Supplement of Earth Syst. Sci. Data, 14, 143–162, 2022
https://doi.org/10.5194/essd-14-143-2022-supplement

Supplement of

Implementation of the CCDC algorithm to produce the LCMAP Collection 1.0 annual land surface change product

George Z. Xian et al.

Correspondence to: George Z. Xian (xian@usgs.gov)

Supplement

1 Data filtering and model fitting

The removal of invalid and cloud-contaminated data points is important for deriving model coefficients that accurately represent the phenology of the surface, and for the correct identification of model break points. After removing invalid data using the pixel quality assessment (QA), an additional procedure is needed to filter pixels with some degree of atmospheric obscuration. The model fitting is performed within the standard procedure to remove additional outliers by using the multitemporal observation record as a further aid in identifying values that deviate from the overall phenology curve using a specific harmonic model to perform an initial fit to the phenology. Model initialization uses an adaptation of the Tmask algorithm (Zhu and Woodcock, 2014) to identify and exclude outliers within the initialization window (where the "window" is the span of observations covered by the procedure). The algorithm uses a specific harmonic model to perform an initial fit to the phenology, defined below:

$$\hat{p} = c_0 + c_1 \cos \omega t + c_2 \sin \omega t + c_3 \cos \frac{\omega t}{N} + c_4 \sin \frac{\omega t}{N} + c_5 t \qquad \text{(S1)}$$

where ω is the base annual frequency as (2π⁄365.2425), N is the number of years of Landsat data. The coefficients c_1 and c_2 capture annual cycles, while the coefficients c_3 and c_4 are used to allow the model to respond to land cover change (Zhu and Woodcock, 2014). When moving through the time series, the model uses a threshold defined by the outlier threshold parameter to exclude values that deviate significantly from the established time series models. The default value of this threshold is derived from an inverse χ^2 distribution. After all input data are filtered, the time series is fit by harmonic models defined by Eq.1.

2 Change detection thresholds

The CCD algorithm checks for a model break (or spectral change) by determining best-fit harmonic model. Change detection sensitivity depends on the value of change threshold that is drawn from an inverse χ^2 distribution, with degrees of freedom equal to the number of used Landsat ARD spectral bands including the green, red, NIR, SWIR1, and SWIR2 bands. The $CHANGE\_THRESHOLD$ is determined in Equations S2 and S3.

$$CHANGE\_THRESHOLD = Inv\chi^2\left(1-(1-0.99)^{size}\right) \quad \text{(S2)}$$

where Invχ^2 is an inverse χ^2 distribution with five degrees of freedom. The *size* is determined by a Peek_Size (six by default) and $Peek\_Size_{adj}$ as size= Peek_Size/$Peek\_Size_{adj}$.

$$Peek\_Size_{adj} = round\left(Peek\_Size \times \frac{16}{median(T_{diff})}\right) \quad \text{(S3)}$$

where $T_{diff} = t_{n+1} - t_n$ is the set of successive differences between observation dates (t) for all $n \in (1, \ldots, N-1)$, where N is the total number of valid observations. If $Peek\_Size_{adj} < Peek\_Size$, then use the original Peek_Size. As the algorithm steps through the time series, it checks for a model break (spectral change) with respect to its last determined best fit harmonic model. Change detection sensitivity depends on the value of $CHANGE\_THRESHOLD$. While there may be at least twelve observations to initialize a new model after the standard procedure (Fig.2), additional procedures are implemented to try to initialize a new harmonic segment for continuous monitoring. The procedures use the green and SWIR bands to identify outliers if a residual exceeds the variability value using observations greater than 30 days apart. The next observations in the time series are needed if there are not at least twelve observations after excluding these outliers. These observations are used to create initial LASSO harmonic fits. A stability check is performed to ensure the initial regression fits are not over a potential landscape disturbance:

$$\sum_{i\in D} \frac{|c_{1,i}t_1 - c_{1,i}t_N| + |resid_{1,i}| + |resid_{N,i}|}{max(var_i, RMSE_i)} < CHANGE - THRESHOLD \quad \text{(S4)}$$

where c_1 is the spectral slope coefficient, t_1 and t_N are the first and last dates for observations (1,⋯,N) covered by the regression fit, $resid_{1,i}$ and $resid_{N,i}$ are the first and last residuals, $RMSE_i$ is the per-band root-mean-square error over the window for each spectral band $i \in D$, where D are the DETECTION_BANDS. If the stability check fails, the initialization needs to start again by adding addition data from the time series. Otherwise, additional checks are performed after refit LASSO regressions.

3 Permanent snow and insufficient clear observations

The procedure for permanent snow suggests that too few clear or water observations exist to robustly detect change, and a large fraction are snow. The algorithm will return at most one segment, fit through the entire time series, provided the filtered observations number at least twelve. Unlike other procedures, snow pixels are not filtered out and are fit as part of the annual pattern. This avoids overfitting the model to a seasonally sparse observation record. The LASSO regression is used to fit for the entire time series. The window of observations covered by the model is used to report the start, end, and observation count. The change Boolean value is set to 0 and the break date is recorded as the last observation date. The magnitude of change is recorded as zero for each band.

The insufficient clear observation procedure is selected when there are too few clear or water observations to robustly detect change. The algorithm will return at most one segment, fit through the entire time series, provided the filtered observations number at least twelve. In addition to excluding snow observations, this procedure differs from the Permanent Snow procedure by the inclusion of an additional filtering step on the GREEN band, designed to further reduce cloud/outlier data points that may negatively affect a model fit through a sparse time series. The LASSO regression is implemented to fit for the entire time series and set the curve QA to 44. Records include the model coefficients, RMSE from the regression, the start and end dates, and observation count. The change Boolean value is set to 0 and the break day recorded as the last observation date.

If fewer than twelve (MEOW_SIZE) observations remain after filtering, no model fit is implemented with zero segment for both procedures of permanent snow and insufficient clear observation.

4 The LCMAP science products

The LCMAP science products consist of five annual land cover and five annual land surface change products that are produced at an annual time step. Land cover products represent the annual status of each pixel on July 1st, a representative date of each year. Land surface change products provide additional information about spectral and temporal change (e.g. magnitude or date) that occurred during an annual period. LCMAP science products are available for CONUS from 1985–2017. The land cover products contain Primary Land Cover (LCPRI), Primary Land Cover Confidence (LCPCONF), Secondary Land Cover (LCSEC), Secondary Land Cover Confidence (LCSCONF), and Annual Land Cover Change (LCACHG). The land surface change products include Time of Spectral Change (SCTIME), Change Magnitude (SCMAG), Spectral

Stability Period (SCSTAB), Time Since Last Change (SCLAST), and Spectral Model Quality (SCMQA).

Fig. S2 from surface reflectance shortwave infrared (SWIR) band time series clarifies the methods used to produce LCMAP Science Products. For simplicity, harmonic regressions are shown without the original pixel observations.

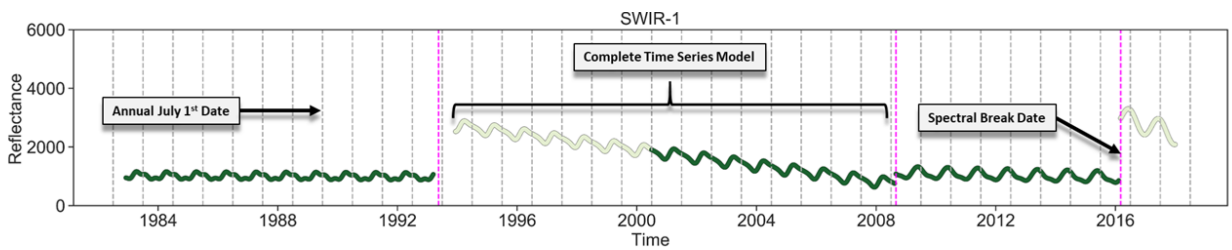


Figure S1. Legend for product definition examples, where green indicates time classified as tree cover, tan indicates time classified as grass/shrub, gray dashed lines are annual July 1st dates, and magenta dashed lines are spectral breaks.

4.1 Time of spectral change

Time of spectral change (SCTIME) represents the timing of a spectral change within the current product year as the day-of-year (DOY) the change occurred (Fig. S2). A spectral change is defined as a "break" in a CCDC time series model where spectral observations have diverged from the model predictions. These breaks may be indicative of a change in thematic land cover (e.g., fire, urbanization) or may represent more subtle conditional surface changes (e.g., forest growth, insect infestation). A value of zero indicates there was no recorded model break in the current year.

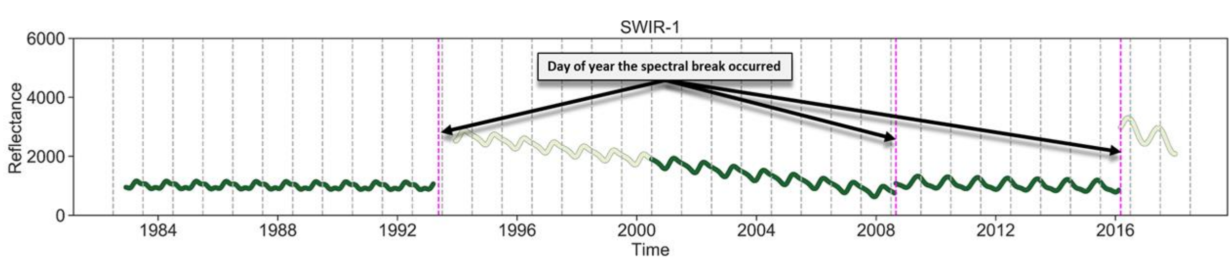


Figure S2. The DOY a spectral change caused a “break” in a CCDC time series model. When a spectral change occurs, causing a divergence from model predictions, a new time series model begins.

4.2 Change magnitude

Change Magnitude (SCMAG) provides information on the spectral strength or intensity of a time series model “break” when spectral observations have diverged from the model predictions (Fig. S3). Change Magnitude is calculated as the square root of the sum of the squared per-band median residuals (excluding the blue and thermal bands) between the observed per-band Landsat surface reflectance (scaled) and CCDC predictions at the time of a detected CCDC model break. SCMAG is unitless and generally ranges between 1–10,000. A value of zero indicates there is no recorded model break in the current year.

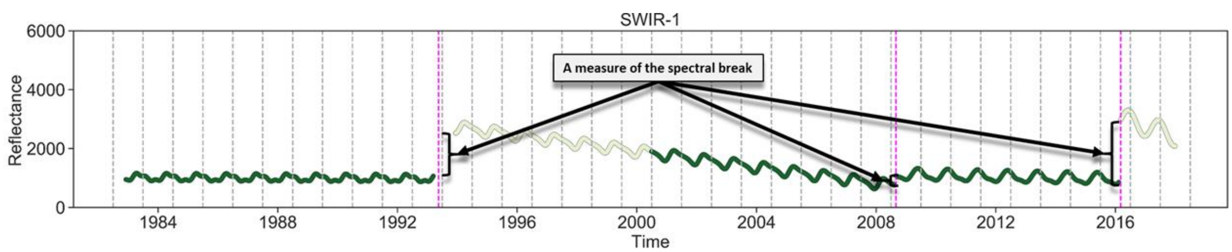


Figure S3. Change magnitude when a spectral break occurs.

4.3 Spectral stability period

Spectral stability period (SCSTAB) is a measure of the amount of time in days a pixel has been in its current spectral state. Spectral state can refer to a pixel’s state during a time series model (Fig. S4) and the temporal ranges between time series models.

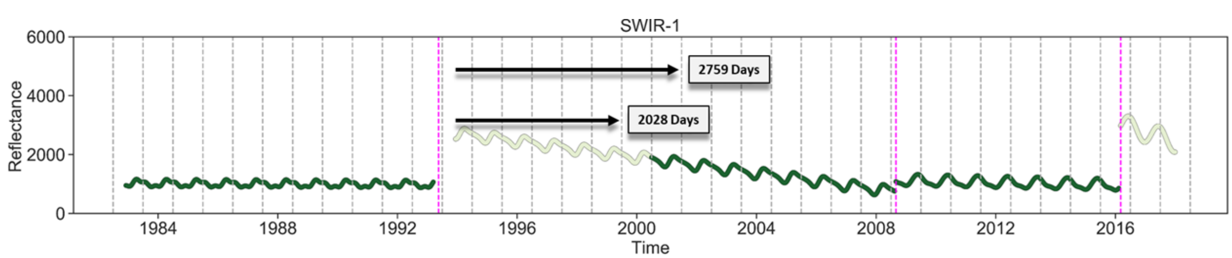


Figure S4. Spectral stability period during a time series model starting on the July 1st start date for the product year.

4.4 Time since last change

Time since last change (SCLAST) represents the time, in days, from July 1st of the current product year back to the most recent time series model "break" where spectral observations diverged from CCDC model predictions. This can also be expressed as the time, in days, since the last recorded result in both SCTIME and SCMAG.

4.5 Spectral model quality

Spectral model quality (SCMQA) provides additional information regarding the type of time series model available in the current product year. SCMQA reflects the type of time series model present on July 1st of the current year and can be useful for interpreting results in other LCMAP science products. SCMQA pixel values and descriptions are provided in Table S1.

Table S1 Description of pixel values for spectral model quality

Pixel Value	Model Type	Description
0	No Model	No model established for July 1st of current year.
4	Simple Model	A partial, 4-coefficient harmonic model.
6	Advanced Model	A partial, 6-coefficient harmonic model.
8	Full Model	A full, 8-coefficient harmonic model.
14	Start Fit	A simple model at the beginning of a time series where sparse and/or highly variable spectral measurements prevent establishment of a harmonic model.
24	End Fit	A simple model at the end of a time series where there are insufficient observations and/or time to establish a new harmonic model following a model break.
44	Insufficient Clear	A simple model for the entire time series in cases where fewer than 25% of input observations are labeled as "Clear" or "Water" by the Landsat ARD per-pixel quality band (PIXELQA).
54	Persistent Snow	A simple model for the entire time series in cases where 75% or more of input observations are labeled as "Snow" by the Landsat ARD per-pixel quality band (PIXELQA).

4.6 Land cover and land cover confidence

Primary and secondary land cover classifications and the associated primary and secondary confidence values are determined through a combination of three approaches: initial classifier, secondary analysis, and rule-based assignment. Both primary and secondary land cover contain values of 1-8, representing the eight different land cover classes defined in Table 1.

Example plots of a time series showing primary land cover (LCPRI), secondary land cover (LCSEC), primary land cover confidence (LCPCONF), and secondary land cover confidence (LCSCONF) are provided in Fig. S5-8. Fig. S5 shows that the time series contains three segments (first, third, and fourth) classified across multiple years using the initial classifier, one segment (second) classified as a Grass/Shrub to Tree Cover transition by secondary analysis (Fig. S6), and several individual years (1982, 1993, and 2018) whose July 1st date does not intersect with a time series model. These values are classified by rule-based assignment. The primary and secondary land cover are assigned according to their confidences (Fig. S7-8).

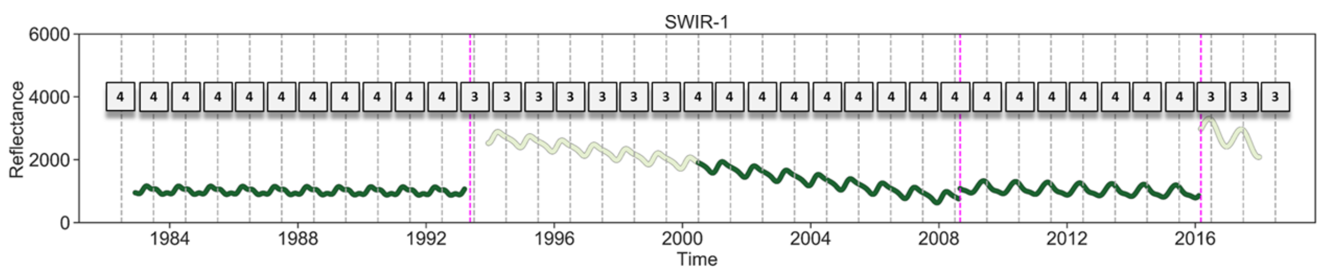


Figure S5 Primary land cover values for the example time series: tree cover (4) and grass/shrub (3).

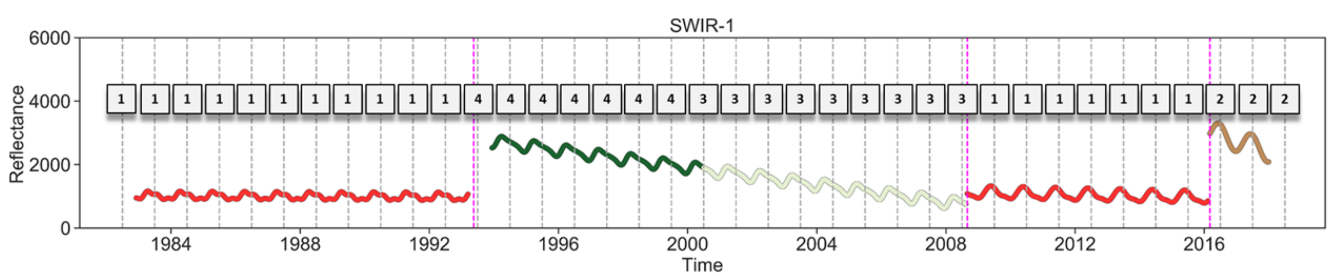


Figure S6 Secondary land cover values for the example time series: developed (1), tree cover (4), grass/shrub (3), and cropland (2).

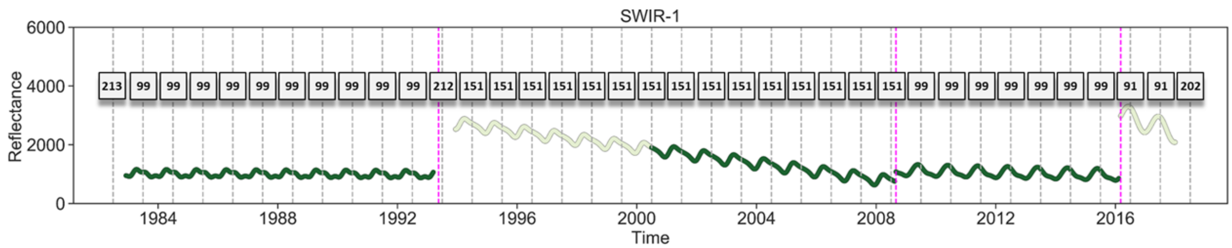


Figure S7 Primary land cover confidence values for the example time series.

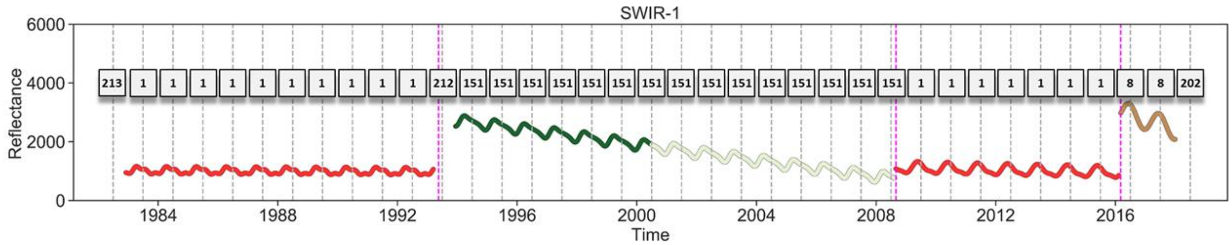


Figure S8 Secondary land cover confidence values for the example time series.

If the assigned land cover value is derived directly from the initial classifier, confidence values will range from 1–100. If from secondary analysis of trends, values will be either 151 or 152. Finally, rule-based assignments will be values of 201–214. Only values from 1–100 represent a continuous confidence score (low confidence to high confidence); values greater than 100 are categorical indicators. Full description of the confidence value is presented in Table S2.

4.6.1 Initial classifier

Land cover is labeled by calculating a per-class mean across all the predictions for a time series model. The land cover classes with the highest and second highest confidence values become the primary and secondary land cover, while the associated confidence (probability) values are used to determine primary and secondary confidence. The 0.0 to 1.0 floating point mean confidence value is then scaled by 100 and converted to an integer for the LCPCONF and LCSCONF product values.

Table S2 Description of pixel values for primary and secondary land cover confidence

Pixel Value	LC Label Source	Description
1-100	Initial classifier	Measure of confidence that the Primary Land Cover label matches the training data.
151	Secondary analysis	Time series model identified as transition from a Grass/Shrub class to a Tree Cover class. Primary Land Cover class assignment based on secondary analysis.
152	Secondary analysis	Time series model identified as transition from a Tree Cover class to a Grass/Shrub class. Primary Land Cover class assignment based on secondary analysis.
201	Rule-based	No stable time series models were produced for this location. Primary Land Cover assigned the land cover class present in NLCD 2001 (cross-walked to LCMAP Level-1 classification schema*)*.
202	Rule-based	Insufficient data available to extend most recent time series model past July 1st of current year. Primary Land Cover assigned the last identified cover class from earlier year.
211	Rule-based	July 1st falls in a gap between two stable time series models of the same land cover class. Primary Land Cover assigned the land cover class of those before/after models.
212	Rule-based	July 1st falls in a gap between two stable time series models of differing land cover class. If July 1st is before the "break date" of the earlier model, Primary Land Cover is assigned the land cover class of that earlier model. Otherwise, Primary Land Cover is assigned the land cover class of the subsequent, later model.
213	Rule-based	Insufficient data available to establish a stable time series model at the beginning of the time series prior to July 1st of the current year. Primary Land Cover assigned the land cover class of 1st subsequent model.
214	Rule-based	Insufficient data available to establish a new stable time series model following a break near the end of the time series prior to July 1st of the current year. Primary Land Cover assigned the last identified cover class from earlier year.

4.6.2 Secondary analysis

Transitions from Grass/Shrub to Tree Cover or from Tree Cover to Grass/Shrub are handled by looking at the first and last set of predictions for the time series model, along with the average reflectance values at the beginning and end of the model for a specific combination of spectral bands. Specifically, we define a normalized band ratio at ordinal date t using the average reflectance values for the NIR and SWIR-1 bands. The difference of the normalized band ratio between start and end dates is calculated as ΔBR. A segment is considered a Grass/Shrub to Tree Cover transition segment if the first annual prediction from the classifier indicates Grass/Shrub, the last indicates Tree Cover, and $\Delta BR > 0.05$. In these cases, all dates for the time series model before the first annual prediction indicating Tree Cover are assigned to Grass/Shrub, and that date and all subsequent dates covered by the model are assigned to Tree Cover. For these segments, land cover confidence (both primary and secondary) is assigned the value 151. For Tree Cover to Grass/Shrub transitions, the reverse set of conditions apply: the first annual prediction must indicate Tree Cover, the last Grass/Shrub, and we require $\Delta BR < 0.05$. For these cases, the primary and secondary land cover confidence is set to 152.

4.6.3 Rule-based assignment

For rule-based assignment, land cover is determined by interpolation or extrapolation of other information available in the time series. This is used when trying to determine land cover that is not immediately attributable to a time series model. This happens either between time series models, at the ends of the time series, or if there are not enough observations to build a time series model.

The following describes how different scenarios are handled for both primary and secondary confidences.

- If there is no stable time series model available for the entire time series (i.e., in cases with few observations), cross-walked NLCD 2001 is used for both primary and secondary land cover for all dates and value 201 assigned to land cover confidence.

- If the model break date falls before all time series model start dates (e.g. July 1st, 1982), the primary and secondary land cover class is assigned from the subsequent time series model with value 213 to land cover confidence.

- If the model break date falls after all time series model end dates (e.g. July 1st, 2017), and the final time series model did not end with a spectral break, the last identified primary and secondary land cover class is assigned value 202 to land cover confidence.

- If the model break date falls after all time series model end dates, and the final time series model did end with a spectral break, the last identified primary and secondary land cover class is assigned value 214 to land cover confidence.

If the model break date falls in a gap between two time series models, then for primary and secondary land cover is assigned *separately*:

- If the surrounding time series models indicate the same land cover, the common land cover value is assigned to primary land cover and value 211 to land cover confidence.
- If the surrounding time series models indicate differing land cover, the break date between the two time series models is used to interpolate the land cover values by assigning the previous land cover value before the break date to primary land cover. At or after the break date, primary land cover is assigned the subsequent land cover class and value 212 is assigned to land cover confidence.

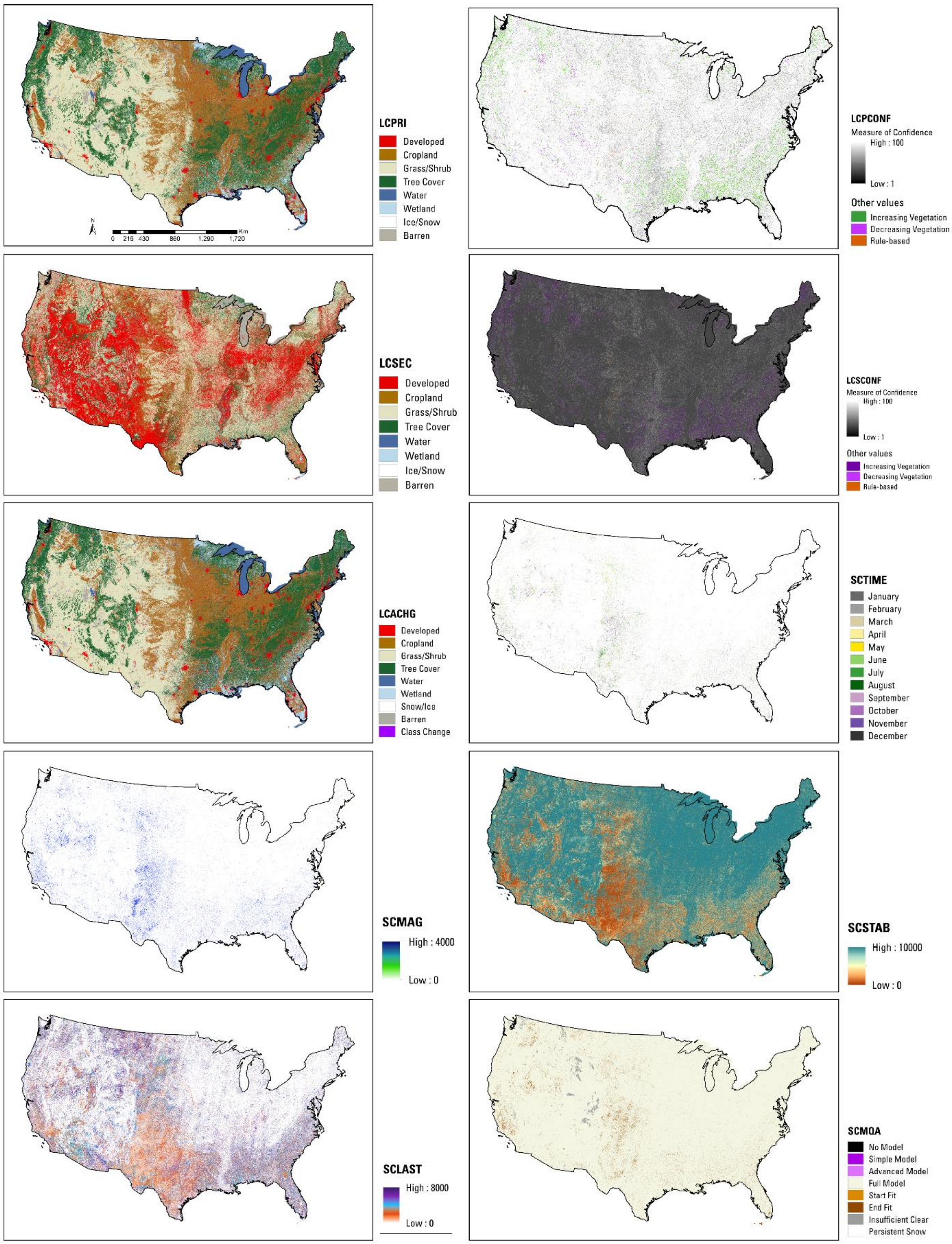
4.7 Annual land cover change

Annual Land Cover Change is a synthesis product derived from the LCPRI of the current product year and the LCPRI of the previous year. Values 1–8 correspond to the integer values representing land cover classes presented in Table 1 and are assigned to pixels if no change was identified between the two years. If change between years was identified, the resulting two-digit pixel value is a concatenation of the previous and current land cover class code (Table S3). For example, a pixel value of 18 is a concatenation of a one (1), representing the Developed class, and an eight (8), representing the Barren class. Therefore, the resulting value of 18 represents a change from Developed to Barren. Table S3 provides a partial list of these concatenated change codes as examples along with the land cover changes they represent.

Table S3 Description of pixel values for annual land cover change

Pixel Value	Previous Land Cover Class	Current Land Cover Class	Land Cover Change
1-8	*(as in Table 1)*	*(as in Table 1)*	No change
18	Developed	Barren	Developed to Barren
21	Cropland	Developed	Cropland to Developed
32	Grass/Shrub	Cropland	Grass/Shrub to Cropland
42	Tree Cover	Cropland	Tree Cover to Cropland
43	Tree Cover	Grass/Shrub	Tree Cover to Grass/Shrub
65	Wetlands	Water	Wetlands to Water
73	Snow/Ice	Grass/Shrub	Snow/Ice to Grass/Shrub

The LCMAP science products for the conterminous United States in 2016 are displayed in Fig.9S.



N
0 215 430 860 1,290 1,720 Km
LCPRI
Developed
Cropland
Grass/Shrub
Tree Cover
Water
Wetland
Ice/Snow
Barren
LCPCONF
Measure of Confidence
High : 100
Low : 1
Other values
Increasing Vegetation
Decreasing Vegetation
Rule-based
LCSEC
Developed
Cropland
Grass/Shrub
Tree Cover
Water
Wetland
Ice/Snow
Barren
LCSCONF
Measure of Confidence
High : 100
Low : 1
Other values
Increasing Vegetation
Decreasing Vegetation
Rule-based
LCACHG
Developed
Cropland
Grass/Shrub
Tree Cover
Water
Wetland
Snow/Ice
Barren
Class Change
SCTIME
January
February
March
April
May
June
July
August
September
October
November
December
SCMAG
High : 4000
Low : 0
SCSTAB
High : 10000
Low : 0
SCLAST
High : 8000
Low : 0
SCMQA
No Model
Simple Model
Advanced Model
Full Model
Start Fit
End Fit
Insufficient Clear
Persistent Snow

Figure S9. The LCMAP 10 science products. From the left to right and the top to bottom are primary land cover (a), primary land cover confidence (b), secondary land cover (c), secondary land cover confidence (d), annual land cover change (e), time of spectral change (f), change magnitude (g), spectral stability period (h), time science last change (i), and spectral model quality (j).

Reference

Zhu, Z. and Woodcock, C. E.: Automated cloud, cloud shadow, and snow detection in multitemporal Landsat data: An algorithm designed specifically for monitoring land cover change, Remote Sensing of Environment, 152, 217-234, 2014.