2 referring to low and high quality, respectively. The combination of these parameters is the resultant QFLAG value for RGB UAV data.

Parameters			
Radiometric Accuracy	**Geometric Accuracy**	**Visual Artifacts**	**QFLAG**
100	10	1	111
		2	112
	20	1	121
		2	122
	30	1	131
		2	132

5.4 QFLAG system for Satellite Data

The satellite derived HRVPP data products i.e., both the seasonal trajectories (ST) and vegetation phenology and productivity (VPP) when downloaded, offers the possibility to also download the quality file associated with each of the products.

5.4.1 Seasonal Trajectories Quality Flag (ST-QFLAG)

A bitwise encoded status map is provided which contains QFLAG data for the seasonal trajectories related satellite tile downloaded. This map is a 8 bit unsigned raster with same spatial and temporal resolution as that of the seasonal trajectory for a given tile ID. The thematic pixel values for the QFLAG raster layer is given in *Table 15* below. For detailed information about the VPP-QFLAG, see the HRVPP data product manual.

Table 15. Detailed description of quality code values for sesaonal trajectory data product. Quality flag layer is a bitwise encoded status map that defines quality of each pixels.

Value	**Quality**	**Definition**
5	High	More than 8 Clear-sky land observations found in a 91-day-window
4	Medium	3 to 8 Clear-sky land observations found in a 91-day-window
3	Low	1 to 2 Clear-sky land observations found in a 91-day-window
2	Filled (Interpolation)	Clear-sky land observation(s) found on both left and right sides but outside a 91-day-window
1	Filled (Extrapolation)	Clear-sky land observation found on one side (left or right) but outside a 91- day-window
0	No data	The time series was not processed

5.4.2 Vegetation Phenology and Productivity Quality Flag (VPP-QFLAG)

A bitwise encoded status map is provided which contains QFLAG data for the vegetation phenology and productivity related satellite data tile downloaded. This map is a 8 bit unsigned raster with same spatial and temporal resolution as that of the VPP data for a given tile ID. The thematic pixel values for the QFLAG raster layer is given in *Table 16* below. For detailed information about the VPP-QFLAG, see the HRVPP data product manual.

Table 16. Detailed description of quality code values for vegetation phenology and productivity data product. Quality flag layer is a bitwise encoded status map that defines quality of each pixels.

Value	**Quality**	**Definition**
10	High	More than 2 clear-sky land observations found during green up (20%-80% of the seasonal PPI amplitude), green down (80%-20% of the seasonal PPI amplitude), and green peak (left 80% - right 80% of the seasonal PPI amplitude)

9	Medium	More than 2 clear-sky land observations found during green down and green peak
8	Medium	More than 2 clear-sky land observations found during green up and green down
7	Medium	More than 2 clear-sky land observations found during green up and green peak
6	Low	More than 2 clear-sky land observations found during green up
5	Low	More than 2 clear-sky land observations found during green peak
4	Low	More than 2 clear-sky land observations found during green down
3	-	Reserved
2	Filled	Cannot find more than 2 clear-sky land observations during green up, green down or green peak
1	No season found	No season found (the time series was processed)
0	No data	The time series was not processed

Chapter 6. Uploading data to SITES data portal

All the different spectral data products generated after multiple steps of processing are uploaded to the SITES data portal. The data products are uploaded to the data portal in different formats. The formats varies across sensor types and data levels. For example the fixed multispectral sensor based data products are all uploaded as a comma separated *.csv* files. The metadata for these sensors are included in the data file itself as multiple header lines. Also, the time series (L3) of GCC and RCC spectral indices for PhenoCam sensors follows the same strategy.

On the other hand, the other remaining data levels of PhenoCams i.e. L0, L1, and L2 as well as the UAV data levels (L0, L1, L2, L3) are all uploaded as a compressed zipped *.zip* files. The same applies to satellite data as well. The zipped file contains all mandatory file extensions for a particular data type, processing reports, and the metadata associated with the data.

There are two ways, a particular data can be uploaded to the data portal: 1) by using the SITES data upload page and 2) by using the automated Python scripts.

The SITES Data upload page can be accessed at https://meta.fieldsites.se/uploadgui/ and only the authorized users can upload the data to the portal using the login credentials. *Figure 29* shows the general layout of the data upload page.

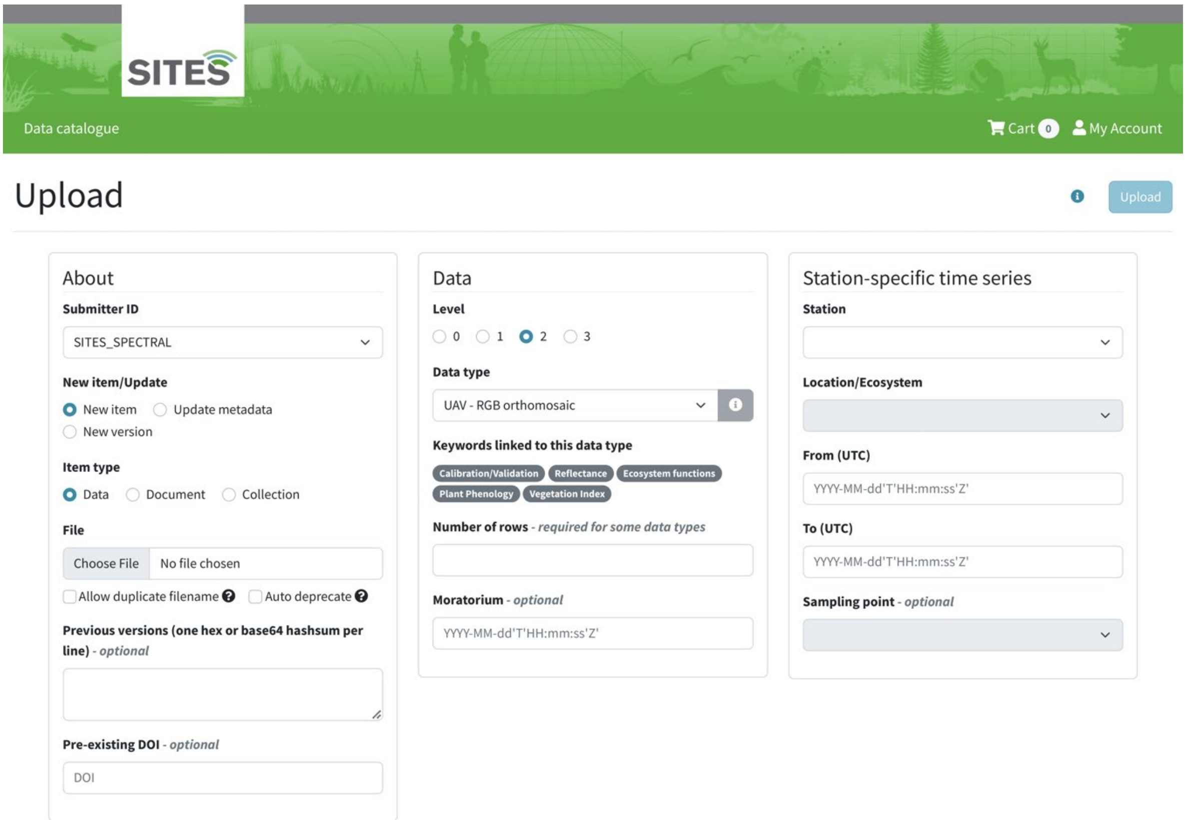


Figure 29. An outlook to SITES data upload page. Accessed on: March 15, 2023.

There are multiple mandatory parameters as you can see on the upload page that needs to be filled before uploading the data. The *Submitter ID* in this case would be SITES_SPECTRAL. The data upload page allows to add a *new item*, just *update metadata* of already existing data or completely upload a *new version* of the data. The upload can be of different *item type*, the one allowed by the portal is *data*, a *document* or a *collection* of various data and documents. Then, navigate to the *file* that is finalized for uploading. Next is to select the data *level* and choose the *data type* defined for the defined data level from the drop down menu. For L3 time series data in *.csv* format, the portal also needs the information on number of rows available in the data being uploaded.

Lastly, the station specific information definition such as the *station* that the data belongs to, *location/ecosystem* for the data, timestamp (*from* and *to* in standard shown in the box) of the data and *sampling location* if any. As soon as all these parameters are defined, hit the *Upload* button on the top right corner and the data will be uploaded to the portal. The metadata page will be accessible directly but, it can take up to an hour for the data to be listed on the data portal.

The update metadata option works slightly differently. To update the metadata, select the *update metadata* option and copy the *metadata url* for the existing data from the data portal. Hit *Get* button and you will see all the metadata in the upload page which you can modify as needed. The similar approach if uploading a *new version* of the existing data. Copy and paste the *metadata url* and hit the *Get* button. Then the upload page will allow to choose and upload the *newer version* of the *file* linked to the *metadata url* pasted before.

The first way of uploading data by using the upload page will not work for large files (especially for phenoCam and UAV data products that are more file size demanding, say approximately around 2GB or more) and will not be convenient to upload several files of the same type. And this is where the second data upload option comes in handy.

The second option of uploading the data to the SITES data portal is by using the automated Python script. The script in the first step asks from the authorized user to provide the complete path to the *.zip* files that are to be uploaded. It will then read and lists all the *.zip* files that are available in the given file path and prepares to upload those data by gathering all the parameters as mentioned in the screenshot (*Figure 29*). In the second and final step, the authorized user should provide the *API token value* for the SITES authorized account. The token value can be obtained from the *My Account* menu available on top left corner of the SITES data portal or the data upload page. This information is required every time the script is run. Note that the *API token value* is valid for 24 hours only. *Figure 30* shows a screenshot of the API token value. Everything for Value parameters should be pasted when the script asks for it.

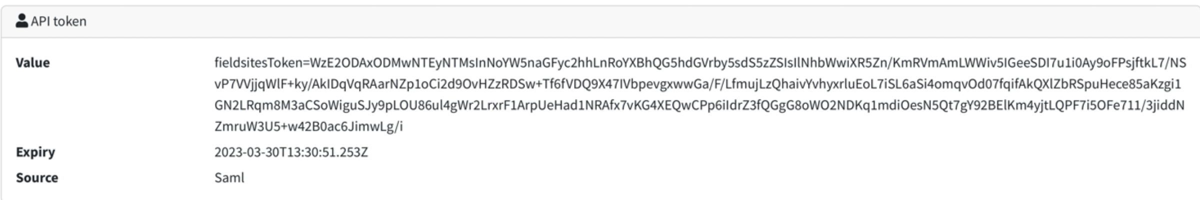


Figure 30. A screenshot of API token value available under My Account menu on the top right corner of the SITES data portal. Accessed on March 15, 2023.

Chapter 7. Navigating and downloading the spectral data in SITES data portal

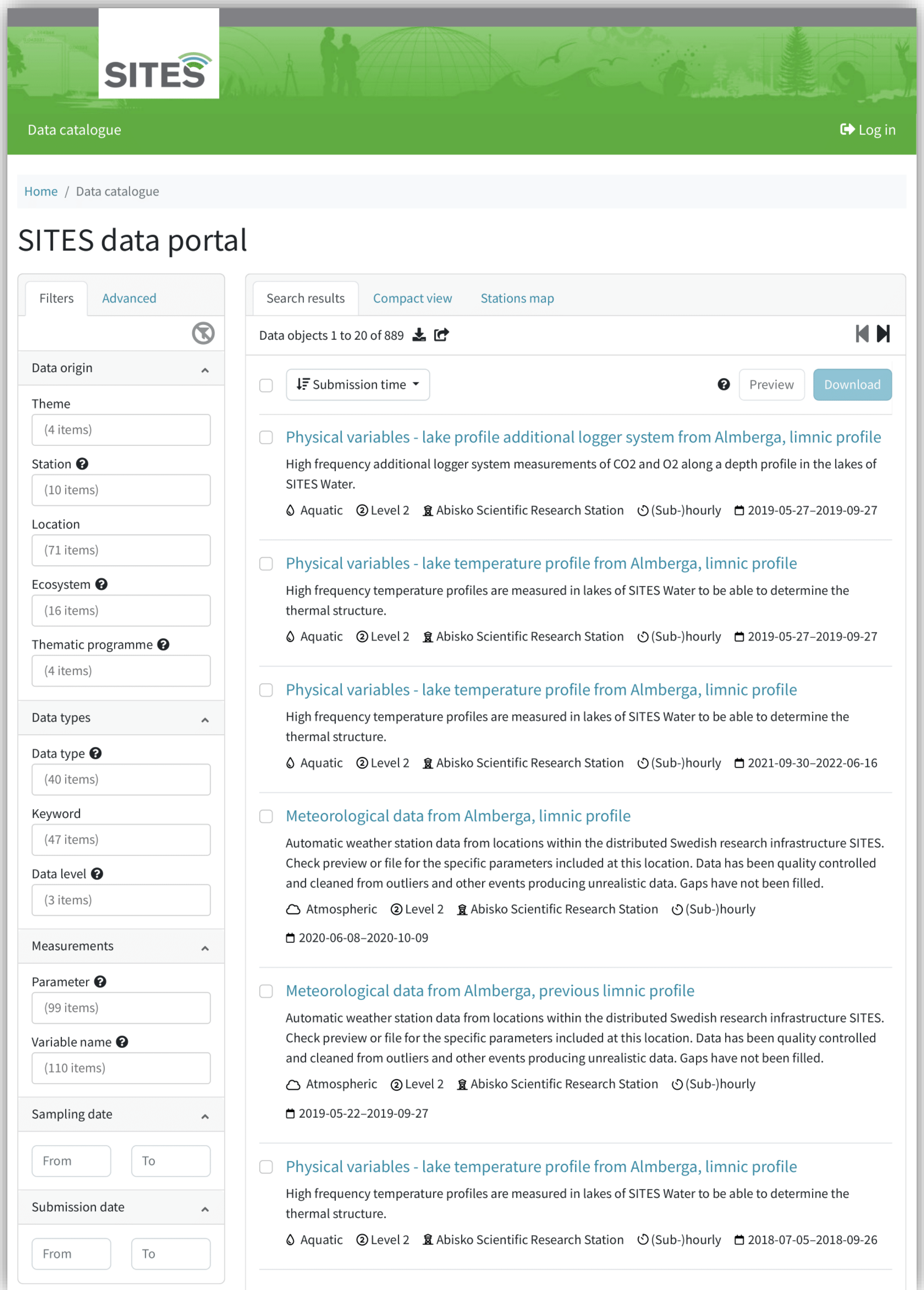


Figure 31. An outlook to the data catalogue page of SITES data portal. Accessed on: March 15, 2023.

The spectral data collected by phenoCams, multispectral sensors, and UAVs at each of the research stations within SITES, together with metadata information are uploaded to the SITES data portal on a regular basis. SITES offers all these data to the researchers free of cost. The URL for the SITES data portal is https://data.fieldsites.se/portal/. When you navigate your browser to this web address, you will see the data catalogue page shown in *Figure 31* above.

The *'About'* tab at the end of the landing page has some basic background information about the SITES project, different thematic programs within it, participating research stations, what the project is focusing on, why and how the collected data could help in answering ecosystem science related questions connected to different spheres of the environment.
The drop down *'SITES Thematic Programs'* tab lists SITES Spectral which contains background information about the project, why the thematic program is interested in spectral data for ecosystem monitoring, and the science questions the project are addressing with data from different proximal and satellite sensors.

Many filters are available on the data catalogue page. Select SITES Spectral in the *'Thematic Programme'* filter. The other filters that could be of interest are the *'Station'*, *'Data type'*, and *'Data level'*. *'Station'* filter helps to choose between the available SITES research station participating in the Spectral thematic program. The available *'Data type'* and *'Data level'* filters could be selected from the drop down as soon as the *'Thematic Programme'* and *'Station'* are defined. The available data that meets the defined filters will show up in the display panel as shown in *Figure 32* where the search is for L1 phenoCam data from Abisko Scientific Research Station.

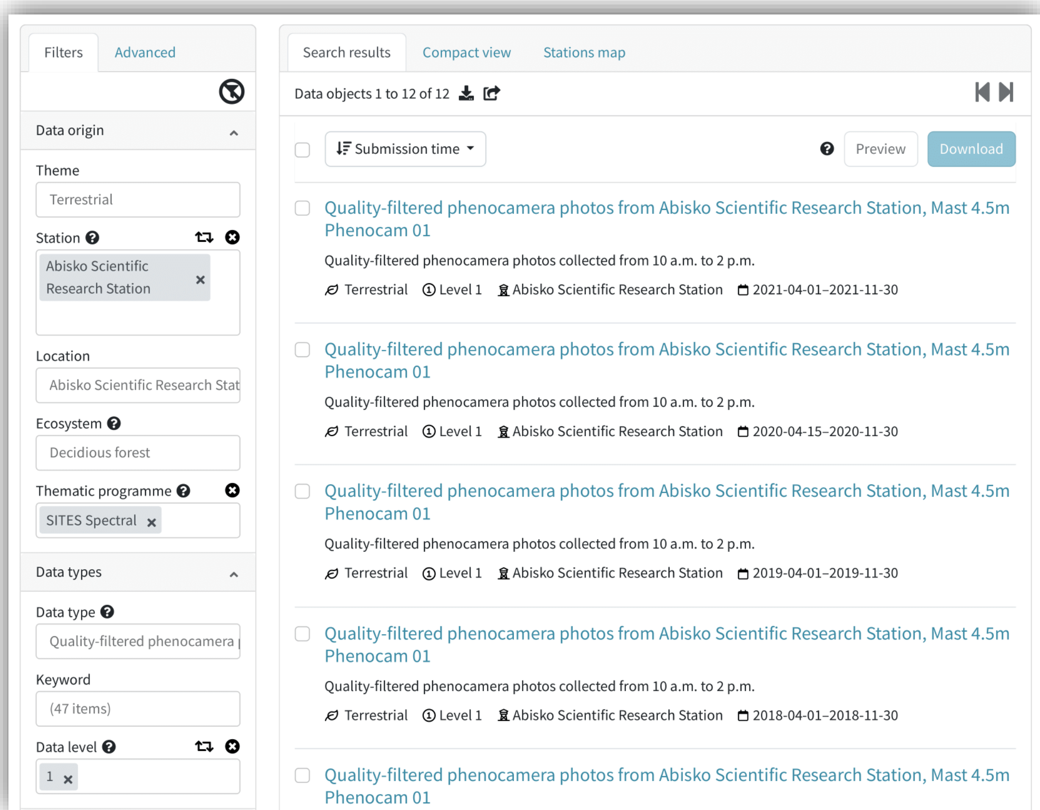


Figure 32. Screenshot showing how different filters can be used for spectral data search for a SITES station. Left pane shows the filters applied and right display pane shows the list of spectral data that match with the filters.

The data portal also offers the facility to search the spectral data using the keywords. On clicking the *'Keyword'* tab, it will list down all the keywords available for data search on the portal.

You can scroll through the lists of available data after the filters are applied. Clicking on the data object of interest on the display pane will open up the metadata landing page for that object. *Figure 33* shows the contents of metadata landing page for a chosen data object.

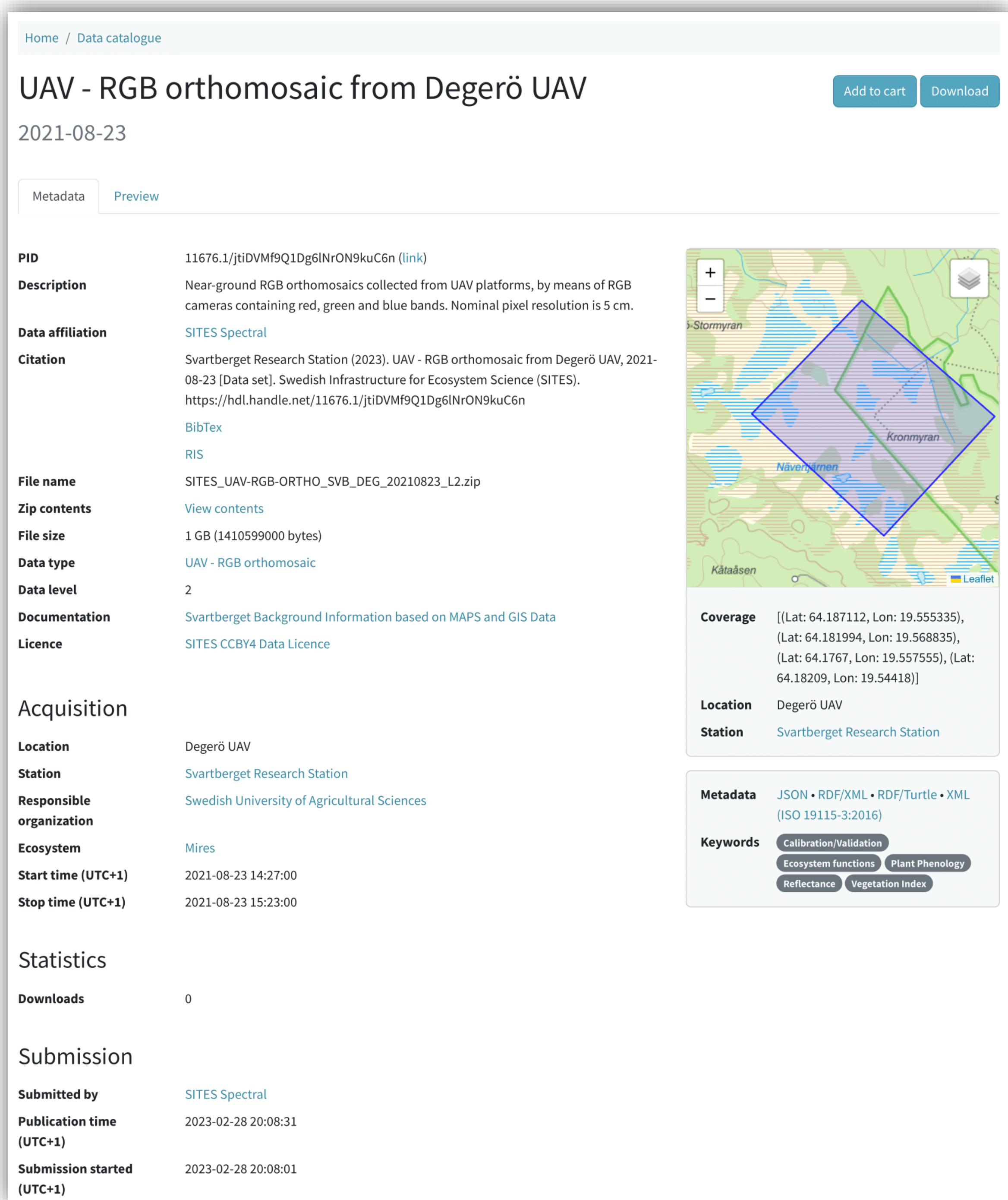


Figure 33. Metadata landing page for a chosen data object. The data object in this example is a RGB UAV orthomosaic for the year 2021 from Degerö location at SITES Svartberget research.

The metadata landing page contains various metadata information associated with a chosen data object. Each data object has a unique persistent identifier (PID) and this can be seen on the metadata landing page for all data objects. Just below the PID, one can find information about the data affiliation followed by a small data description. In addition to this, there is a specific citation for the data, just like the publication citation available in the same metadata landing page. Apart from this, the landing page consists of several other key information such as the name of the data file, file size, data type and level, documentation, licence type, data acquisition station and timestamp, ecosystem as well as download statistics. If the data object is a *.zip* file, the data portal allows to view and download the individual contents of the file without downloading the original *.zip* file. The *'Documentation'* parameter links to the background information and all GIS files associated with the research station and the data.

On the same metadata landing page, there is a preview tab which allows the users to preview the data. *Figure 34* shows the preview funtionality of the data portal allowing the users to scroll down the contents of the *.zip* file and visualize it even before downloading it.

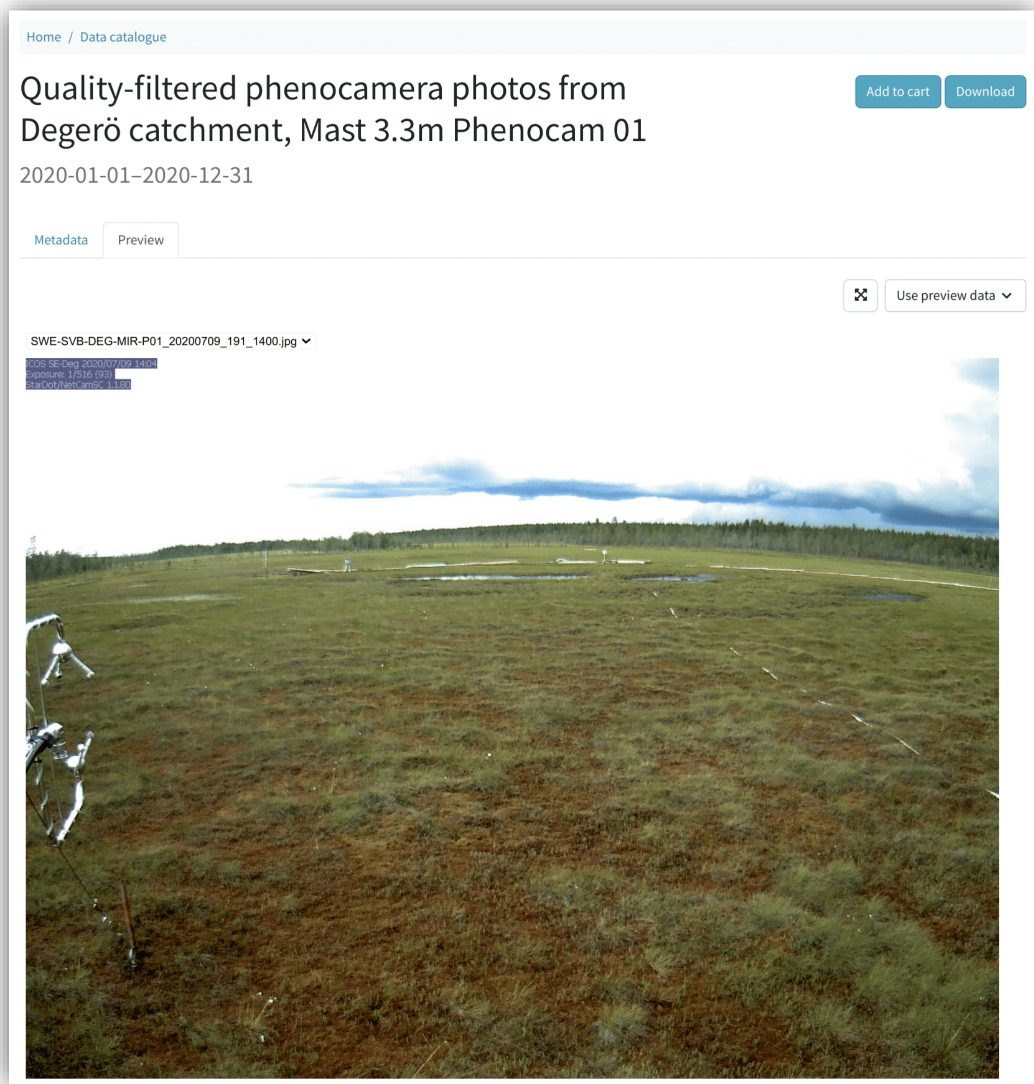


Figure 34. Preview tab offers data visualization of each individual entries within a quality filtered phenoCam data object (.zip) in this case without even downloading it.

Only the data objects that contains .jpg files and the L3 time series data in .csv file format can be previewed without downloading the data. The L3 time series data especially GCC-RCC from phenoCam sensor and L2 time series data from fixed multispectral sensors can be previewed. *Figure 35* shows how the L3 time series *.csv* files can be previewed.

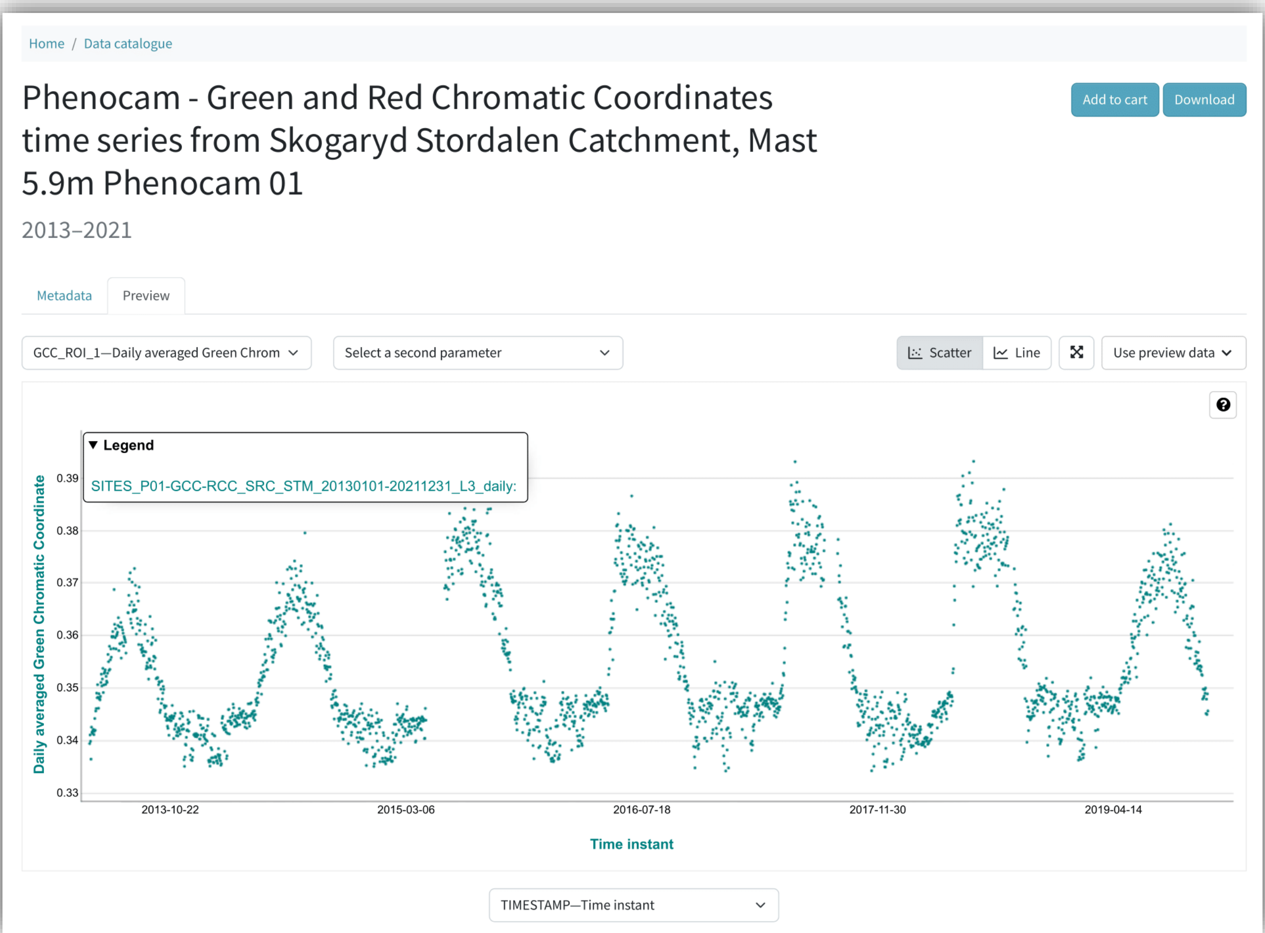


Figure 35. L3 daily time series GCC data visualization through the preview tab. The daily GCC time series data (2013-2019) is from a phenoCam mounted on a mast located at stordalen catchment location in the SITES Skogaryd research station.

The preview function for time series data offers the possibility to plot one or two variables of the data and check how the time series data looks like throughout the year and after. One can switch between different plot styles such as scatter or line plots. The graph can be zoomed, panned and reset to the original scale, the information to which is available in the small help symbol *'?'* on the top right corner of the plotting window. Under *'Use preview data'* drop down menu, the data portal allows to copy the link to embed the current plot to any other web services. In addition, there is also provision to parse the previewed data to a program like R and Python, the documentation to this functionality is also provided on clicking the help symbol *'?'* like mentioned before.
Finally, after checking the contents of the *.zip* file or the quality of the time series data, one can download the spectral data of interest from the data portal. To download any spectral data, just click on the download button on the upper right corner of the metadata landing page. On clicking the download button, you will see the SITES data policy page (*Figure 36*) where all the informations regarding summary to data licence, how to cite the data as well as a short note on PIDs are available. In order to download the data, one needs to accept the SITES data policy and press the download button once again. The file gets downloaded locally to your computer after that.

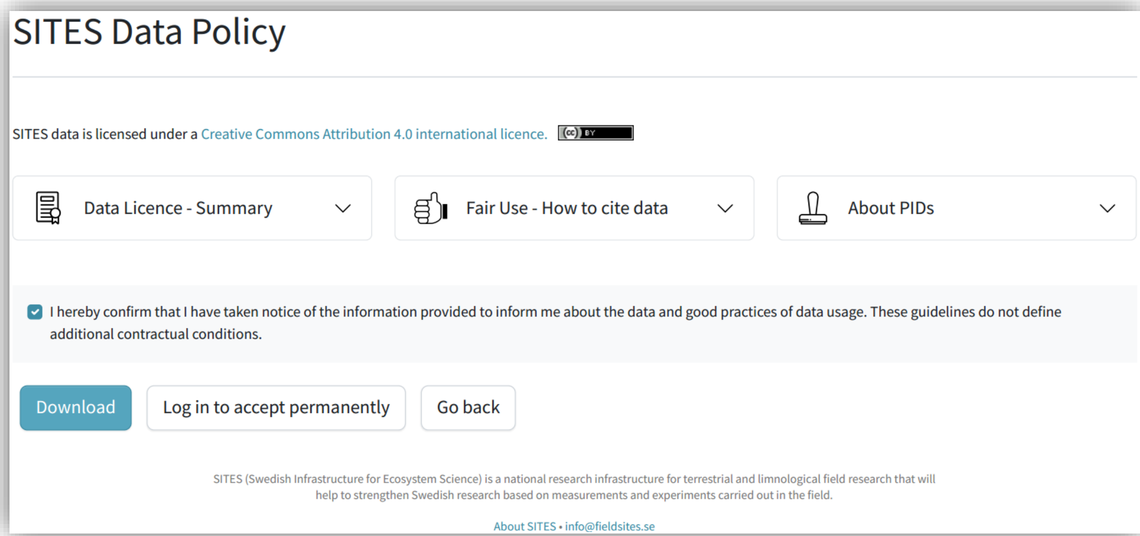


Figure 36. SITES data policy page showing information about the data licence, how to cite data as well as terms and conditions to download the data.

Acknowledgement

When using SITES Spectral data, please acknowledge the infrastructure in any publications, for example by using a modification of this sentence:

"This [study/report/dataset/etc.] has been made possible by the Swedish Infrastructure for Ecosystem Science (SITES), (in this case [at/by station/stations]). SITES receives funding through the Swedish Research Council under the grant no 2017-00635."

Data from the SITES Data Portal (link: https://data.fieldsites.se/portal/) must be properly cited and acknowledged.

Each dataset has a specific citation, like that of a publication citation. The citation is provided on the dataset metadata page.

Example for data specific citation provided on the data portal:
Skogaryd Research Catchment (2022). Phenocam - RGB composites from Skogaryd Central, Mast 3 phenocam, 2019-01-01–2019-12-31 [Data set]. Swedish Infrastructure for Ecosystem Science (SITES). https://hdl.handle.net/11676.1/ZRx1o3XnoqAGSqQuER9zy4xy

The persistent identifier (PID) will always lead you back to the metadata page where you are able to download the data object again or check for updates.

You may also use or link to pictures and visualizations obtained from the SITES Data Portal if you add the link to the original and the text "SITES, licensed under CCBY" followed by the link to the original source.

We suggest that you inform the data providers when the data are used for publications and that you offer them the possibility to comment and/or offer them co-authorship or acknowledgement in the publication when this is justified by the added value of the data for your results.

References

Agisoft Metashape User Manual: Professional Edition, Version 1.6. Publication Date 2020. Available online: https://www.agisoft.com/downloads/user-manuals/ (accessed on 20 July 2021).

Awasthi, B., S. Karki, P. Regmi, D. S. Dhami, S. Thapa, and U. S. Panday. 2019. Analyzing the Effect of Distribution Pattern and Number of GCPs on Overall Accuracy of UAV Photogrammetric Results. In *International Conference on Unmanned Aerial System in Geomatics*, 339-354. Springer.

Cai, Z., P. Jönsson, H. Jin, and L. Eklundh (2017) Performance of Smoothing Methods for Reconstructing NDVI Time-Series and Estimating Vegetation Phenology from MODIS Data. *Remote Sensing,* 9.

Eklundh, L., H. Jin, P. Schubert, R. Guzinski, and M. Heliasz (2011) An Optical Sensor Network for Vegetation Phenology Monitoring and Satellite Data Calibration. *Sensors,* 11**,** 7678-7709.

Fensholt, R., I. Sandholt, and M. S. Rasmussen. 2004. Evaluation of MODIS LAI, fAPAR and the relation between fAPAR and NDVI in a semi-arid environment using in situ measurements. *Remote Sensing of Environment,* 91**:** 490-507. DOI: https://doi.org/10.1016/j.rse.2004.04.009

Gamon, J. A., L. Serrano, and J. S. Surfus. 1997. The photochemical reflectance index: an optical indicator of photosynthetic radiation-use efficiency across species, functional types, and nutrient levels. *Oecologia,* 112**:** 492-501. DOI: 10.1007/s004420050337

Gao, B.-c. 1996. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment,* 58**:** 257-266. DOI: https://doi.org/10.1016/S0034-4257(96)00067-3

Gillespie, A. R., A. B. Kahle, and R. E. Walker. 1987. Color enhancement of highly correlated images. II. Channel ratio and "chromaticity" transformation techniques. *Remote Sensing of Environment,* 22**:** 343-365. DOI: https://doi.org/10.1016/0034-4257(87)90088-5

Gitelson, A., and M. N. Merzlyak. 1994. Spectral Reflectance Changes Associated with Autumn Senescence of Aesculus hippocastanum L. and Acer platanoides L. Leaves. Spectral Features and Relation to Chlorophyll Estimation. *Journal of Plant Physiology,* 143**:** 286-292. DOI: https://doi.org/10.1016/S0176-1617(11)81633-0

Goldman, D. B. 2010. Vignette and exposure calibration and compensation. *IEEE transactions on pattern analysis and machine intelligence,* 32**:** 2276-2288.

Jiang, Z., A. R. Huete, K. Didan, and T. Miura. 2008. Development of a two-band enhanced vegetation index without a blue band. *Remote Sensing of Environment,* 112**:** 3833-3845.

Jin, H., and L. Eklundh. 2015. In Situ Calibration of Light Sensors for Long-Term Monitoring of Vegetation. *IEEE Transactions on Geoscience and Remote Sensing,* 53**:** 3405-3416. DOI: 10.1109/TGRS.2014.2375381

Keenan, T., B. Darby, E. Felts, O. Sonnentag, M. Friedl, K. Hufkens, J. O'Keefe, S. Klosterman, et al. 2014. Tracking forest phenology and seasonal physiology using digital repeat photography: a critical assessment. 24**:** 1478-1489.

Liu, Y., C. Wu, O. Sonnentag, A. R. Desai, and J. Wang. 2020. Using the red chromatic coordinate to characterize the phenology of forest canopy photosynthesis. *Agricultural and Forest Meteorology,* 285-286**:** 107910. DOI: https://doi.org/10.1016/j.agrformet.2020.107910

Migliavacca, M., M. Galvagno, E. Cremonese, M. Rossini, M. Meroni, O. Sonnentag, S. Cogliati, G. Manca, et al. 2011. Using digital repeat photography and eddy covariance data to model grassland phenology and photosynthetic CO2 uptake. 151**:** 1325-1337.

Olsson, P.-O., A. Vivekar, K. Adler, V. E. Garcia Millan, A. Koc, M. Alamrani, and L. Eklundh (2021) Radiometric Correction of Multispectral UAS Images: Evaluating the Accuracy of the Parrot Sequoia Camera and Sunshine Sensor. *Remote Sensing,* 13.

Olsson, P. O., M. Heliasz, H. Jin, and L. Eklundh. 2017. Mapping the reduction in gross primary productivity in subarctic birch forests due to insect outbreaks. *Biogeosciences,* 14**:** 1703-1719. DOI: 10.5194/bg-14-1703-2017

Peltoniemi, M., M. Aurela, K. Böttcher, P. Kolari, J. Loehr, J. Karhu, M. Linkosalmi, C. M. Tanis, et al. 2018. Webcam network and image database for studies of phenological changes of vegetation and snow cover in Finland, image time series from 2014 to 2016. 10**:** 173-184.

Rahman, A. F., V. D. Cordova, J. A. Gamon, H. P. Schmid, and D. A. Sims. 2004. Potential of MODIS ocean bands for estimating CO2 flux from terrestrial vegetation: A novel approach. *Geophysical Research Letters,* 31. DOI: https://doi.org/10.1029/2004GL019778

Richardson, A. D., K. Hufkens, T. Milliman, D. M. Aubrecht, M. Chen, J. M. Gray, M. R. Johnston, T. F. Keenan, et al. 2018a. Tracking vegetation phenology across diverse North American biomes using PhenoCam imagery. *Scientific Data,* 5**:** 180028. DOI: 10.1038/sdata.2018.28

Richardson, A. D., K. Hufkens, T. Milliman, D. M. Aubrecht, M. E. Furze, B. Seyednasrollah, M. B. Krassovski, J. M. Latimer, et al. 2018b. Ecosystem warming extends vegetation activity but heightens vulnerability to cold temperatures. *Nature,* 560**:** 368-371. DOI: 10.1038/s41586-018-0399-1

Richardson, A. J., and C. L. Wiegand. 1977. Distinguishing vegetation from soil background information. *Photogrammetric Engineering and Remote Sensing,* 43**:** 1541-1552.

Rouse, J. W., Jr., R. H. Haas, J. A. Schell, and D. W. Deering. 1973. Monitoring vegetation systems in the Great Plains with ERTS. *"Third ERTS Symposium, NASA SP-351, Washington DC, Dec. 10-14 1973"***:** 309-317.

Seyednasrollah, B., A. M. Young, K. Hufkens, T. Milliman, M. A. Friedl, S. Frolking, and A. D. Richardson. 2019. Tracking vegetation phenology across diverse biomes using Version 2.0 of the PhenoCam Dataset. *Scientific Data,* 6**:** 222. DOI: 10.1038/s41597-019-0229-9

Smith, G. M., and E. J. J. I. J. o. r. s. Milton. 1999. The use of the empirical line method to calibrate remotely sensed data to reflectance. 20**:** 2653-2662.

Stow, D., C. J. Nichol, T. Wade, J. J. Assmann, G. Simpson, and C. J. D. Helfter. 2019. Illumination geometry and flying height influence surface reflectance and NDVI derived from multispectral UAS imagery. 3**:** 55.

Thapa, S., V. E. Garcia Millan, and L. Eklundh. 2021. Assessing Forest Phenology: A Multi-Scale Comparison of Near-Surface (UAV, Spectral Reflectance Sensor, PhenoCam) and Satellite (MODIS, Sentinel-2) Remote Sensing. *Remote Sensing,* 13. DOI: 10.3390/rs13081597

Tinkham, W. T., and N. C. Swayze (2021) Influence of Agisoft Metashape Parameters on UAS Structure from Motion Individual Tree Detection from Canopy Height Models. *Forests,* 12.