User Guide to Collection 6.1 MODIS Land Cover Dynamics (MCD12Q2) Product

Josh Gray, Damien Sulla-Menashe, and Mark A Friedl

March 14, 2022

The Collection 6.1 (C6.1) MODIS Land Cover Dynamics Product (MCD12Q2) provides science data sets (SDSs) that map global land surface phenology metrics (hereafter: "phenometrics") at 500 meter spatial resolution and annual time step. Phenometrics are derived from time series of MODIS observed land surface greenness, specifically: time series of the 2-band Enhanced Vegetation Index calculated from MODIS nadir BRDF adjusted surface reflectances (NBAR-EVI2).

This user guide provides the following information related to the C6.1 product:

1. An overview of the MCD12Q2 algorithm

2. Important differences from the C5 product

3. Guidance on data portals, projections, and formats, to help users access and use the data.

4. Contact information for users with questions that cannot be addressed through information or websites provided in this document.

5. A table describing the different data sets and associated scaling and fill values.

C6.1 provides a minor update to C6 that resolves one main issue in the C6 product. Specifically, in C6.1, a fix was implemented to resolve spuriously early detection of greenup dates in a small proportion of pixels caused by a discontinuity in vegetation index time series across calendar years arising from the way that splines are fit to the data (see Section 2.1). In addition, the processing pipeline was adjusted to allow product generation 6 months after the end of the calendar year, instead of 12 months, which allows the product to be generated in a more timely fashion.

1   Product Overview

The MODIS C6.1 MCD12Q2 Land Surface Dynamics Product provides global maps of phenometrics at 500 m and an annual time step since 2001. The product contains 25 SDS (REF TABLE) that record the timing of phenometrics such as the onset of greenness increase, peak greenness, senesence, and dormancy; derivative features of the vegetation cycles such as the NBAR-EVI2 amplitude, NBAR-EVI2 minimum, and the integrated NBAR-EVI2 over a vegetation cycle; as well as the total number of vegetation cycles detected for the product year, and overall and phenometric-specific QA/QC information. Information is provided for up to two detected vegetation cycles to record multicropping, precipitation-driven greenup events, and other phenological regimes that deviate from the common single annual greenup/greendown event pattern.

1.1 Differences from Collection 5 Product

The C6.1 methodological approach and SDSs differ substantially from the previous C5 product. Changes were designed to better capture phenometrics in systems with multiple vegetation cycles per year, to increase the reliability of retrieved phenometrics in tropical, arid, and semi-arid ecosystems, to more accurately represent phenometrics in systems where NBAR-EVI2 time series do not closely resemble logistic growth patterns, to provide working QA/QC SDS, and to deliver phenometrics in a more usable and intuitive fashion when vegetation cycles cross calendar boundaries, especially in the Southern Hemisphere. Users of C5 products should note the additional SDS, the provision of overall and bit-packed, phenometric-specific QA/QC SDSs, the change to a days since 1-1-1970 standard for date delivery, and the adoption of a new standard for assigning vegetation cycles to product years: all phenometrics are delivered for vegetation cycles where the date of peak NBAR-EVI2 is within the product's calendar year. This particular change should ease difficulties in collating phenometrics across product years when vegetation cycles cross calendar year boundaries by ensuring that all phenometrics for a vegetation cycle are fully contained in a single product year.

2 Methodology

The C6.1 MCD12Q2 product is created by assembling time series of NBAR-EVI2, eliminating outliers and filling dormant period values, fitting a QA/QC-weighted penalized cubic smoothing spline to the time series, identifying valid vegetation cycles within the time series, and then extracting and recording phenometrics for each vegetation cycle. All subsequent functions are applied to a time series of three consecutive years: the product year and preceding and subsequent years.

2.1 Spline Fit

NBAR-EVI2 observations identified as being snow contaminated by the MCD432 product, or where the corresponding NBAR-NDSI > -0.2, are filled with a dormant period value. The dormant NBAR-EVI2 is taken to be the 5th percentile snow-free value across the product year and its preceding and subsequent year, unless this differs by more than 25% from the 10th percentile snow-free value for the central year, in which case only the central year's data are used to compute the background NBAR-EVI2 value. This is to accommodate land cover changes that may alter the true dormant period NBAR-EVI2 value. Missing NBAR-EVI2 values between snow-contaminated observations are also filled with the dormant NBAR-EVI2 value. Finally, a penalized cubic smoothing spline is fit to the value with the weights proportional to the MCD43A2 QA/QC flags.

2.2 Identifying Valid Vegetation Cycles

Valid vegetation phases in the time series are identified as periods of sustained increase in NBAR-EVI2 followed by sustained periods of decrease, subject to ecologically meaningful heuristics concerning the length of these periods and magnitude of the NBAR-EVI2 amplitude. First, "peaks" in the NBAR-EVI2 time seres are identified as points where the first derivative (numerically evaluated) changes sign (Fig. 1A). These candidate peaks are sorted by NBAR-EVI2 magnitude and then analyzed with a recursive function that attempts to identify associated start and end dates for the greenup/down segment.
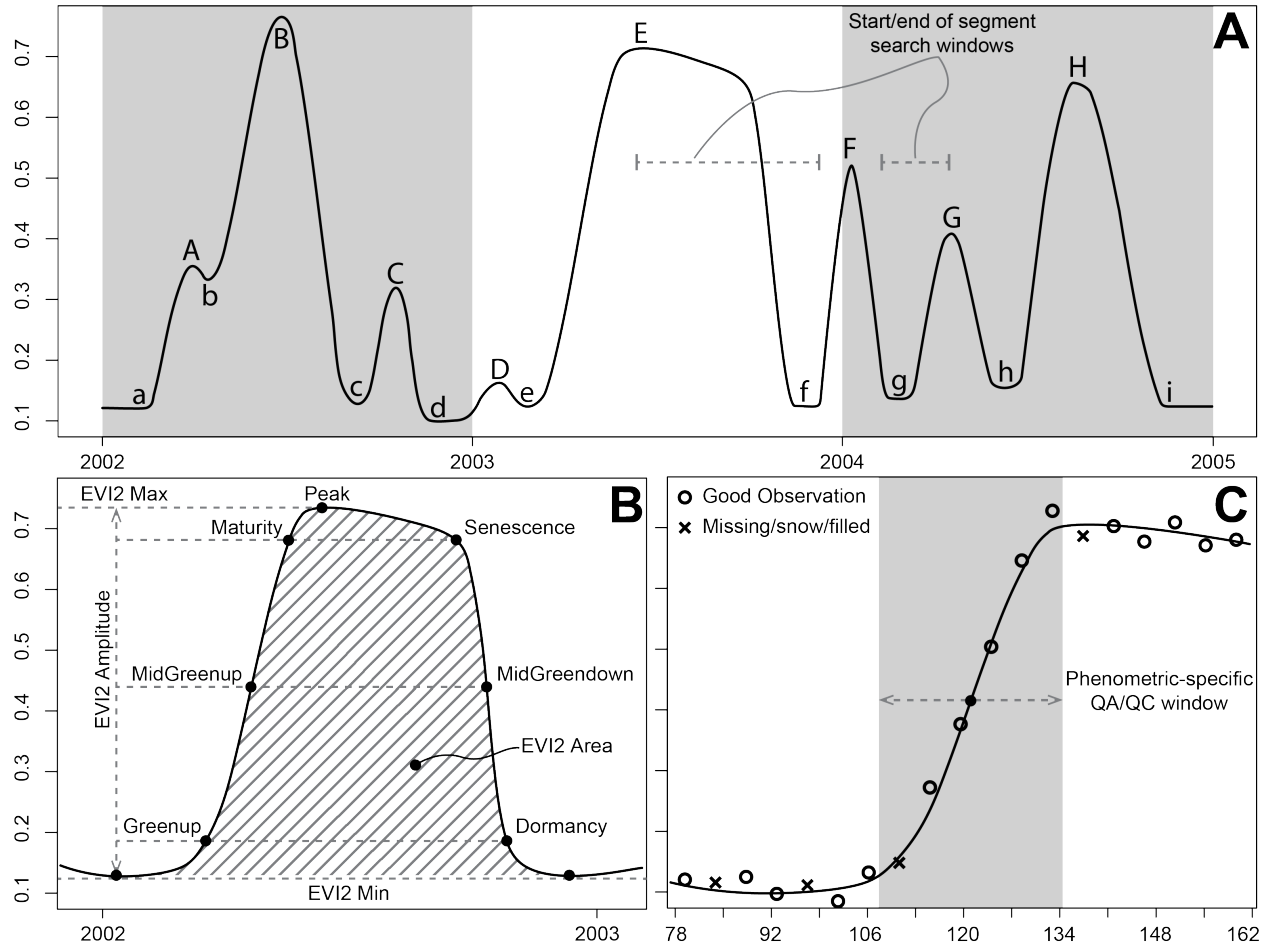
Figure 1: Three years of a hypothetical splined NBAR-EVI2 time series with potential peaks (upper-case letters), troughs (lower-case letters), and search windows for start/end of the segment corresponding to peak F marked (A). Diagram of phenometrics retrieved for a single hypothetical vegetation cycle (B). Illustration of the methodology used to determine the phenometric-specific QA/QC score (C).

Starting with the smallest peak, the start of the greenup segment is sought as the date of the minimum NBAR-EVI2 value closest to the candidate peak between 30 days prior to peak and the maximum of: the start of the time series, the closest of any preceding candidate peaks, or the candidate peak date minus a maximum greenup period length parameter (185 days). This search windows are depicted in Fig. 1A for the segment associated with peak F. If the NBAR-EVI2 amplitude between this minimum value and the peak is greater than or equal to 0.1, and greater than or equal to 35% of the time series maximum NBAR-EVI2 minus the minimum NBAR-EVI2, then the identified greenup period is considered valid and an associated greendown period is sought. If these conditions are not met, then the candidate peak is removed from the list of candidate peaks. The end of the greendown period is sought in an identical manner, but in the opposite temporal direction. A candidate end date is identified as the date of the closest minimum NBAR-EVI2 value between 30 days after the candidate peak and the minimum of: the end of the time series, the closest of any following peaks, or the candidate peak date plus a maximum greendown length parameter (set to 185 days; Fig. 1A). The entire greenup and greendown segment is considered a valid vegetation cycle if the NBAR-EVI2 amplitude between the candidate segment end date and candidate peak is greater than or

3

equal to 0.1. Otherwise, the candidate peak is eliminated. This process is repeated, each time updating the candidate peak list, until all candidate peaks are either eliminated or identified as belonging to a valid vegetation cycle. The full 3-year time series is thus partitioned into valid vegetation cycles.

For example, in Fig. 1A, the potential peaks would first be sorted by NBAR-EVI2 magnitude: **D**, **C**, **A**, **G**, **F**, **H**, **E**, and **B**. A segment start for potential peak **D** would be identified as trough **d** and the NBAR-EVI2 amplitude between points **d** and **D** calculated. Because this amplitude is less than the global minimum of 0.1, peak **D** would be eliminated from the list of potential peaks. Similarly, the segment associated with peak **C** would be eliminated because although the NBAR-EVI2 magnitude between points **c** and **C** exceeds 0.1, it is less than 35% of the maximum NBAR-EVI2 variation exhibited over the three-year interval (minimum point **d** and maximum point **B** depict an amplitude of 0.67, so the 35% threshold is 0.234, but **c** to **C** is 0.2). Likewise, the potential segment associated with peak **A** would be eliminated because while the NBAR-EVI2 amplitude between **a** and **A** exceeds 0.1 and the 35% threshold, the greendown portion (points **A** to **b**) does not. The segment start/end search windows shown in Fig. 1A are for the segment associated with peak **F**. A segment start is sought as a NBAR-EVI2 minimum point between the preceding peak **E** and 30 days prior to **F**. The segment end date is sought between a period 30 days after **F** and the next valid peak (**G**). Valid vegetation cycles would be retrieved for the segments associated with peaks **B**, **E**, **F**, **G**, and **H**.

2.3 Phenometric Extraction

Identification of phenophase transition dates is performed on each of the identified vegetation cycles for which the peak date falls within the calendar year of the product year (e.g. phenometrics associated with peak **F** in Fig. 1A would be delivered with product year 2004). The first metric: number of vegetation cycles, is determined by simply counting the number of such peaks. Subsequent phenophase date retrieval, and quantification of associated QA, are performed only for up to two of the valid segments with highest segment NBAR-EVI2 amplitude. That is, if there are three cycles, the number of vegetation cycles metric will reflect the correct number, but phenophase transitions will not be retrieved for the cycle with the lowest NBAR-EVI2 amplitude (e.g. product year 2004 for the time series represented in Fig. 1A would have 3 for number of cycles, but only phenometrics for segments associated with peaks **F** and **H** would be delivered). The start of greenup, greenup midpoint, and maturity dates are retrieved as the first date within the greenup segment where the NBAR-EVI2 time series crosses 15, 50, and 90% of the greenup segment NBAR-EVI2 amplitude (peak NBAR-EVI2 - segment start EVI; Fig. 1B). Similarly, start of senescence, senescence midpoint, and dormancy are retrieved as the last date within the greendown segment where the NBAR-EVI2 time series crosses 90, 50, and 15% of the greendown segment NBAR-EVI2 amplitude (peak NBAR-EVI2 - segment end NBAR-EVI2; Fig. 1B). All dates are converted to UNIX-epoch time: days since Jan 1, 1970. Segment integrated NBAR-EVI2 is calculated as the sum of daily NBAR-EVI2 values, minus the greenup segment minimum value, between the segment start and end dates (not from the start of greenup and dormancy; Fig. 1B). NBAR-EVI2 amplitude and NBAR-EVI2 minimum values are also recorded as distinct product layers (NBAR-EVI2 maximum can be retrieved as NBAR-EVI2 amplitude - NBAR-EVI2 minimum).

Quality assurance (QA) scores are calculated for each phenometric (detailed QA), and for the entire greenup/down segment (overall QA). These scores are a weighted combination of the fraction of not missing (e.g. due to snow) and not filled (e.g. with dormant NBAR-EVI2 value) values (80% weight), and the spline goodness-of-fit (R^2; 20% weight), in the vicinity of each phenometric (2 weeks before and after; detailed QA; Fig. 1C) or within the entire segment (overall QA). These 0–1 scores are mapped to categories as: [1, 0.75)=0 (best), [0.75, 0.5)=1 (good), [0.5,0.25)=2 (fair), and [0.25, 0]=3 (poor). Detailed QA scores are then

bit-packed into a single 16-bit integer value, whereas the overall QA score is retained as a separate product layer (details on detailed QA below).

3   Data Formats and Projection

MCD12Q2 data are provided as tiles that are approximately 10°x 10°at the Equator using a Sinusoidal grid in HDF4 file format. Information related to the MODIS sinusoidal projection and the HDF4 file format can be found at:

- MODIS tile grid: `http://modis-land.gsfc.nasa.gov/MODLAND_grid.html`

- MODIS HDF4: `http://www.hdfgroup.org/products/hdf4/`

Several parameters are needed to reproject the Sinusoidal HDF4 files to other projections using widely used software such as GDAL. Here we provide the values used for the upper left corner of the grid, the size of a single pixel, and the Sinusoidal projection string in Cartographic Projections Library (PROJ4) and Well-Known Text (WKT) formats.

- ULY Grid = 10007554.677, ULX Grid = -20015109.354

- Pixel Size (m) = 463.312716525

- Number of Pixels per Tile = 2400

- Projection Information
  PROJ4: '+proj=sinu +a=6371007.181 +b=6371007.181 +units=m'
  WKT:

  ```
  PROJCS["Sinusoidal", GEOGCS["GCS_unnamed ellipse",
  DATUM["D_unknown", SPHEROID["Unknown",6371007.181,0]],
  PRIMEM["Greenwich",0], UNIT["Degree",0.017453292519943295]],
  PROJECTION["Sinusoidal"], PARAMETER["central_meridian",0],
  PARAMETER["false_easting",0], PARAMETER["false_northing",0],UNIT["Meter",1]
  ```

3.1   Accessing MODIS Data Products

Several ways to access the MODIS data products are listed below. More info about the data sets, data formats, and quality information are available from the Land Processes DAAC. For MCD12Q2 the link is `https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mcd12q2_v006`

- Bulk download: LP DAAC Data Pool and DAAC2Disk.

- Search and browse: USGS EarthExplorer and NASA Earthdata Search.

3.2 Known Issues and Sources of Uncertainty

- *Greenup* and *Dormancy* phenometrics are anomalously early/late in some high-latitude regions with low-amplitude NBAR-EVI2 variation. This likely results from the splining approach assuming a dormant period NBAR-EVI2 value that is too low. The cause and a fix are being investigated, but until a fix is devised, users are encouraged to use the more realistic and stable *MidGreenup* and *MidGreendown* metrics to capture season start/end in these regions (e.g. high-latitude evergreen forests).

- In some semi-arid and arid environments exhibiting low-amplitude NBAR-EVI2 variation, there are large differences in filled pixels between the Collection 5 and Collection 6 products. Collection 6 typically retrieves fewer pixels because it applies a global minimum NBAR-EVI2 amplitude requirement of 0.1 in addition to the relative amplitude requirements in Collection 5 (valid segments must have NBAR-EVI2 amplitude that is at least 35% of the 3-year max - min.)

- For pixels with more than two valid vegetation cycles, the two cycles that are documented in the product may not be temporally adjacent. Instead, they are the two cycles with the largest NBAR-EVI2 amplitudes. That is, the two cycles may have a temporally intervening cycle of lower amplitude.

- Because heterogeneous land cover/use results in NBAR-EVI2 time series that are a mixture of perhaps distinct phenologies, users should be careful in interpreting multi-cycle results. They are not, for example, guaranteed to represent two distinct cropping cycles, but may instead capture two different sub-pixel scale fields with very different sow/harvest dates.

- Users may notice substantial differences in patterns of missing data between Collections 5 and 6 in the Southern Hemisphere. This is mostly a result of the difference in methodology whereby phenometrics are assigned to particular product years. Collection 5 phenometrics are delivered with the product year associated with the year of the particular phenometric, and so the phenometrics associated with a particular vegetation cycle may be spread across multiple product years. In Collection 6, the phenometrics for a particular cycle are always kept together, and delivered with the product year associated with the calendar year in which the *Peak* phenometric falls.

4 Contact Information

User Contact:

- Josh Gray (jmgray2@ncsu.edu)

- Mark Friedl (friedl@bu.edu)

5 Science Data Sets

Table 1 describes the Science Data Sets contained in the MCD12Q2 Collection 6 product. All SDS have a 16-bit integer data type. All except "NumCycles" contain two data layers corresponding to up to two valid vegetation cycles. Overall QA can be interpreted directly as an integer, but Detailed QA is bit-packed and so must be unpacked for proper interpretation (details and examples below).

The QA_Overall SDS provides a simple representation of the quality of retrievals for a particular vegetation cycle. As previously described, it is calculated from a weighted combination of the fraction of missing

Table 1: MCD12Q2 Science Data Sets.

SDS Name	Description	Layers	Data Type	Units	Valid Range	Fill Value	Scale Factor
NumCycles	Total number of valid vegetation cycles with peak in product year	1	INT16	none	1–7	32767	1
Greenup	Date when EVI2 first crossed 15% of the segment EVI2 amplitude	2	INT16	days since 1-1-1970	11138–32766	32767	1
MidGreenup	Date when EVI2 first crossed 50% of the segment EVI2 amplitude	2	INT16	days since 1-1-1970	11138–32766	32767	1
Maturity	Date when EVI2 first crossed 90% of the segment EVI2 amplitude	2	INT16	days since 1-1-1970	11138–32766	32767	1
Peak	Date when EVI2 reached the segment maximum	2	INT16	days since 1-1-1970	11138–32766	32767	1
Senescence	Date when EVI2 last crossed 90% of the segment EVI2 amplitude	2	INT16	days since 1-1-1970	11138–32766	32767	1
MidGreendown	Date when EVI2 last crossed 50% of the segment EVI2 amplitude	2	INT16	days since 1-1-1970	11138–32766	32767	1
Dormancy	Date when EVI2 last crossed 15% of the segment EVI2 amplitude	2	INT16	days since 1-1-1970	11138–32766	32767	1
EVI_Minimum	Segment minimum EVI2 value	2	INT16	NBAR-EVI2	0–10000	32767	0.0001
EVI_Amplitude	Segment maximum - minimum EVI2	2	INT16	NBAR-EVI2	0–10000	32767	0.0001
EVI_Area	Sum of daily interpolated EVI2 from Greenup to Dormancy	2	INT16	NBAR-EVI2	0–3700	32767	0.1
QA_Overall	QA code for entire segment	2	INT16	none	0–3	32767	1
QA_Detailed	Bit-packed, SDS-specific QA codes	2	INT16	none	0–16383	32767	1

or filled NBAR-EVI2 data in the cycle, and the goodness of fit of the smoothing spline. Values are in the range 0–3 corresponding to "best", "good", "fair", and "poor". The QA_Detailed SDS contains bit-packed QA information for each date phenometric (e.g. Greenup, MidGreenup, etc.). Individual QA values have an identical interpretation as QA_Overall. However, accessing these scores requires the user to convert the 16-bit integer value to binary, and then retrieve specific bit pairs that may then be converted to an integer in the range 0–3. Figure 2 provides a visual example of this conversion, and Table 2 provides some examples of arbitrary QA scores and their 16-bit integer value to aid users in verifying their interpretation.

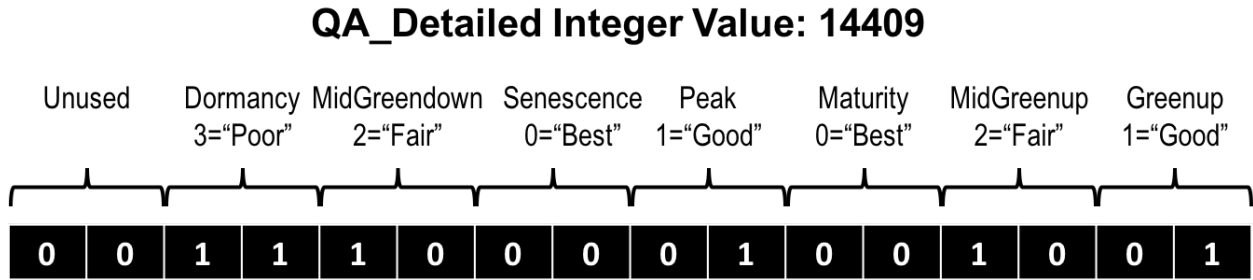## QA_Detailed Integer Value: 14409

Figure 2: An example QA_Detailed 16-bit integer value, its binary representation, and associated phenometric-specific bit pair interpretation.

Table 2: Detailed QA example values

Greenup	MidGreenup	Maturity	Peak	Senescence	MidGreendown	Dormancy	Value
0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	5461
3	2	1	1	2	3	3	15963
1	2	0	1	0	2	3	14409
3	3	3	3	3	3	3	16383

Here is an example R function that converts raw values of QA_Detailed to a vector of 0–3 quality scores in the order: Greenup, MidGreenup, Maturity, Peak, Senescence, MidGreendown, Dormancy:

```
UnpackDetailedQA <- function(x){
  bits <- as.integer(intToBits(x))
  quals <- sapply(seq(1, 16, by=2), function(i) sum(bits[i:(i+1)] * 2^c(0, 1)))[1:7]
  return(quals)
}
```