

## User Guide for MOD44A Vegetative Cover Conversion (VCC)

### **Introduction:**

Changes in land cover can be seen in remotely sensed data as changes in spectral response of the land cover. These changes can be characterized visually by looking at a series of images or in an automated way through the use of computer programs and models. The version 004 Vegetative Cover Conversion (VCC) product is designed to be a global alarm product for these land cover conversions. Land cover conversion is defined here as a cumulative effect of human and natural events over time, not including changes related to seasonal or phenological changes in vegetation.

**It is important to note that this product is not intended to be an areal estimate for land cover change.** This means that we expect to correctly identify that change has occurred in an area, but do not necessarily expect to quantify the exact amount of change that has occurred in that area. The intention is for further investigation to be performed with higher resolution data such as Landsat or ASTER. The VCC product is designed to be conservative, in an effort to minimize errors of commission (areas flagged as having changed which have not actually changed)

The VCC product is distributed in HDF-EOS file format, in which most users will find the Science Data Set (SDS) “Labeled Land Cover Change” is the most pertinent layer. This SDS contains the information showing what areas were identified to have changed or not during the processing period. There is also a Quality Assurance (QA) layer associated with this called “Labeled Land Cover Change QA” giving the user a confidence level at which the change was detected.

### **Inputs:**

VCC uses an intermediate composite product derived from 250-meter resolution 16-day composites from the MODIS instrument aboard the Terra satellite. These 16-day composites are generated from daily, level 2G, surface reflectance data (MOD09), and are composited based on the quality of the daily observations as determined by the QA flags in that product. The criteria used to generate the intermediate composite are:

- Cloud free
- Low aerosol contamination
- Cloud shadow free
- Low cirrus contamination
- Near nadir (nadir is the point at which the instrument is “looking” straight down)

The QA flags associated with the observation that is selected are retained in a separate layer for use later.

The compositing process identifies the observation, based on the quality of bands 1 and 2 natively at 250-meter resolution, that best represents the 16-day period. The compositing algorithm then utilizes the sub-pixel geolocation information from the daily L2G observation pointer files (MODPT) to select the appropriate observation from the daily bands 3 – 7, to correspond with the observation selected for the composite. This allows us to retain bands 3 – 7, which are natively at 500-meter resolution, in the same 250-meter grid as bands 1 and 2. The new data layers do not simply have duplicated 500-

meter values in a 250-meter grid; they contain the most relevant information based on the 250-meter compositing process. This composited dataset is an internal product called MOD44C or the intermediate data collector.

### **Training Data:**

In order to generate the VCC product, it is necessary to have a training data set to use in the creation of the decision tree models. Using MOD44C a sample of the spectral values is generated. This is accomplished by sampling every tenth pixel of every tenth line into an output file, retaining all 7 bands, calculated NDVI, and the quality flags. The NDVI used here is calculated from bands 1 and 2 using the formula  $(\text{Band2} - \text{Band1}) / (\text{Band 2} + \text{Band 1})$ , rather than using an existing NDVI product. This sampled data set is stitched together to yield a global data set with representative spectral values. This process is repeated for each 16-day composite period until there is at least 1 year of data. In the current version, 2 full years of sampled data are used in the training data set.

The MODIS Vegetation Continuous Fields (VCF) Percent Tree Cover product (Hansen, 2003) is used to determine generalized “classes” of tree cover. (this will be explained in more detail in the algorithm section) These classes are incorporated into the training data so that spectral values can be sampled by class. The quality flags are then used to eliminate the poor quality pixels from each time period. A representative sample of 20,000 pixels is then taken on a per class basis for each time period. This data is then used to generate a decision tree model for each time period.

### **Decision Trees:**

The decision tree models are generated in S-plus statistical software. In order to limit the contamination of the model by seasonal effects, a different tree is created for each month (two 16-day periods) of the year. In addition, the data are stratified by latitude into different zones:

10° N to 30° N

10° N to 10° S

10° S to 30° S

In future releases stratification for north of 30° N and south of 30° S will be addressed.

### **Algorithm:**

The VCC algorithm has evolved over time. The original algorithm employed several different types of methods and combined them at the end to improve accuracy and confidence in the product. A description of the original method can be seen in Zhan 2002. Through the evolution it has been determined that the space partitioning method showed the best results. (Zhan 2002) In this version of the algorithm we focus on the space partitioning method, which is performed by the use of decision tree models to define physical phenomena by location within spectral space.

In order to partition the spectral space, the phenomena under investigation must first be defined. As referenced above, the MODIS VCF Percent Tree Cover product is used to identify classes of forested and non-forested areas. Initial attempts were

performed aggregating “classes” that are representative of the type of ground cover based on the IGBP definition of a closed canopy as 60% tree cover:

Scheme 1

60 – 100%	Tree Cover
40 – 59%	Tree Cover
20 – 39%	Tree Cover
0 – 19%	Tree Cover

The results from the initial method, described above, and the results from the following method:

Scheme 2

73 – 100%	Tree Cover
55 – 72%	Tree Cover
37 – 54%	Tree Cover
19 – 36%	Tree Cover
0 – 18%	Tree Cover

were combined and the results compared. The combined method yielded improved results over any single method and retained the spirit of the original VCC product.

With the classes defined, the process by which change is detected can be addressed. On the very basic level this product takes two inputs from the same area, from the same 16-day time period, one year apart; performs a classification on each image (excluding any poor quality and cloudy data) and then does a difference between those two images. A change is recorded if:

- Scheme 1
  - Time 1 is classified as 60 – 100% tree cover and time 2 is classified as 0 – 19% or 20 – 39% tree cover.
  - Time 1 is classified as 40 – 59% tree cover and time 2 is classified as 0 – 19% tree cover.
- Scheme 2
  - Time 1 is classified as 73 – 100% tree cover and time 2 is classified as 0 – 18%, 19 – 36%, or 37 – 54% tree cover.
  - Time 1 is classified as 55 – 72% tree cover and time 2 is classified as 0 – 18%, or 19 – 36% tree cover.
  - Time 1 is classified as 37 – 54% tree cover and time 2 is classified as 0 – 18% tree cover.

In order to avoid confusion in areas where the forest dynamic changed slightly from one year to the next a full class is skipped before change is indicated. This allows flexibility for inter-annual variability in the tree cover. This is necessary to accommodate differences in rainfall and other possible environmental factors.

**Outputs:**

Outputs are generated four times per year (quarterly). The most important layer for the user is the second Science Data Set (SDS) which is called the “Labeled Land Cover Change” layer. This is an image layer that shows the pixels that have been identified as either changed or not changed. See the table in Appendix B for the description of values. For any pixel where change was identified in more than 1 16-day time period by either of the 2 methods it is considered to have changed and is indicated in the “Labeled Land Cover Change” layer. Quality of the input data is used to define the confidence with which the change was detected and is recorded in the “Labeled Land Cover Change QA layer”. Appendix C shows the range of valid values and their interpretations for the QA layer.

For internal use and diagnostic purposes information from individual 16-day periods is retained. This information helps developers understand how the algorithm is performing during intermediate steps. Appendix A contains a table that shows the interpretation of the DN’s (Digital Number) shown on the screen and how that relates to the original 16-day time periods. The value shown when viewing any of the 12 planes in the first SDS, which are called metrics planes, can be referenced in the table in Appendix A to determine the original intermediate classification. The table’s vertical axis shows the “class” from time 1 and the horizontal axis shows the “class” from time 2. These correspond to the “classes” described in the algorithm section. The 12 planes represent the 6 time periods for each of the two schema; planes 1 – 6 correspond to Scheme 2 and planes 7 – 12 correspond to Scheme 1. The fourth production quarter has only five 16 – day time periods (due to the number of days in the year not being evenly divisible). This means that planes 6 and 12 are empty for the last time period.

The metrics planes could be used in two ways. The metrics allow the user to determine when a change that is represented in the Labeled Land Cover Change layer was first detected and also allow the user to determine which scheme identified the change.

### **Known Issues:**

There are a number of items that can manifest as errors in this product. See the list below for the known issues with the product. Any of these items can create a false positive result in our product.

- Clouds that are not flagged in the QA layer of the input data and occur over forested regions in time 2 will invariably cause a change to be detected even though there was no change.
- Inter-annual variability in vegetation can cause issues. Specifically:
  - Deciduous forests can lose their leaves at slightly different rates from year to year. If they hold their leaves particularly long in 1 year and drop them early the next year it could appear as a change.
  - Understory in open canopy forests can change dramatically from year to year. This can also cause a false change to be detected.

Reference:

Hansen, M. C., DeFries, R. S., Townshend, J. R. G., Carroll, M., Dimiceli, C., Sohlberg, R. A.. 2003: **Global Percent Tree Cover at a Spatial Resolution of 500 Meters: First Results of the MODIS Vegetation Continuous Fields Algorithm.** *Earth Interactions*: Vol. 7, No. 10, pp. 1–15.

Zhan, X., Sohlberg, R., Townshend, J.R.G., DiMiceli, C., Carroll, M., Eastman, J.C., Hansen, M., DeFries, R. S.. 2002: **Detection of land cover changes using MODIS 250-meter data.** *Remote Sensing of Environment*, Vol. 83, pp. 336-350.

### Appendix A (Metrics legend)

Class Time Period 1	Class Time Period 2								
	0	1	2	3	4	5	10	110	255
0	0	1	2	3	4	5	10	110	255
1	256	257	258	259	260	261	266	366	511
2	512	513	514	515	516	517	522	622	767
3	768	769	770	771	772	773	778	878	1023
4	1024	1025	1026	1027	1028	1029	1034	1134	1279
5	1280	1281	1282	1283	1284	1285	1290	1390	1535
10	2560	2561	2562	2563	2564	2565	2570	2670	2815
110	28160	28161	28162	28163	28164	28165	28170	28270	28415
255	65280	65281	65282	65283	65284	65285	65290	65390	65535

### Appendix B (Labeled Land Cover Change)

Dn	Interpretation
0	no change
1	change
255	fill

### Appendix C (Labeled Land Cover Change QA)

Dn	Interpretation
0	High Confidence change
1	High Confidence no change
22	Undetermined due to bad data or initial condition was non-forest
2	Undetermined other
10	Moderate confidence change
11	Moderate confidence no change
12	Undetermined, high sensor zenith in inputs (>=45)
32	Undetermined due to bad data or initial condition was non-forest