OPERA DIST product

Matthew C. Hansen[1], Amy Pickens[1], and Zhen Song[1]

[1] University of Maryland, Department of Geographical Sciences, Global Land Analysis and Discovery (GLAD) laboratory

Corresponding Author(s): Amy Pickens (ahudson2@umd.edu)

Key Points:

Abstract

The earth's land surface is continually changing due to natural seasonal and interannual cycles, to direct human action, and to changing climate. Land disturbance outside of the natural cycle impacts habitats, climate, hydrology, food supply, and other critical systems. Improved understanding and monitoring of disturbances can provide tools to land managers as well as aid scientific investigation of processes. Here, we introduce a global system, DIST-ALERT, that flags all vegetation cover loss relative to its near-term historical range. With a median revisit rate of <2 days, Harmonized Landsat Sentinel-2 (HLS) imagery provides the highest cadence medium spatial resolution data set available and an ideal data source for monitoring the earth's surface. The vegetation fraction of each new HLS observation is estimated per pixel and compared to the last 3 years within a 31-day window surrounding the date of observation. Vegetation fractions less than the baseline minimum are flagged as disturbance and tracked through subsequent observations to build or decrease confidence. The system is agnostic to vegetation type and to resulting land cover, requiring further analysis for many applications. In addition to vegetation loss, there is a secondary disturbance detection algorithm based on spectral anomalies relative to the same baseline. This general disturbance detection is intended to capture all kinds of other land changes including crop extensification or lava flows. Along with DIST-ALERT, annual summaries of alerts are generated (DIST-ANN) to facilitate downstream science through the production of synoptic annual records of global land disturbance.

Plain Language Summary

Human activity alters the land surface as we seek to produce fuel, food, fiber, and dwelling space for expanding human populations and associated economic growth. Such changes impact the functioning of natural systems, sometimes replacing them wholesale through the expansion of cities, croplands, mines, and other land uses. Climate change may also impact the seasonal and interannual patterns of vegetation, altering the cycles of plant growth and decay. A better understanding of these changes will aid in the scientific study of their impacts on the Earth system and provide information for improved management of land resources. Satellite remote sensing provides a unique opportunity to monitor these changes at scale almost as they happen. This task is already being done for tropical forests. Here we extend the idea to the entire Earth land surface for all vegetation with a new system: DIST-ALERT. All kinds of vegetation loss are mapped and tracked through time to build or decrease confidence of change based on the severity and duration. Also, anything that looks different in the current observation compared to a baseline is marked and evaluated in the subsequent observations. The whole earth is imaged every 1-5 days and the maps are updated daily, though cloud cover blocks some regions from being analyzed as frequently. All of the disturbance alerts from a year are summarized within an annual product, DIST-ANN. Results will reflect the balance of direct and indirect human impacts in the pursuit of development with the maintenance of natural systems.

Version Description

This is the ATBD for DIST-ALERT and DIST-ANN v1 products released March, 2024.

1. Introduction

Land disturbance represents a host of dynamics that impact the earth system and may be due to human-induced or natural causes. For example, urbanization impacts local climate and hydrology; deforestation is a major source of carbon emissions; and drought inhibits food production. Information from earth observation data on land disturbance can help managers, agencies, and governments understand and respond to land changes in a timely manner. Here, we define disturbance as a loss of vegetation cover outside of near-term historical variability. Vegetation cover is mapped per pixel, evaluated against a seasonal baseline to detect vegetation disturbance, and then tracked through the time series. To account for land disturbances unrelated to vegetation loss, we also include a secondary algorithm that employs a spectral distance measure. Both disturbance algorithms are applied to near-real time HLS data and deliver low latency disturbance results DIST-ALERT and annual summary alert DIST-ANN. Stratified random samples are employed to validate DIST-ALERT and DIST-ANN. Overall accuracy for the disturbances ≥50% of DIST-ALERT is specified to exceed 80% and of DIST-ANN, 90%.

2. Context/Background

2.1. Historical Perspective

Transformation of the Earth's surface has increased over time, with impacts shifting from local to global scales, altering the fundamental flows of chemical and energy that sustain life on the only inhabited planet we know(Kates et al., 1990). Land disturbance due to human activity impacts a range of earth system functions, including climate regulation, hydrologic function, biodiversity richness, and more. For example, land use change accounts for 23% of total anthropogenic forcing of climate warming (Shukla et al., 2019). Direct human action in the form of land use change has accounted for roughly two-thirds of all observed land change over the last 30+ years (Song et al., 2018). Deforestation, agricultural intensification, urbanization and other dynamics reveal an increasing appropriation of natural lands for economic use and the increasing intensification of existing land uses (Foley et al., 2005). Considerable international policy effort has focused on slowing tropical deforestation in an attempt to reduce carbon emissions and limit damage to forest co-benefits such as the maintenance of terrestrial biodiversity, largely with little impact to date. Monitoring land change is a prerequisite to measuring its impacts, both in the policy and scientific domains.

Climate change itself is also a driver modifying land cover and land use. Tree lines are changing, mortality events increasing, droughts intensifying. Such land changes over time may become a larger fraction of the overall dynamic compared to other proximate drivers of disturbance, including tipping points resulting in large scale collapse of ecosystems. (Lenton et al., 2008) state that “Climate change and other human activities risk triggering biosphere tipping points across a range of ecosystems and scales" and that "Researchers need to improve their understanding of these observed changes in major ecosystems…" For both direct land use change and more distal climate-driven land change, data on land disturbance data can offer invaluable observational data and insights in support of scientific inquiries concerning global environmental change.

Disturbance is defined as any event that occurs outside the range of natural variability (Mildrexler et al., 2009) and may be due to human-induced or natural causes. Disturbance indicates an impact on land cover or land use that may result in a complete conversion or only a modification of the pre-existing condition. Disturbances can be instantaneous or long-lived, limited in area or regional in scale. Differentiating the limit of natural variability is a key requirement in assigning disturbance, as many ecosystems and associated land covers can be highly interannually variable, as are land uses. Additionally, many land covers and land uses are defined by disturbance, such as fire regimes in boreal forests, or rotations in forestry land uses.

Disturbance is often characterized in ecological terms as events that alter ecosystem extent, structure, communities, and other variables. Deforestation, mining, fire, drought and other dynamics, whether human-induced or natural in cause, result in a reduction in vegetation cover with concomitant impacts on ecosystem function. For this application, we focus on such dynamics and define disturbance as a loss of vegetation cover outside of near-term historical variability. In this way, we modify the general definition of disturbance to being a cover change of more to less vegetation cover. The presented method and associated disturbance algorithm is applied in near-real time as new imagery are available, delivering low latency results to facilitate land management decision-making. Annual summaries of alerts are then delivered, facilitating downstream science through the production of synoptic annual records of global land disturbance.

The potential uses of vegetation disturbance alerts at medium spatial resolution (10-30 m from globally acquired, publicly available sensing systems such as Landsat and Sentinel 2) range from enforcement to management applications. Monitoring road building, logging, forest clearing for agriculture and other dynamics can have added value if reported in near-real time. The DETER alerts of Brazil were critical to increasing the capacities of law enforcement and land management agencies in reducing illegal deforestation in the Brazilian Amazon (Nepstad et al., 2014). The deployment of such a system pan-tropically by Global Forest Watch through the University of Maryland's Global Land Analysis and Discovery (GLAD) lab, has offered such possibilities to other countries. Studies have showed GLAD alerts to have been used to reduce deforestation in community forests in Peru and Central Africa (Slough et al., 2021)**;** (Moffette et al., 2021). Here, we advance this approach by applying it at global scale with a continuous measure, characterizing generic vegetation loss instead of only forests, and employing the highest cadence medium spatial resolution (30 m) data set available, in the form of Harmonized Landsat Sentinel-2 data (HLS), as the input. The HLS tiling system is shown in ***Figure 1***. The integration of both medium spatial resolution systems greatly enhanced the temporal resolution (2-4 day repeat) of these data, which improves alert capabilities.

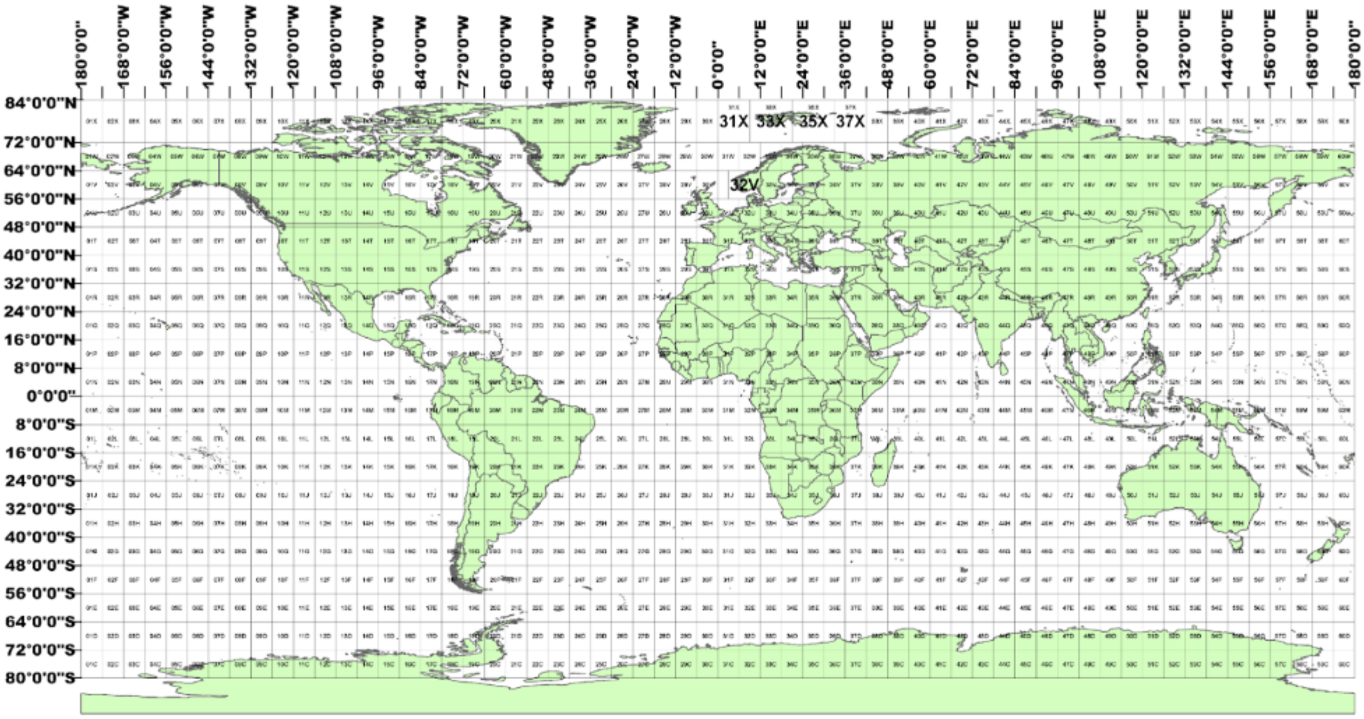

Figure 1: The global map of tile IDs for the HLS products (same as original Sentinel-2 tiling system).

3. Algorithm Description

3.1. Scientific Theory

Operational disturbance alert systems can be signal or land cover theme-based. Signal-based systems use a radiometric measure, such as greenness or brightness temperature, as the primary input to the alert system, delineating change outside normal variation of these bio-geophysical variables. By comparison, theme-based alerts characterize a specific land cover change dynamic over a time series, such as forest cover loss, i.e., the removal of tree cover, or flooding, i.e., an increase in the expanse of surface water beyond the norm. As such, land cover theme-based alert systems provide a more intuitive physical meaning and a resulting ability to map and validate area estimates more easily than bio-geophysical measures.

Fractional vegetative cover is a theme-based measure and the basis of the OPERA Surface Distburbance (DIST) algorithm. Fractional cover estimations from satellite data have a long history, employing a host of algorithms, from simple linear endmember mixture models (Adams et al., 1995); (Settle & Drake, 1993), multiple endmember mixture models (Roberts et al., 1998), empirical modeling (DeFries et al., 1997); (Zhu & Evans, 1994), and distribution-free machine learning methods (Hansen et al., 2002). The advantages of continuous dependent variables such as percent tree or vegetation cover include improved sensitivity to change compared to categorical labels, greater flexibility for users to adjust definitions, and more realistic depictions of ecotones.

Operational alerts of land change have been employed in a variety of modes, ranging from illegal deforestation monitoring in Brazil with the Real-Time System for Detection of Deforestation (DETER) (Shimabukuro et al., 2012), to active fire monitoring with the NASA's Fire Information for Resource Management System (FIRMS)(Davies et al., 2008), to food security with the Famine Early Warning System (FEWS) of USAID (Ross et al., 2009). Newer products include the use of medium spatial resolution data, for example Global Forest Watch's deforestation alerts (Hansen et al., 2016) made from Landsat and Sentinel-2 data. To advance this capability, we will implement a global low latency alert, DIST-ALERT, and an annual summary, DIST-ANN, product suite using NASA's Harmonized Landsat Sentinel-2 data (HLS) (Claverie et al., 2018) as inputs. The combined capability of these Earth observing systems results in a 2-4 day repeat visit cadence globally (Li & Roy, 2017), facilitating the application of near-real time disturbance mapping.

Vegetation fraction is a suitable variable for monitoring global land change. Our team has developed global algorithms for mapping per pixel percent vegetation cover using MODIS and Landsat data (Carroll et al., 2010); (Hansen et al., 2014); (Ying et al., 2017). Results with Landsat demonstrate the utility of the measure in mapping the dynamic of vegetation loss and employing the maps to sample-based and econometric methods to estimate land use outcomes/drivers and apply the measure as a leading economic indicator, respectively (Ying et al., 2017); (Ying et al., 2019). Vegetation loss as a generic dynamic can inform specific downstream applications from local to global scales and we will apply percent vegetation to HLS time-series imagery (Claverie et al., 2018) in mapping land disturbance.

3.1.1. Scientific Theory Assumptions

The first assumption in monitoring land disturbance, as we have defined it, concerns disturbances that involve vegetation loss. In terms of global environmental change, vegetation loss is a key indicator, whether the dynamic is deforestation, desertification, overgrazing, or fire. However, a limited number of disturbance dynamics do not involve vegetation loss, for example redevelopment of a commercial parcel, or lava flow superposed on old lava fields. An open question is the proportion of disturbance, as defined by generic surficial change events, that is omitted when targeting vegetation loss. It is the assumption that the vast majority of land disturbance relevant to policy, management and science applications will be observable using vegetation cover as the indicator variable. To confirm this assumption, we will add a Mahalanobis distance measure to delineate generic changes outside of the vegetation cover loss theme.

Another assumption concerns the ability of optical time-series data to discriminate relevant vegetation loss events accurately and in a timely fashion. Large conversion events, such as deforestation, have been shown to be reliably characterized (Hansen et al., 2013), while modification, or degradation of land cover types, has more mixed results. Conversions represent a high contrast, typically long-lived spectral change. Modifications represent low contrast, often ephemeral spectral change. The manner in which we plan to add signal for detecting modifications, in effect to improve contrast, is to exploit the density of the HLS time-series. Repeated alert detections, even if low in contrast individually, can in concert enable accurate assignment of low intensity land disturbance. As a safeguard, our product specification and definition of disturbance is for 50% or greater vegetative cover loss events, or more suited for conversion than modification.

3.2. Mathematical Theory

Vegetation Fraction Algorithm

We have been working on Vegetaion Continuous Field (VCF) maps for years,including MODIS VCF (Hansen et al., 2002), (Hansen et al., 2003) and Landsat-based VCF maps (Hansen et al., 2014); (Ying et al., 2017). In this product, we employ 8cm drone images and K Nearest Neighbor (KNN) model to characterize vegetation fraction. Our model relies on the following assumptions for successful estimation:

1. Consistent radiometric characterization of input imagery. The HLS time-series data feature state of the practice pre-processing including surface reflectance estimation and bi-directional reflectance distribution function correction, resulting in a reliable, scale set of independent variables for inputs into a turn-key KNN model.
2. Accurate quality assessment flags in screening inputs. The HLS data come with a quality assessment flag that must accurately screen unviable observations. No quality assessment layer is perfect, but too many omission errors in terms of passing haze/cloud/smoke/shadow-impacted observation leads to errors in mapping land change. However, the unprecedented density of the HLS time-series mitigates against occasional errors in quality flags.

Given consistent spectral inputs and quality assessment, a drone data based KNN model is built to estimate the per-HLS pixel vegetation fraction. Drone images were collected across different biomes at different seasons, covering representative land cover and land use. NDVI was selected to calculate vegetation fraction over other indices that employ additional shorter wavelength bands that are impacted by greater scattering effects which we sought to avoid in a global application. A linear translation of NDVI to vegetation fraction was applied for the range of 0.10 to 0.80 based on the studies of (Jiang et al., 2006), (Tucker & Nicholson, 1999), and (Gitelson et al., 2002). Modifying the model to account for varying illumination effects and background soil variation was not feasible, particularly for a global application. For this product, the linear translation was applied to the 8cm drone NDVI values and then aggregated to 30m HLS -pixel level. The averaging of these data to 30m is assumed to ameliorate less well characterized mixed pixels at the 8cm scale. Then we matched the coincident clear-sky HLS data with drone images and employed four bands of HLS data, including red, nir, swir 1.6 and swir 2.1 bandwidths, and drone-derived vegetation fraction as the training inputs (Figure 2). Only bands that are present in both the HLS Landsat and HLS Sentinel collections were included. Additionally, to avoid residual atmospheric contamination, we did not employ shorter wavelength blue and green bands as independent variables in estimating vegetation fraction. We converted the four HLS bands to three principal components through principal component analysis and built the KNN model based on the three principal components using a K-value of 100. The result is a turn-key model which can be applied to any HLS image using the aforementioned bands. The main assumption of the resulting model is that the samples cover all ranges of vegetative cover and may be applied to all the HLS images. A global composite can be seen in Figure 3. While many cover types require a time-series for identification, whether forests or croplands, vegetation fraction can be mapped instantaneously, much like water. As such, time-series of HLS observations in the form of vegetation cover can be used to monitor land change.

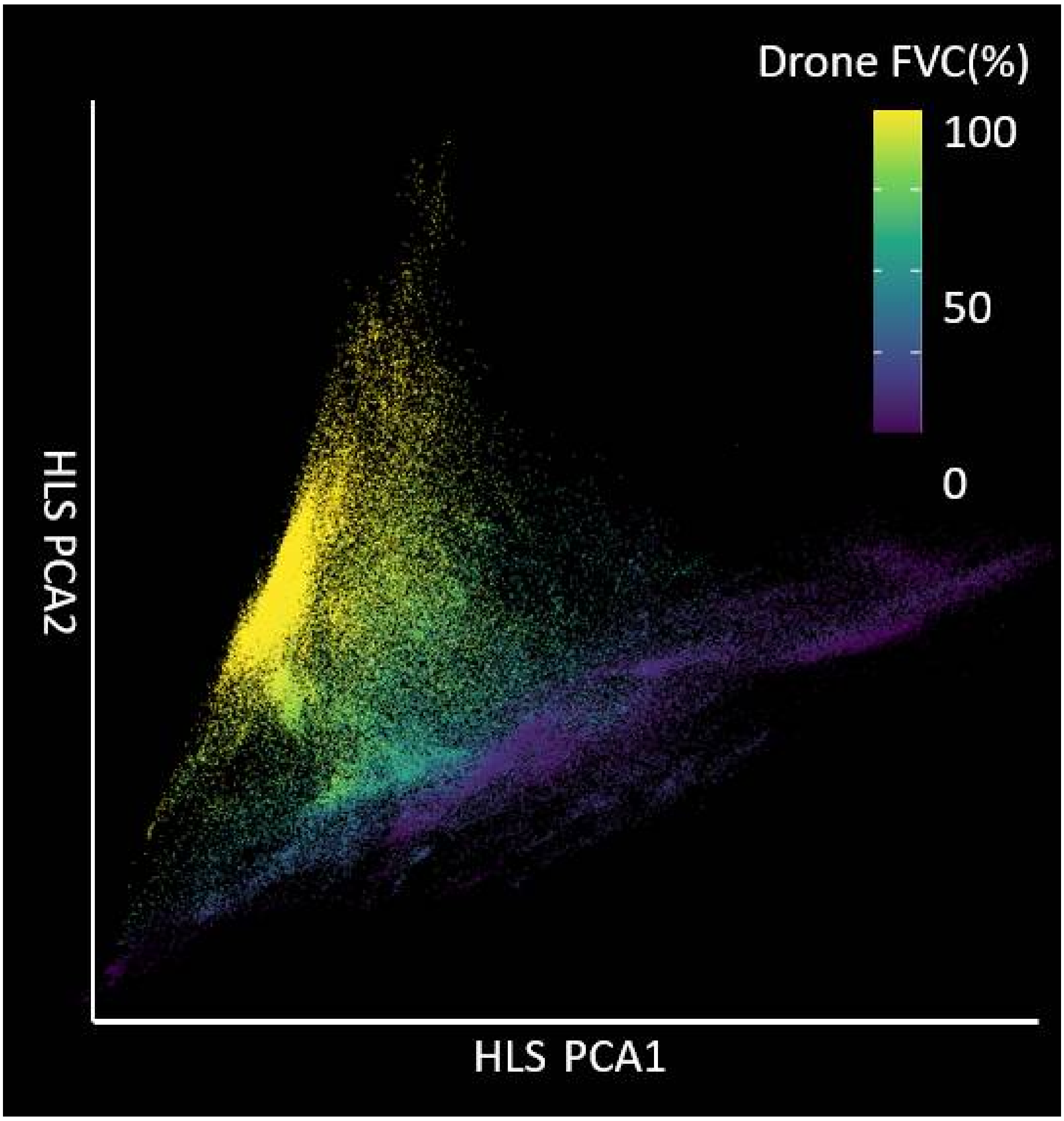

Figure 2: Training data collected from drone-derived Fractional Vegetation Cover (FVC) and coincident HLS data. The data is showed at the first two principle components derived from HLS red, nir, swir 1.6 and swir 2.1 bandwidths, denoted as HLS PCA1 and PCA2.

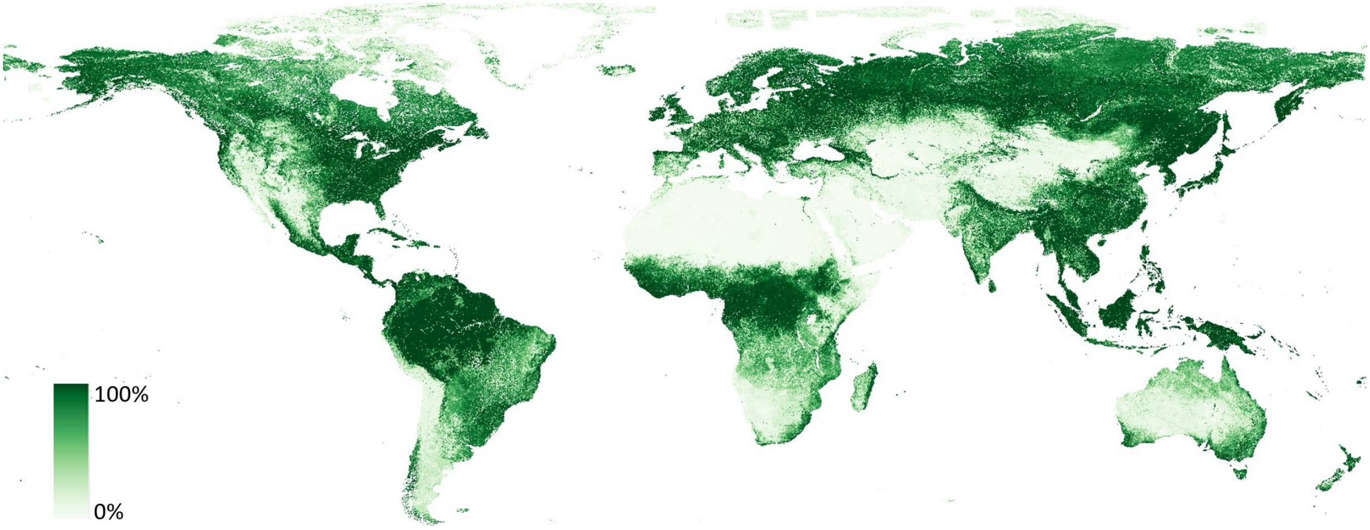

Figure 3: Global vegetation fraction estimates from HLS data from July 2023, gap-filled by surrounding months. Vegetation fraction calculated as percent vegetation cover with the KNN model derived from drone data.

Vegetation Disturbance Algorithm

The DIST algorithm employs a vegetation cover indicator as the input variable. Vegetation cover can be mapped per pixel, recording the natural phenological or managed land use dynamic of the land surface. Near-term historical variation can then be used as a reference for detecting anomalous vegetative cover estimates. Vegetation cover is defined as “the amount of skylight orthogonal to the surface that is intercepted by the cover trait of interest” (Carroll et al., 2010) and includes all plant life over land including both woody and herbaceous (i.e., non-woody) vegetation as with the MODIS VCF product (Hansen et al., 2013) (MODIS VCF ATBD, n.d.). Vegetation disturbance is mapped when there is an indicated decrease in vegetation cover within an HLS pixel, formally defined to be 50% vegetation cover decrease when the scene is compared to the previous calendar years as in (Ying et al., 2017), though the algorithm will report a continuous record of vegetation cover loss. The number of calendar years used as a reference is three years.

Applying the per scene vegetation cover model will result in a time-series of per pixel vegetation cover. ***Figure 4*** shows three dates of HLS-derived vegetation cover over the Bootleg Fire in Oregon, USA. The dates were chosen as they were cloud/smoke-free and outline the scale of the fire extent. ***Figure 5*** depicts a time-series of vegetation cover for a pixel at the western edge of the fire, illustrating the start of the fire for this pixel sometime after the morning of July 6. Disturbance, or vegetation loss, is quantified by the next cloud/smoke-free image on July 16. In algorithm implementation, disturbances in vegetation cover will be identified by comparing each current HLS scene to a summary of cover estimates from previous years representing a lower bound of observed vegetation cover. The composite historical reference is derived from the minimum vegetation fraction of all observations in the previous three years within a 31-day window surrounding the calendar date of the current HLS scene. In order to capture more representative conditions, at least four historical observations are required to calculate the vegetation cover anomaly. In this manner, we may account for intra-annual and seasonal variation in quantifying anomalously low vegetation cover conditions. Regions with high cloud frequency may always have four observations within the baseline period, such as regions of humid tropical forests. In order to enable greater monitoring capcity in these regions, we also employ a three-year, 12-month baseline minimum from the three previous calendar years to identify areas with stable high vegetation presence. When there are not four seasonal baseline estimates available and the annual baseline is ≥85%, then the annual baseline cover estimate is used as the reference baseline to calculate the vegetation cover anomaly.

Strongly seasonal environments may have periods where detection of land disturbance is precluded. For example, low sun angles and winter conditions will lead to fewer observations and lower vegetation contrast when observed. However, the vegetation cover model is sensitive to leaf-off non-photosynthetic woody cover in semi-deciduous and deciduous environments, meaning forests and woodlands will register a positive leaf-off vegetative cover. This outcome is due to the fact that dense leaf-off tree cover has a similar spectral signature to peak greenness transitional shrublands in semi-arid ecotones. ***Figure 6*** illustrates leaf-on, leaf-off vegetation cover estimates as compared to a clearing event in northern Virginia, USA.

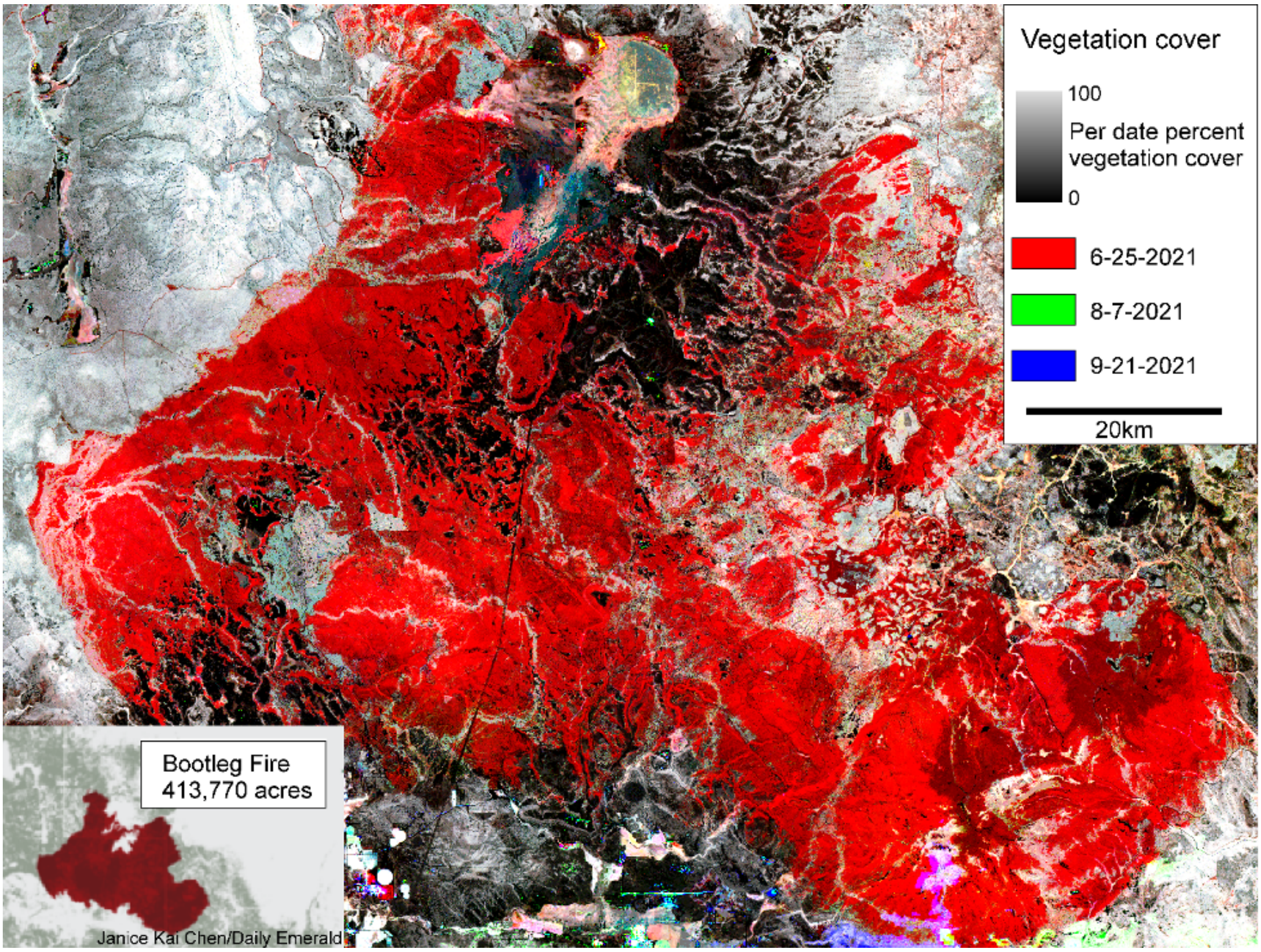

Figure 4: HLS-derived vegetation cover for three dates in r-g-b color composite. Red indicates vegetation loss after June 25, 2021. The three dates were chosen as they were cloud-free and graphically capture fire extent.

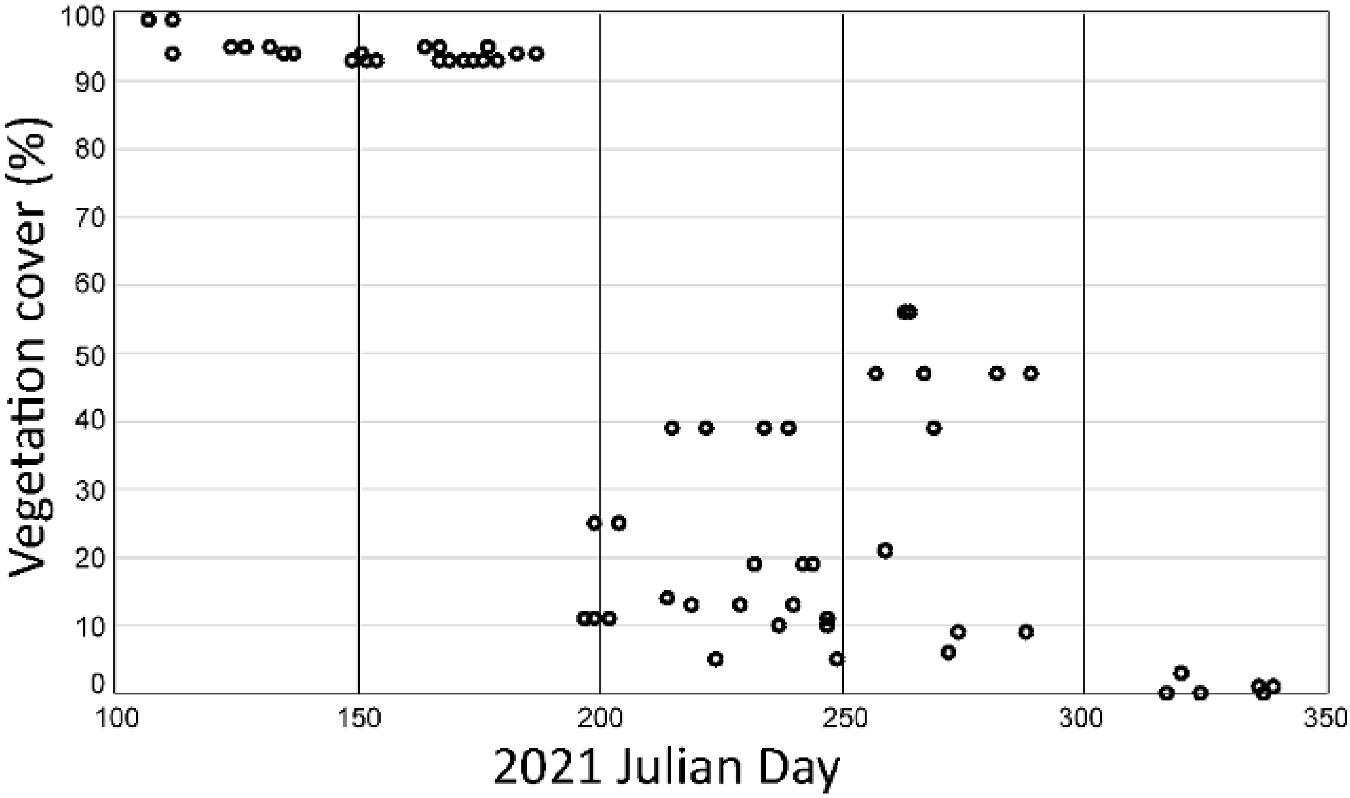

Figure 5: HLS-derived percent vegetation cover for fire-disturbed pixel on the western edge of the Bootleg Fire, Oregon, USA. Pixel is located at 121°24'51"W, 42°38'31"N.

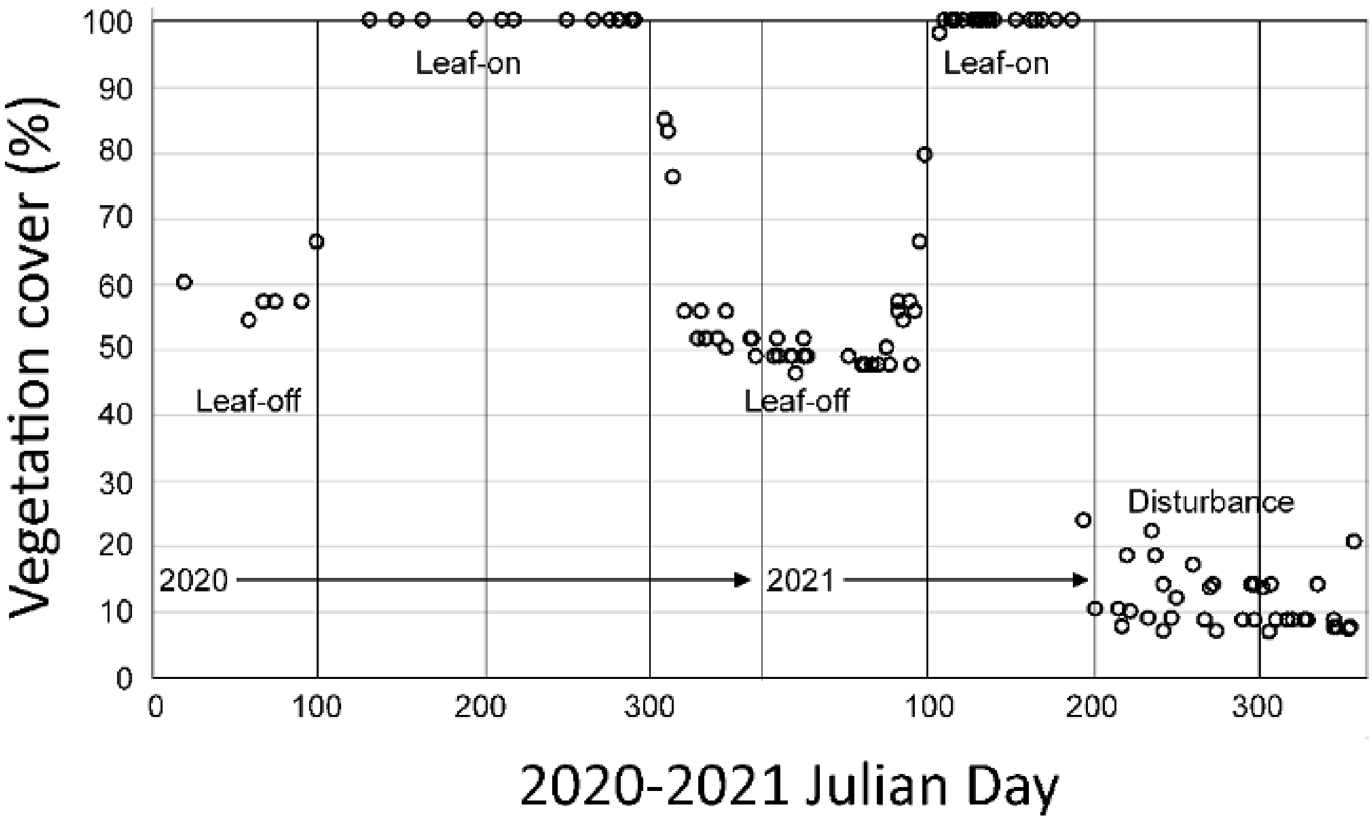

Figure 6: HLS time-series of vegetation cover for deciduous forest pixel in Virginia, USA, cleared between July 9th and 16th, 2021.

Figure 7 shows an example pixel from a shrubland outside of Fort Worth, Texas, USA being converted to a residential land use. In this simplified example, two years of Landsat Analysis Ready Data are composited and the historic per-composite ranges used as a reference for the current year. Observations of the current year are compared to the range of their respective composite periods and vegetation cover estimates below the lower bounds indicate loss. The anomaly magnitude and current vegetation cover estimates are recorded. Both the magnitude and frequency of alerts will be used to integrate time-series information with repeated alert observations resulting in confirmed land disturbance assignation. Beyond confirmed status, a confidence layer which weights repeated alerts will be generated, where confidence equals the mean vegetation loss over a series of alerts multiplied by the number of alerts squared. ***Figure 8*** illustrates scenarios for a range of mean vegetation loss values for this measure. As the system moves forward, the reference data are updated and, in the case shown in ***Figure 7***, a new range of reference vegetation cover will preclude alerts from being repeated in the following year. The full set of time-series layers, including confidence, date, and duration are listed in ***Table 1***.

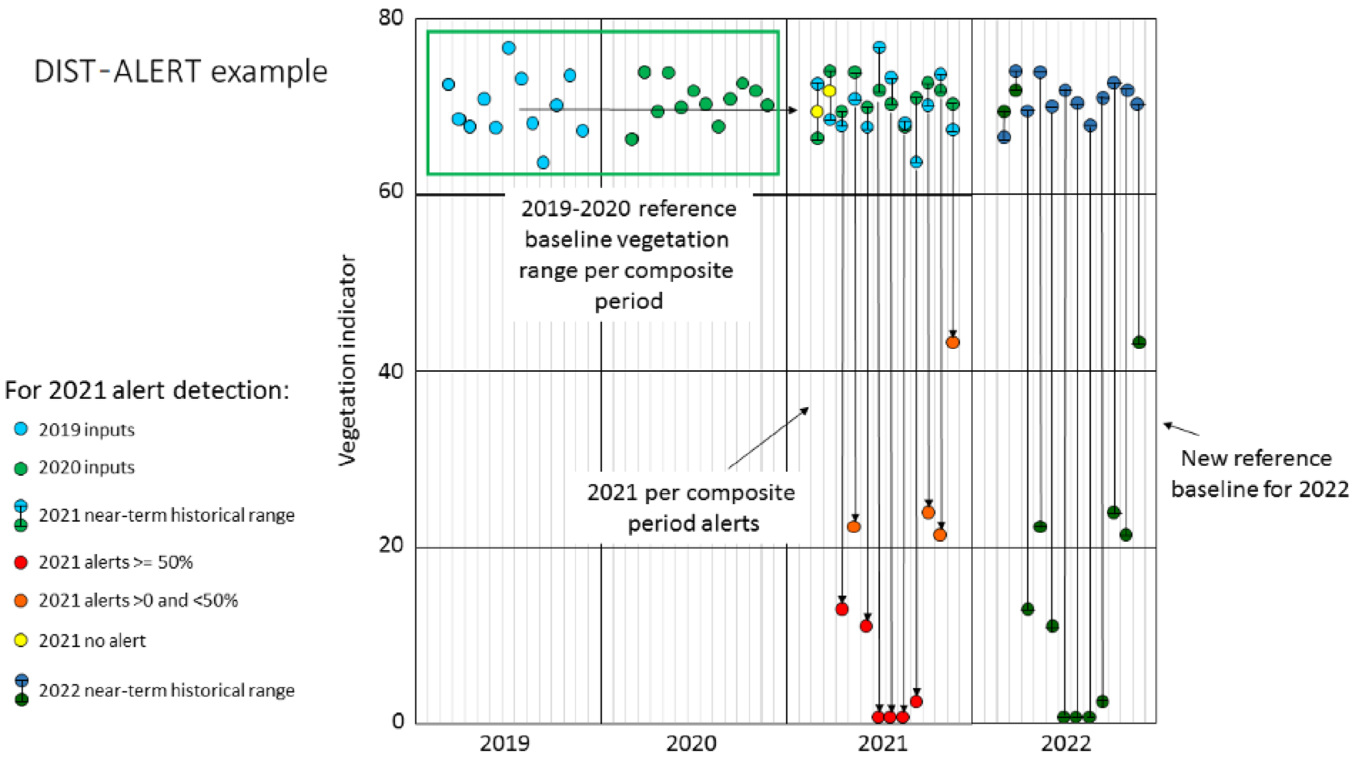

Figure 7: DIST-ALERT example from Analysis Ready Data Landsat observations for a sub-tropical shrubland conversion event near Fort Worth, Texas.

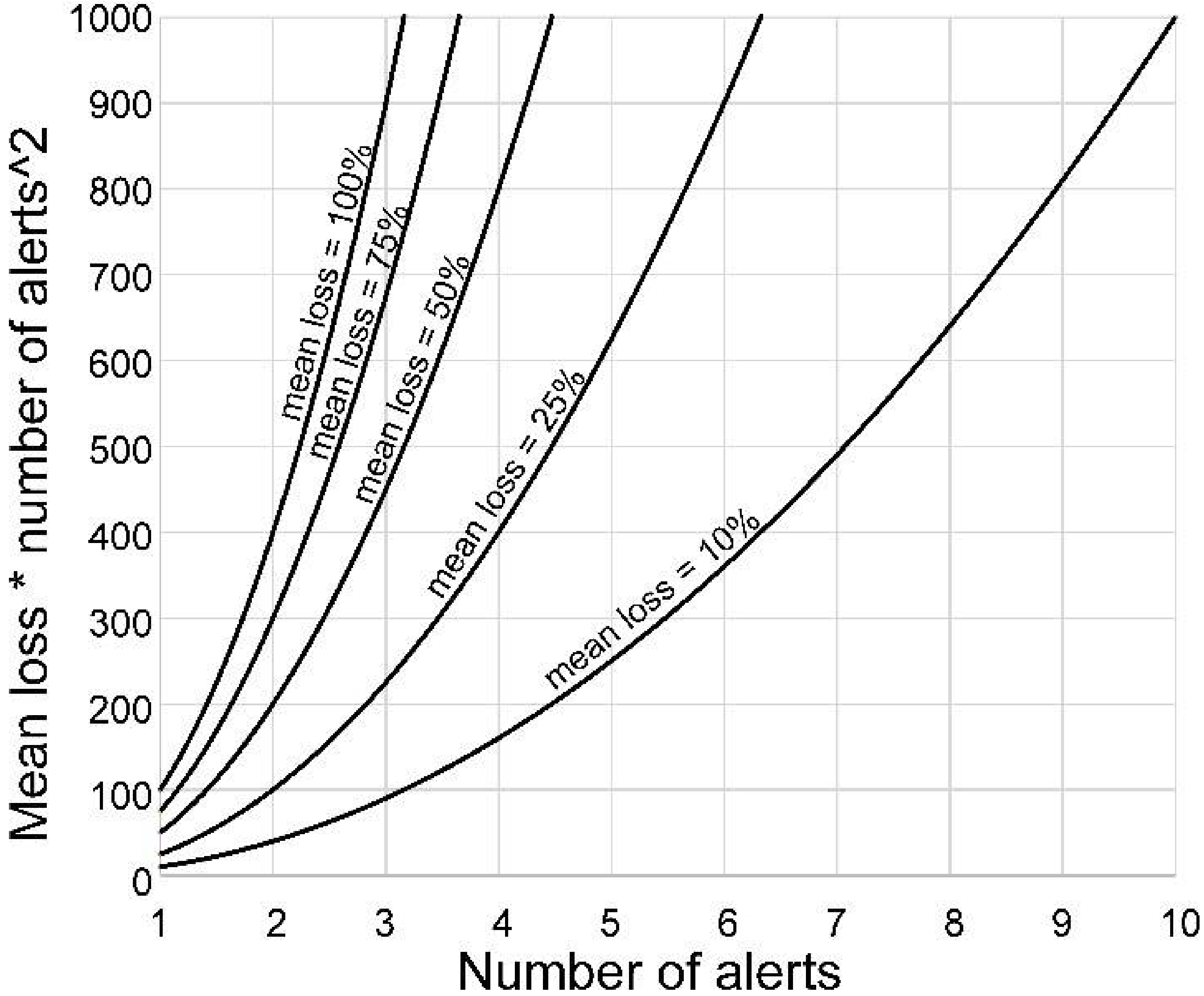

Figure 8: Confidence layer approach (y-axis) combining alert magnitude and frequency. For example a mean 10% loss for 6 alerts would have a confidence of 360.

Spectral Distance Secondary Algorithm

To account for land disturbances unrelated to vegetative cover loss, we include a secondary algorithm that employs a spectral distance measure. Near-time historical data, as with the vegetation fraction algorithm, will be used to calculate a spectral envelope that delimits a normal range of spectral variation. Current HLS spectral signatures will be compared to the normal historical range and outliers calculated. The measure chosen is Mahalanobis distance, as it accounts for co-variance in the near-term historical range, unlike Euclidean distance. ***Figure 9*** shows the difference between Mahalanobis and Euclidean distance in graphical form, comparing the two measures. Each pixel will have a Mahalanobis function based on historical data for the current temporal window from two or more years, as with the vegetation fraction algorithm. A minimum of six historical clear land observations are needed to calculate the baseline envelope. The function will be applied to the current red, near-infrared, and two shortwave infrared bands, again as with the vegetation fraction algorithm, and the value recorded for all valid land observations per HLS scene. Time-series layers derived from the spectral distance alerts will mimic those of the vegetation fraction model, and are listed in ***Table 1***.

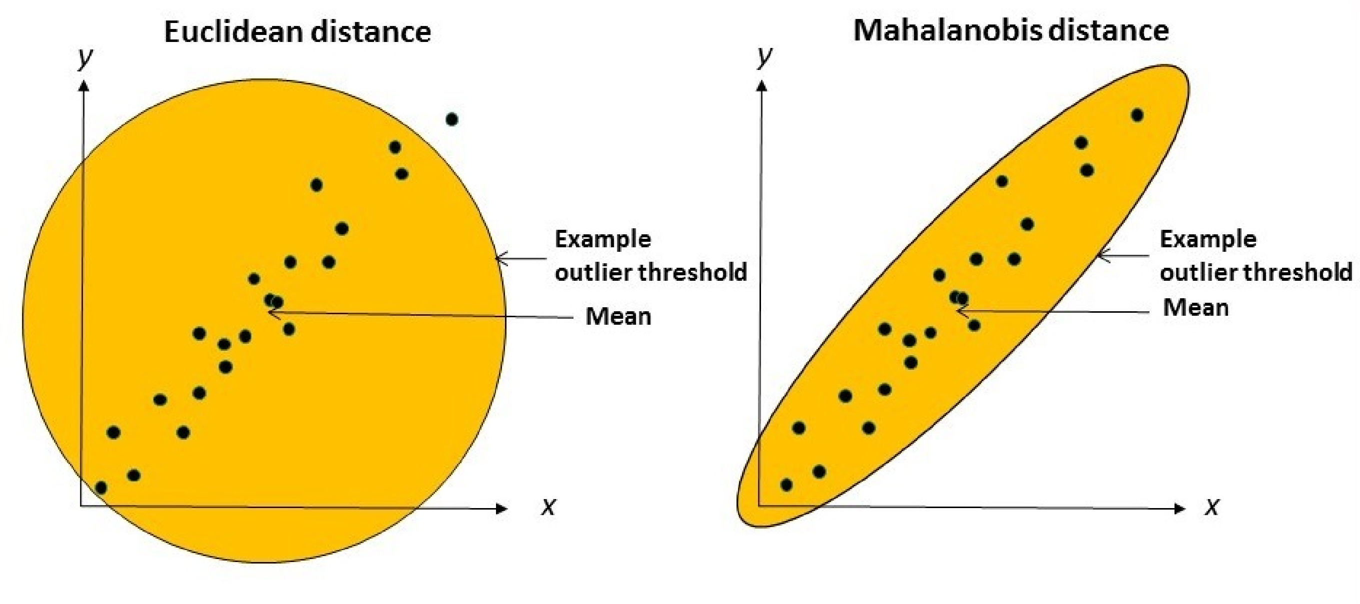

Figure 9: Graphical comparison of Euclidean and Mahalanobis spectral distance measures. Mahalanobis functions calculated from historical data, which better account for co-variance, will be applied to new observations and spectral distances from the mean recorded per HLS scene.

Algorithm Implementations

Two DIST products are generated with respect to their temporal relevance: a) the DIST-ALERT product capturing disturbance at the cadence of HLS sampling (median average 2.9 days for HLS (Li & Roy, 2017)) and b) the DIST-ANN product summarizes changes of the DIST-ALERT products from the previous year. The date of the first disturbance is tracked within both products. Each DIST-ALERT product is associated with an HLS scene and is used to track vegetation disturbances at the temporal frequency of the input HLS dataset. The DIST-ANN tracks changes at the annual scale, aggregating changes identified in the DIST-ALERT product. ***Figure 10*** shows the general flow of operations and products for both algorithms.

The DIST-ALERT product tracks disturbances from initial detection through subsequent observations to increase or decrease confidence. Disturbances identified with the vegetation loss algorithm are tracked independently from those identified with the spectral distance algorithm, but they are each monitored with a parallel set of rules. Disturbances are marked as provisional or confirmed and as low or high intensity. All disturbances begin as provisional alerts starting with the date of initial detection. Through repeated anomaly detections in subsequent observations, alerts can move to confirmed status. The precise number of valid land observations required for a confirmed status will be determined during the algorithmic calibration and are a function of both the magnitude of the anomaly and the number of anomaly detections. This confirmation will come from a variable number of HLS scenes due to invalid observations contaminated by cloud or shadow. If a pixel marked provisional disturbance has no observed loss in subsequent images, then this label will be removed and this pixel's vegetation cover will continue to be analyzed for future vegetation cover losses. Additional contextual layers are provided for disturbed pixels including: the date of initial disturbance, vegetation disturbance confidence, number of observed anomalies, and disturbance duration. An example of provisional versus confirmed alerts is shown in Figure 10. For vegetation loss disturbances, pixels are also marked as low or high intensity based on whether the estimated vegetation cover loss is ≥50%. For general disturbances identified by the spectral distance algorithm, the distance threshold delineating low and high intensity will be determined during algorithm calibration.

In DIST-ANN, only confirmed disturbances from the associated year are reported together with the date of initial disturbance. As confirmed disturbances are determined using subsequent cloud-free observations to determine if the loss detections persist, the required number of HLS scenes depends on visibility of the target. Due to this, summarizing the DIST-ALERT in the DIST-ANN product will have some latency depending on the algorithmic calibration and detailed in subsequent documentation. Additional contextual layers are provided for disturbed pixels including: the date of initial disturbance, vegetation disturbance confidence, number of observed anomalies, and disturbance duration. An example of provisional versus confirmed alerts is shown in ***Figure 11.***

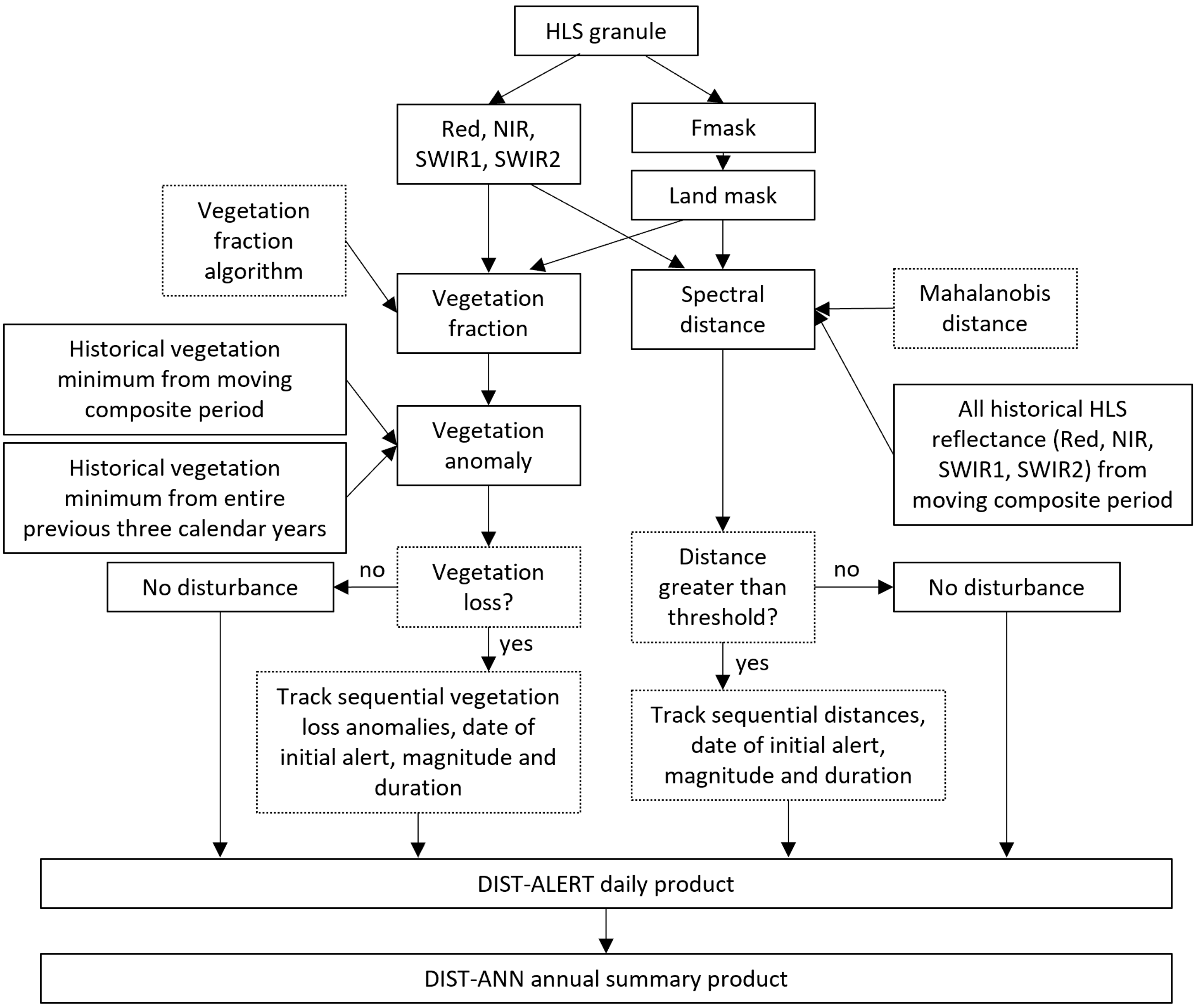

Figure 10: Flow of processes and outputs of the DIST product suite.

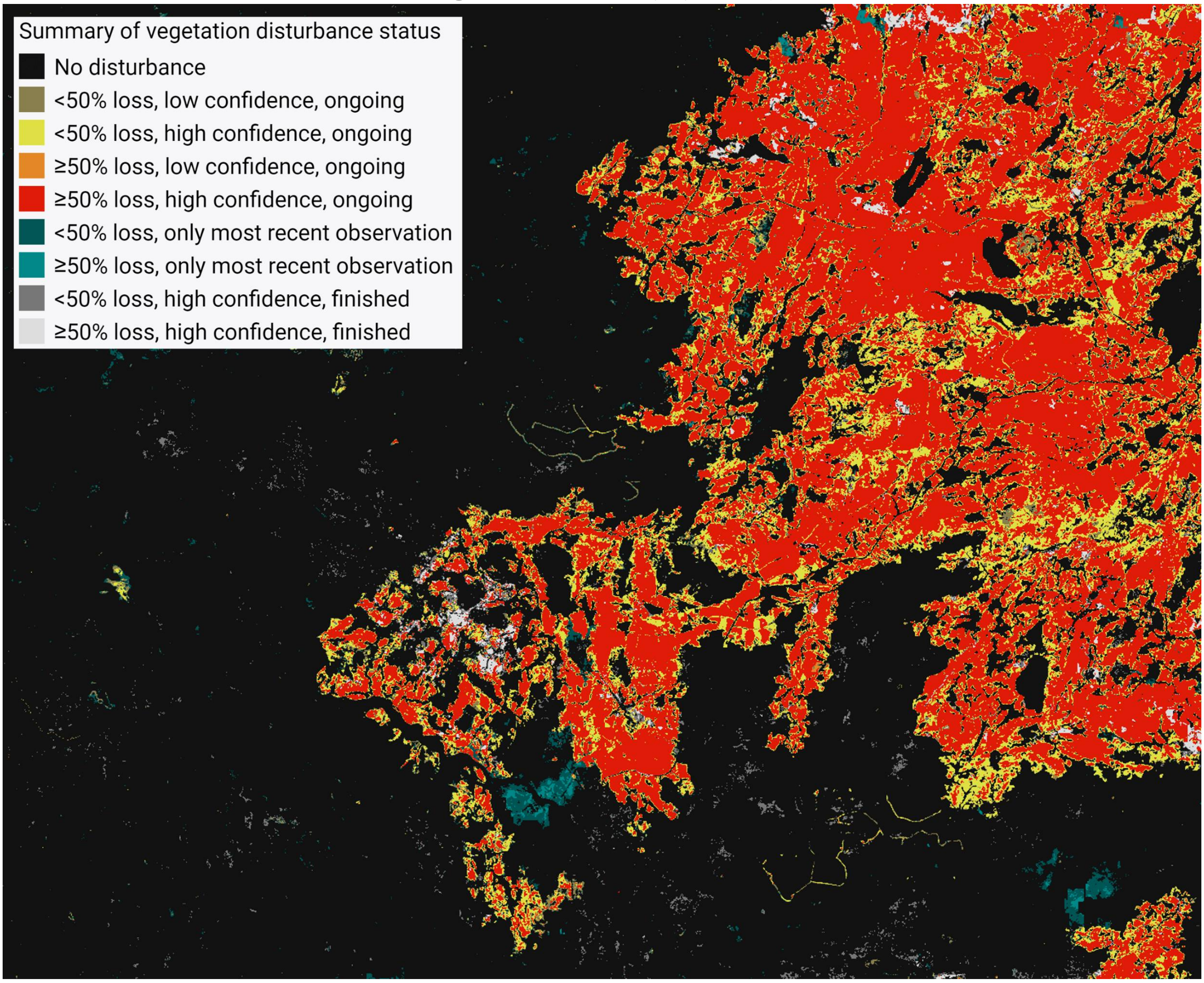

Figure 11: Map showing a Quebec wildfire captured in the DIST-ALERT vegetation disturbance status from September 13, 2023.

Output Product Layers

The layers detailed in ***Tables 1*** are output for each HLS scene/tile for which vegetation cover is estimated. For every HLS scene a per-pixel estimate of the current percent vegetation cover indicator and the current anomaly value are provided within the DIST-ALERT product. The anomaly value is defined as the difference in estimated percent vegetation cover between the seasonally normalized lower bound of historical vegetation cover (historic vegetation cover indicator) and the percent vegetation cover estimate from the current HLS scene. Only anomalies of vegetation loss are reported. Although disturbances must be reported for ≥50% vegetation cover loss per the project requirements and validation activities, all disturbances with vegetation cover loss ≥10% are tracked in the time-series. The maximum anomaly and duration layers can be leveraged to assess the magnitude and duration of disturbances. Given potential rapid vegetation recovery, the anomaly value corresponding to the date of maximum anomaly as well as the historical lower bound from that date are reported. As the historical lower bound corresponds to the date of maximum anomaly it is not reported for pixels without recorded anomalies. The vegetation cover estimate for the current year at the date of maximum anomaly can be calculated from these two values. ***Table* 2** shows all the metadata of DIST-ALERT product, which can be found in the cmr.json metadata file in the package of DIST-ALERT product.

Table 3 lists all the layers in the annual summary DIST-ANN product. And ***Table 4*** lists all the the metadata of DIST-ANN. The metadata file can be found in a cmr.json file of the DIST-ANN package. There will also be a comma delieted file(CSV) listing all the DIST-ALERT input files, HLS source files and image metadata.

Table 1: Product Raster Layers for DIST-ALERT

DIST-ALERT Raster Layer	**Description**	**File name**	**Data type**	**Layer values**
Vegetation disturbance status	Indication of vegetation cover loss (vegetation disturbance). The status label is based on the maximum anomaly value, confidence level, and whether it is ongoing or finished. “First” means the pixel has had an anomaly detection but no subsequent observations whether anomalous or not. “Provisional” means there have been two consecutive disturbance detections but not yet high confidence. “Confirmed” means that vegetation disturbance is detected with high confidence (≥400). The label “finished” is applied to confirmed disturbances that have had two consecutive no-anomaly observations or one 15 days or more after the last anomaly detection. If a new disturbance is detected, it will overwrite those in a “finished” state. These labels are reported for both above and below the 50% disturbance threshold based on the maximum anomaly value.	VEG-DIST-STATUS	UInt8	0: No disturbance 1: first <50% 2: provisional <50% 3: confirmed <50% 4: first ≥50% 5: provisional ≥50% 6: confirmed ≥50% 7: confirmed <50%, finished 8: confirmed ≥50%, finished

				255: No data
Current vegetation cover indicator	The percent vegetation cover estimated for the current HLS scene for all land and water pixels.	VEG-IND	UInt8	0-100: Estimated percent vegetation 255: No data
Current vegetation anomaly value	Difference between historical baseline and observed vegetation cover at the current date (vegetation loss of 0-100%). When >0, the sum of this anomaly value and the current vegetation cover indicator will be the historical vegetation cover estimate.	VEG-ANOM	UInt8	0-100: Estimated loss of percent vegetation 255: No data
Historical vegetation cover indicator	Historical percent baseline value at the time of the maximum anomaly for disturbance pixels. A fill value is used for all non-disturbance pixels. Historical vegetation is calculated from all HLS scenes within a synchronous temporal window (±15 days) from previous three calendar years to capture intra-annual/seasonal variation.	VEG-HIST	UInt8	0-100: Vegetation percent 200: No disturbance 255: No data
Max vegetation anomaly value	Difference between historical and current year observed vegetation cover at the date of maximum decrease (vegetation loss of 0-100%).. This layer can be used to threshold vegetation disturbance per a given sensitivity (e.g. disturbance of ≥20% vegetation cover loss). The sum of the historical percent vegetation and the anomaly value will be the vegetation cover estimate for the current year.	VEG-ANOM-MAX	UInt8	0-100: Maximum loss of percent vegetation 255: No data
Vegetation	Mean anomaly value since initial anomaly detection multiplied by the number of loss anomalies squared. Confidence is	VEG-	Int16	-1: No data

Disturbance Confidence Layer	calculated until the anniversary date is reached, or a fixed number of consecutive non-anomalies are observed causing the status (VEG-DIST-STATUS) to change to “finished”.	DIST-CONF		0: No disturbance >0: Disturbance confidence
Date of initial vegetation disturbance	Day of first loss anomaly detection of the most recent disturbance event. Day denoted as the number of days since December 31, 2020.	VEG-DIST-DATE	Int16	-1: No data 0: No vegetation anomalies in the last year >0: Day of initial anomaly detection in the last year
Number of detected vegetation loss anomalies	Total number of observations with anomalous low vegetation since initial anomaly detection (inclusive). Maximum of 254.	VEG-DIST-COUNT	UInt8	0: No disturbance 1-254: Count of loss anomalies 255: No data
Vegetation disturbance duration	Number of days of ongoing loss anomalies since initial anomaly detection (inclusive). Maximum duration is one year.	VEG-DIST-DUR	Int16	-1: No data 0: No disturbance 1-366: number

				of days from first anomaly to most recent anomaly detection
Date of last observation assessed for vegetation disturbance	Day of last quality assessed HLS observation flagged as land or water that also had sufficient observations for baseline calculation for vegetation disturbance algorithm. Day denoted as the number of days since December 31, 2020.	VEG-LAST-DATE	Int16	-1: No data ≥1: Last day assessed
Generic disturbance status	Indication of generic spectral difference. The status label is based on the maximum anomaly value, confidence level, and whether it is ongoing or finished. “First” means the pixel has had an anomaly detection but no subsequent observations whether anomalous or not. “Provisional” means there have been two consecutive disturbance detections but not yet high confidence. “Confirmed” means that disturbance is detected with high confidence. The label “finished” is applied to confirmed disturbances that have had two consecutive no-anomaly observations or one 15 days or more after the last anomaly detection. If a new disturbance is detected, it will overwrite those in a “finished” state. These labels are reported for both above a low and high threshold based on the maximum spectral anomaly.	GEN-DIST-STATUS	UInt8	0: No disturbance 1: first low 2: provisional low 3: confirmed low 4: first high 5: provisional high 6: confirmed high 7: confirmed low, finished

				8: confirmed high, finished 255: No data
Current generic disturbance anomaly value	Spectral distance between current HLS scene reflectance and the reflectance of the previous three calendar years within ±15 calendar days. Calculated by Mahalanobis distance.	GEN-ANOM	Int16	-1: No data 0: No disturbance >0: Spectral distance
Generic disturbance maximum anomaly value	Maximum spectral distance between a current year HLS scene reflectance and the composite reflectance of previous calendar years.	GEN-ANOM-MAX	Int16	-1: No data 0: No disturbance >0: Spectral distance
Generic Disturbance Confidence Layer	Mean spectral distance since initial spectral anomaly detection times the number of spectral anomalies above a threshold, until the anniversary date is reached, or a fixed number of consecutive non-anomalies are observed.	GEN-DIST-CONF	Int16	-1: No data 0: No disturbance >0: Disturbance confidence
Date of initial generic disturbance	Day of first spectral anomaly detection of the most recent disturbance event. Day denoted as the number of days since December 31, 2020.	GEN-DIST-DATE	Int16	-1: No data 0: No spectral anomalies in the last year

				>0: Day of initial anomaly detection in the last year
Number of detected spectral anomalies	Total number of observations with spectral reflectance anomalies (inclusive). Maximum of 254.	GEN-DIST-COUNT	UInt8	0: No disturbance 1-254: Count of loss anomalies 255: No data
Generic disturbance duration	Number of days of ongoing spectral reflectance anomalies since initial anomaly detection (inclusive). Maximum duration is one year.	GEN-DIST-DUR	Int16	-1: No data 0: No disturbance 1-366: number of days from first loss anomaly to most recent anomaly detection
Date of last observation assessed for generic disturbance	Day of last quality assessed HLS observation flagged as land that also had sufficient observations for baseline calculation for generic disturbance algorithm. Day denoted as the number of days since December 31, 2020.	GEN-LAST-DATE	Int16	-1: No data ≥1: Last day assessed

Data mask	Mask of pixels the algorithms are applied to in the current HLS scene. Based on the Fmask layer of the source HLS granule.	DATA-MASK	UInt 8	0: Not land 1: Land

Table 2: DIST-ALERT metadata

Attribute	**Description**
GranuleUR	The granule ID for each DIST-ALERT. Format: OPERA_L3_DIST-ALERT-HLS_Tile_YYYYMMDDTHHMMSSZ_ YYYYMMDDTHHMMSSZ_S2A_30_v1
TemporalExtent: RangeDateTime	The DIST-ALERT product versionTemporal extent of the HLS data, flagged as BeginningDateTime and EndingDateTime. Format: YYYY-MM-DDTHH:MM:SS.SSSSSSZ
ProviderDates	The date of DIST-ALERT product be provided
CollectionReference:ShortName	The short name of the collection, OPERA_L3_DSIT-ALERT-HLS_V1
CollectionReference:Version	The DIST-ALERT product version
DataGranule: DayNightFlag	Flag if the image is during the day or night
DataGranule: ProductionDateTime	DIST-ALERT product processing date. Format: YYYY-MM-DDTHH:MM:SS.SSSSSSZ
Platforms	Name of the sensor platform (e.g. Landsat-8/9 or Sentinel-2 A/B)
Instruments	Name of the sensor instrument (e.g. OLI or MSI)
SpatialExtent	The longitude and latitude boundary of the image
CloudCover	The percentage of cloud and cloud shadow in the DIST-ALERT product (copied from HLS)
Input_DIST-ALERT_granule	The input DIST-ALERT granule ID

BaselineCalendarWindow	Number of days before and after the calendar date used to create the baseline
BaselineYearWindow	Number of previous years used to create the baseline
BaselineImageIds	List of the input HLS granules used to create the baseline
ValidationLevel	The validation level of the product
HLSGranuleUR	Name of the input HLS granule used to generate the DIST-ALERT product
SENSOR_PRODUCT_ID	The source Landsat or Sentinel-2 data ID
SPATIAL_COVERAGE	The area percentage of the tile with data (copied from HLS)
MGRS_TILE_ID	The tile ID
HLS_PROCESSING_TIME	The input HLS granule processing date. Format: YYYY-MM-DDTHH:MM:SS.SSSSSSZ.
SENSING_TIME	The sensing time provided with the source Landsat or Sentinel-2 image. Format: YYYY-MM-DDTHH:MM:SSZ.
HORIZONTAL_CS_CODE	The code for the coordinate system, eg: “EPSG:32655”
HORIZONTAL_CS_NAME	The name of the coordinate system, eg :"UTM, WGS84, UTM ZONE 55"
ULX	The E-W coordinate of the upper left within the given coordinate system
ULY	The N-S coordinate of the upper left within the given coordinate system

Table 3: Product Raster Layers for DIST-ANN

DIST-ANN Raster Layer	**Description**	**File name**	**Data type**	**Layer values**
Vegetation disturbance status	Status corresponding to the highest confidence vegetation disturbance confirmed within the year. Status classes identify confirmed ongoing disturbance, confirmed finished disturbance, and	VEG-DIST-STATUS	UInt 8	0: No disturbance 3: confirmed <50% ongoing

	confirmed disturbance initially detected in previous year for both <50% and ≥50%, and no disturbance.			6: confirmed ≥50% ongoing 7: confirmed <50% finished 8: confirmed ≥50% finished 9: confirmed previous year <50% 10: confirmed previous year ≥50% 255: No data
Historical vegetation cover indicator	Historical percent vegetation from composite of HLS scenes during the same time period of the maximum anomaly for disturbance pixels. A fill value is used for all non-disturbance pixels. Historical vegetation is calculated from a synchronous temporal window from previous calendar years to capture intra-annual/seasonal variation.	VEG-HIST	UInt 8	0-100: Vegetation percent 200: No disturbance 255: No data
Maximum vegetation cover indicator	For non-disturbance pixels, maximum annual vegetation fraction from the HLS time-series data will be reported. For disturbance pixels, the vegetation fraction from the date of maximum anomaly will be reported.	VEG-IND-MAX	UInt 8	0-100: Estimated loss of percent vegetation 255: No data
Maximum vegetation anomaly value	Difference between historical vegetation cover and vegetation cover at the date of maximum decrease (vegetation loss of 0-100%). This layer can be used to threshold vegetation disturbance per a given sensitivity (e.g. disturbance of ≥20% vegetation cover loss).	VEG-ANOM-MAX	UInt 8	0-100: Maximum loss of percent vegetation 255: No data
Vegetation Disturbance Confidence Layer	Mean anomaly value since initial anomaly detection times the number of loss anomalies squared, until the anniversary date is reached, or a fixed number of consecutive non-anomalies are observed.	VEG-DIST-CONF	Int16	-1: No data 0: No disturbance >0: Disturbance confidence

Date of initial vegetation disturbance	Day of first loss anomaly. Day denoted as the number of days since December 31, 2020.	VEG-DIST-DATE	Int16	-1: No data 0: No disturbance >0: Day of first loss anomaly detection
Number of detected vegetation loss anomalies	Total number of loss anomalies since initial anomaly detection(inclusive). Maximum of 254.	VEG-DIST-COUNT	UInt 8	0: No disturbance 1-254: Count of loss anomalies 255: No data
Vegetation disturbance duration	Number of days of ongoing loss anomalies since initial anomaly detection (inclusive). Maximum duration is one year.	VEG-DIST-DUR	Int16	-1: No data 0-366: number of days from first anomaly to most recent anomaly detection
Indicator of vegetation disturbance from previous year	Indicator of whether the highest confidence vegetation disturbance event confirmed within the year (corresponding to the above layers) was initially detected in the previous calendar year.	VEG-CONF-PREV	UInt 8	0: no disturbance 1: confirmed low previous year, 2: confirmed high previous year, 255: no data
Count of confirmed vegetation disturbance events	Count of distinct confirmed vegetation disturbance events.	VEG-CONF-COUNT	UInt 8	≥0: count of confirmed vegetation disturbance events 255: no data
Minimum three year vegetation	The minimum vegetation cover of the current year and two previous years with stricter aerosol filtering. Becomes input to the following year's DIST-ALERT product.	VEG-IND-3YR-MIN	UInt 8	0-100: Vegetation percent 255: No data

cover indicator				
Date of last observation assessed for vegetation disturbance	Day of last quality assessed HLS observation flagged as land that also had sufficient observations for baseline calculation for vegetation disturbance algorithm. Day denoted as the number of days since December 31, 2020.	VEG-LAST-DATE	Int16	-1: No data 0: Never flagged as land >0: Day of last land observation
Generic disturbance status	Status corresponding to the highest confidence generic spectral difference confirmed within the year. Status classes identify confirmed ongoing disturbance, confirmed finished disturbance, and confirmed disturbance initially detected in previous year for both above a low and high threshold and no disturbance.	GEN-DIST-STATUS	UInt8	0: No disturbance 3: confirmed low, ongoing 6: confirmed high, ongoing 7: confirmed low, finished 8: confirmed high, finished 9: confirmed low, previous year 10: confirmed high, previous year 255: No data
Generic maximum disturbance anomaly value	Maximum spectral distance between a current year HLS scene reflectance and the composite reflectance of previous calendar years.	GEN-ANOM-MAX	Int16	-1: No data 0: No disturbance >0: Spectral distance
Generic Disturbance Confidence Layer	Mean spectral distance since initial spectral anomaly detection times the number of spectral anomalies above a threshold, until the anniversary date is reached, or a fixed number of consecutive non-anomalies are observed.	GEN-DIST-CONF	Int16	-1: No data 0: No disturbance >0: Disturbance confidence
Date of generic	Day of first spectral anomaly. Day denoted as the number of days since December 31, 2020.	GEN-DIST-	Int16	-1: No data

initial disturbance anomaly		DATE		0: No disturbance >0: Day of first anomaly detection
Number of detected spectral anomalies	Total number of observations with a spectral anomaly since initial anomaly detection (inclusive). Maximum of 254.	GEN-DIST-COUNT	UInt 8	0: No loss anomalies 1-254: Count of loss anomalies 255: No data
Generic disturbance duration	Number of days of ongoing spectral anomalies since initial anomaly detection (inclusive). Maximum duration is one year.	GEN-DIST-DUR	Int16	-1: No data 0-366: number of days from first anomaly to most recent anomaly detection
Indicator of generic disturbance from previous year	Indicator of whether the highest confidence generic disturbance event confirmed within the year (corresponding to the above GEN layers) was initially detected in the previous calendar year.	GEN-CONF-PREV	UInt 8	0: no disturbance 1: confirmed low previous year, 2: confirmed high previous year, 255: no data
Count of confirmed generic disturbance events	Count of distinct confirmed generic disturbance events.	GEN-CONF-COUNT	UInt 8	>0: count of confirmed generic disturbance alert 255: no data
Date of last observation assessed for generic disturbance	Day of last quality assessed HLS observation flagged as land that also had sufficient observations for baseline calculation for generic disturbance algorithm. Day denoted as the number of days since December 31, 2020.	GEN-LAST-DATE	Int16	-1: No data 0: Never flagged as land

				>0: Day of last land observation

Table 4: DIST-ANN Metadata

Attribute	Description
GranuleUR	The granule ID for each DIST-ANN. Format: OPERA_L3_DIST-ANN-HLS_Tile_YYYY_ YYYYMMDDTHHMMSSZ_ 30_v1
TemporalExtent: RangeDateTime	Temporal extent of the HLS data of the input year, flagged as BeginningDateTime and EndingDateTime. Format: YYYY-MM-DDTHH:MM:SS.SSSSSSZ
ProviderDates	The date-time of when the granule was sent to LP-DAAC. Format: YYYY-MM-DDTHH:MM:SS.SSSSSSZ
CollectionReference:ShortName	OPERA_L3_DIST-ANN-HLS_V1
CollectionReference:Version	The DIST-ANN product version
SpatialExtent	The longitude and latitude boundary of the image
CloudCover	The percentage of no-data in the DIST-ANN product
Platforms	Names of the input sensor platforms (e.g. Landsat 8/9 and Sentinel-2 A/B)
Instruments	Names of the input sensor instruments (e.g. OLI and Sentinel-2 MSI)
ValidationLevel	The validation level of the product
SPATIAL_COVERAGE	The area percentage of the tile with data
MGRS_TILE_ID	The tile ID
HORIZONTAL_CS_CODE	The code for the coordinate system, eg: “EPSG:32655”
HORIZONTAL_CS_NAME	The name of the coordinate system, eg :"UTM, WGS84, UTM ZONE 55"
ULX	The E-W coordinate of the upper left within the given coordinate system

ULY	The N-S coordinate of the upper left within the given coordinate system
PROCESSING_DATETIME	DIST-ANN product processing date. Format: YYYY-MM-DDTHH:MM:SSZ.

3.2.1. Mathematical Theory Assumptions

As with any machine learning algorithm, its performance is only as good as its input data sets. High resolution drone images are collected as the training data for vegetation fraction estimation. We have iterated to collect the representative land cover and land use to fill different compositions in the principle components space and to cover all ranges of vegetation cover.

3.3. Algorithm Input Variables

Variable #1

NAME	**HLS FMASK**
LONG NAME	**HLS quality assessment layer**
UNIT	**Unit8**

Variable #2

NAME	**HLS RED**
LONG NAME	**HLS red band**
UNIT	**Int16**

Variable #3

NAME	**HLS NIR**
LONG NAME	**HLS near infrared band**
UNIT	**Int16**

Variable #4

NAME	**HLS SWIR1**
LONG NAME	**HLS shortwave infrared band 1**
UNIT	**Int16**

Variable #5

NAME	HLS SWIR2
LONG NAME	HLS shortwave infrared band 2
UNIT	Int16

Variable #6

NAME	Ocean mask
LONG NAME	Global ocean mask
UNIT	Unit8

3.4. Algorithm Output Variables

Variable #1

NAME	VEG-DIST-TATUS
LONG NAME	Vegetation disturbance status
UNIT	UInt8

Variable #2

NAME	VEG-HIST
LONG NAME	Historical vegetation cover indicator
UNIT	UInt8

Variable #3

NAME	VEG-IND-MAX
LONG NAME	Maximum vegetation cover indicator
UNIT	UInt8

Variable #4

NAME	VEG-ANOM-MAX
LONG NAME	Maximum vegetation anomaly value
UNIT	UInt8

Variable #5

NAME	VEG-DIST-CONF
LONG NAME	Vegetation Disturbance Confidence Layer
UNIT	Int16

Variable #6

NAME	**VEG-DIST-DATE**
LONG NAME	**Date of initial vegetation disturbance**
UNIT	**Int16**

Variable #7

NAME	**VEG-DIST-COUNT**
LONG NAME	**Number of detected vegetation anomalies**
UNIT	**UInt8**

Variable #8

NAME	**VEG-DIST-DUR**
LONG NAME	**Vegetation disturbance duration**
UNIT	**Int16**

Variable #9

NAME	**VEG-LAST-DATE**
LONG NAME	**Date of last observation assessed for vegetation disturbance**
UNIT	**Int16**

Variable #10

NAME	**GEN-DIST-STATUS**
LONG NAME	**Generic disturbance status**
UNIT	**UInt8**

Variable #11

NAME	**GEN-ANOM-MAX**
LONG NAME	**Generic maximum disturbance anomaly value**
UNIT	**Int16**

Variable #12

NAME	**GEN-DIST-CONF**
LONG NAME	**Generic Disturbance Confidence Layer**
UNIT	**Int16**

Variable #13

NAME	**GEN-DIST-COUNT**
LONG NAME	**Number of detected spectral anomalies**
UNIT	**UInt8**

Variable #14

NAME	**GEN-DIST-DUR**
LONG NAME	**Generic disturbance duration**
UNIT	**Int16Int16**

Variable #15

NAME	**GEN-LAST-DATE**
LONG NAME	**Date of last observation assessed for generic disturbance**
UNIT	**Int16**

Variable #16

NAME	**VEG-IND**
LONG NAME	**Current vegetation cover indicator**
UNIT	**UInt8**

Variable #17

NAME	**VEG-ANOM**
LONG NAME	**Current vegetation anomaly value**
UNIT	**UInt8**

Variable #18

NAME	**GEN-ANOM**
LONG NAME	**Current generic disturbance anomaly value**
UNIT	**Int16**

Variable #19

NAME	**GEN-DIST-DATE**
LONG NAME	**Date of initial generic disturbance**
UNIT	**Int16**

Variable #20

NAME	**LAND-MASK**
LONG NAME	**Land mask**
UNIT	**UInt8**

Variable #21

NAME	**VEG-CONF-PREV**
LONG NAME	**Indicator of vegetation disturbance from previous year**
UNIT	**UInt8**

Variable #22

NAME	**VEG-CONF-COUNT**
LONG NAME	**Count of confirmed vegetation disturbance events**
UNIT	**UInt8**

Variable #23

NAME	**VEG-IND-3YR-MIN**
LONG NAME	**Minimum three year vegetation cover indicator**
UNIT	**UInt8**

Variable #24

NAME	**GEN-CONF-PREV**
LONG NAME	**Indicator of generic disturbance from previous year**
UNIT	**UInt8**

Variable #25

NAME	**GEN-CONF-COUNT**
LONG NAME	**Count of confirmed generic disturbance events**
UNIT	**UInt8**

4. Algorithm Usage Constraints

The principal limitation to the DIST products is there is no attribution of the causes of disturbances and the type is only defined as vegetation loss or general. The vegetation algorithm is a generic indicator of vegetation loss, whether due to logging, landslides, development or any other event resulting in reduced vegetation cover. Example vegetation loss dynamics to be detected with the DIST products are shown in ***Figure 12***, with a resulting need for users to place appropriate context on the outputs. The dynamics detected with the spectral distance algorithm are spread across an even further range including both vegetation loss and gain and change in non-vegetated surfaces, such as a parking lot replaced with a building or a lava flow in the desert. Future iterations of the algorithm may include reference cover state or supplementary algorithms to assign dynamics of antecedent land cover or land use, change factor, and resulting cover.

Additionally, some land disturbances will not be detected by the vegetation disturbance layer, including vegetation recovery, phenological and intra-annual vegetation changes, urban development within urban sprawl (e.g. buildings being replaced or being demolished) or more generally, any urban changes of non-vegetated areas (e.g. a highway being built over a desert landscape), lava flows over a rocky, non-vegetated terrain. However, some of these will be able to be detected with the spectral distance algorithm, but areas with more spectral variation in the baseline will be less likely to detect changes. Land use dynamics that are part of an annual practice, for example crop rotations or tilling practices, or fire as maintenance of vegetative cover, for example savanna fires in tropical Africa, will not be flagged as disturbance if they occur within the same temporal window (±15 days) in the previous three years. Such changes are part of a regular and repeated interannual land use, and as such within normal near-term variability in vegetative cover. That said, very often these regular land changes may be shifted by more than 15 days between years and then they will appear within DIST-ALERT.

The algorithms of DIST-ALERT require accurate detection of land observations uncontaminated by cloud, haze, shadow, or snow/ice. If contaminated pixels pass through to the disturbance algorithms then these masking errors propagate through the product. For example, when pixels with cloud cover are not masked out, the vegetation fraction model will be applied to that pixel and result in a low vegetation cover estimate. If these pixels are over a vegetated area then they will be marked as vegetation loss. Tracking these pixels through subsequent observations can mitigate these commission errors as cloud omission errors are not likely to regularly repeat over the same pixels, in which case they would be removed from the status layers. Cloud commission errors can also propagate omission errors in DIST-ALERT. Some bright targets such as the white roofs of buildings, are regularly flagged as cloud and these pixels then never pass through to the DIST-ALERT disturbance algorithms. For now, the product is relying on the identification of cloud, shadow, and snow/ice provided within the Fmask layer of the HLS input.

Although measuring vegetation cover is beyond the scope of the DIST product, the auxiliary vegetation cover indicator layers that are used by the internal models for identifying areas of disturbance can be used for additional correlative analysis directly. For example, the maximum vegetation anomaly can be harnessed to threshold vegetation cover loss at a higher sensitivity (i.e., loss smaller than 50%) and the current vegetation cover indicator can be tracked over time to evaluate possible recovery trends. Constraints on the use of the vegetation cover indicator will be determined via validation efforts and result in guidance to users in applying the time-series vegetation cover estimates.

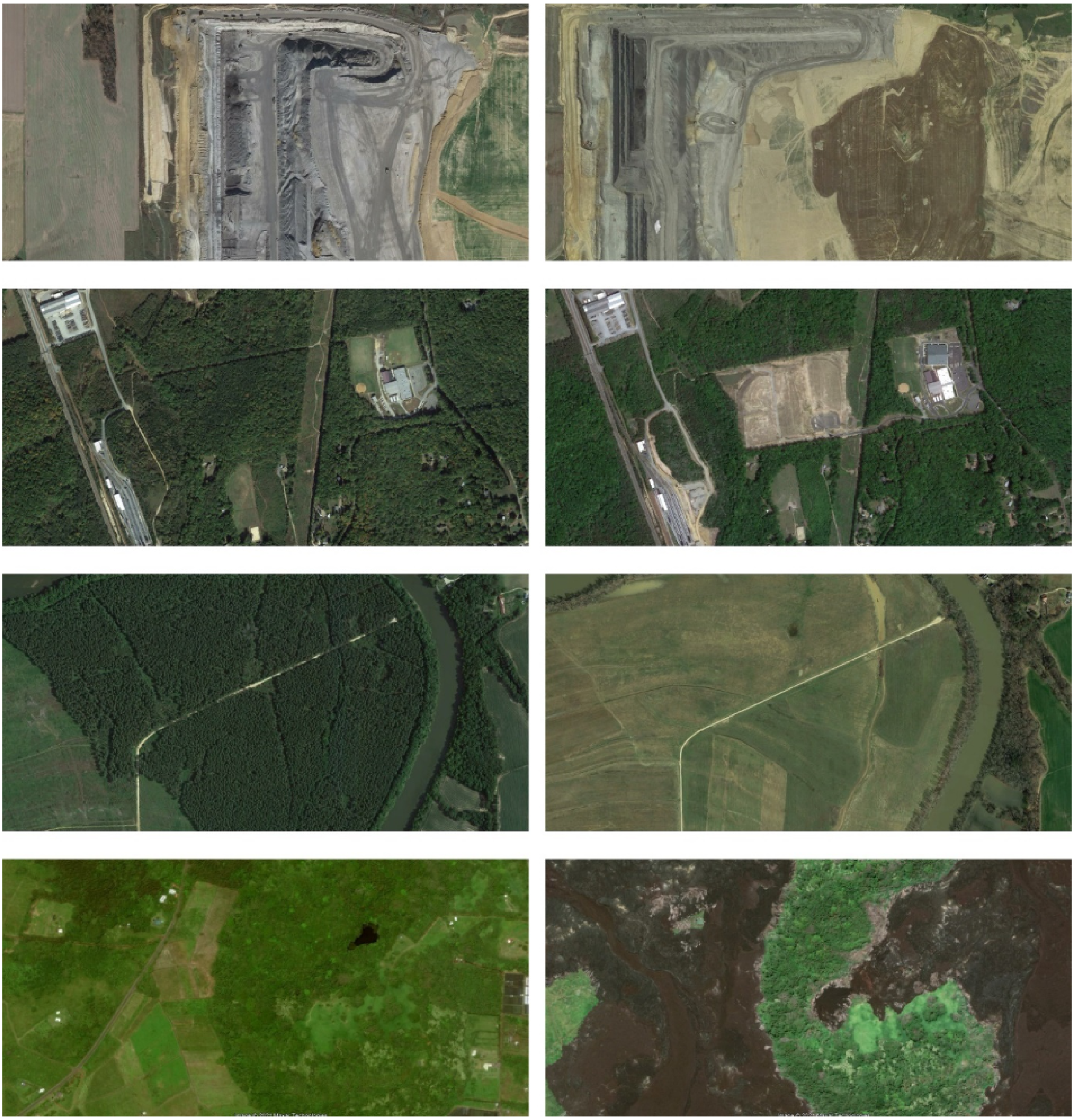

Figure 12: Example dynamics that will be detected by the DIST algorithm: from top to bottom, open pit coal mining in southwest Indiana (centered at 87.308W, 39.026N), commercial land use expansion in northern Virginia (77.433W, 38.215N), conversion of secondary forest to cropland in northern Alabama (85.796W, 34.124N), and loss of vegetation due to lava flow on Hawaii (154.842W, 19.500N). Imagery from GoogleEarth.

5. Performance Assessment

5.1. Validation Methods

Validation of DIST product will have two important categories of activities:

1. Validation of the disturbance detection layers in the DIST-ALERT and DIST-ANN products using high-resolution derived disturbance data.
2. Assessment activities related to the current vegetation cover indicator within the DIST-ALERT product.

The first category will be used to determine the accuracy of the vegetation disturbance. The second category is related to reporting the statistical relationship of the intermediate vegetation cover indicator layer used by the disturbance algorithm and in-situ vegetation cover as determined by field work.

To validate both the DIST-ALERT and the DIST-ANN we will employ a stratified random sample of reference data in a manner similar to that of the study of (Ying et al., 2017). The global population of all 30m pixels aligned with the HLS pixel grid will be stratified based on disturbance presence and an equal area sample of 30m pixels will be selected. For each pixel randomly selected as a validation site, 3m PlanetScope data will be employed to create reference data of disturbance status (***Figure 13***) and the initial date of disturbance events, where feasible. This reference data will then be compared to the selected pixel in the DIST-ALERT time-series and the DIST-ANN product to ensure the requirement is met. The reference data for each validation site are created by mapping or visually interpreting 30m HLS pixel footprints from time-series high-resolution PlanetScope data. An analyst marks all time steps of the high-resolution data that have a ≥50% loss of vegetation cover, resulting in a yes/no reference time-series of ≥50% vegetation cover loss. The analyst will also mark the initial date of any discernible loss events <50% or whether there was no disturbance.

For vegetation disturbance validation, per the requirements, the vegetation disturbances with anomalies ≥50% mapped in the DIST-ALERT will be considered disturbed and all other land observations will be considered no-disturbance. As the DIST-ALERT product is released at the cadence of the HLS input dataset, the accuracy of the DIST-ALERT product is calculated from all HLS time-steps with respect to the reference disturbance time-series so that the reported accuracy will apply to all observations and thus all phenological stages. The DIST-ALERT product will be compared to the reference time-series by matching the respective dates. There is an associated date for each disturbance/no-disturbance label in the reference time-series which corresponds to the time of the high-resolution Planet acquisitions. For each HLS acquisition date, DIST-ALERT will be compared with the reference data label from the same date when possible. For HLS dates without a coincident reference label, if the closest preceding and following reference date have the same label, this is compared with the DIST-ALERT label. If the two reference dates have different labels this time step is excluded as it is unknown when the disturbance occurred between these two dates. Given that the validation assessment covers an entire year with all HLS acquisition dates evaluated, all seasons are assessed and any errors due to intra-annual variation will be quantified.

Figure 13: Example validation data from ~3-4m PlanetScope imagery over a residential expansion site near Dallas, Texas, centered at 32.654N,97.500W. From left to right, image from 12-31-2019, image from 12-29-2020, and mapped vegetation loss in black overlay. All cloud-free PlanetScope data (not shown here) will be used to refine data of disturbance estimation. Sample 30m HLS pixel shown in red outline exhibits vegetation loss.

The DIST-ALERT has two types of disturbances: provisional and confirmed. Both types of disturbances in DIST-ALERT will be evaluated against the high-resolution derived time-series. Typically, the confirmed disturbances in the DIST-ALERT product are expected to have a higher accuracy as they have been repeatedly observed. However, given that for minimum latency DIST-ALERT products may be used as soon as HLS scenes are characterized and before a new disturbance can be confirmed from repeat observations, we will also validate provisional alerts. The overall user's and producer's accuracies of both products will be reported and provide a globally representative measure of DIST product performance.

The assessment of the vegetation cover indicator layer pertains to the vegetation layers of the DIST suite used as inputs to the disturbance detection algorithms and distributed with the DIST product. Although this layer is without a formal requirement, providing a general assessment of this intermediate layer's utility is valuable for users and for increasing the transparency of the disturbance detection algorithm. The assessment uses 3 by 3 pixel grids of vegetation cover derived from field data collected contemporaneously with a Landsat 8 or Sentinel 2 overpass. Each grid is aligned with the DIST-ALERT product pixels and is produced using sub-meter maps from field data collected using a drone-based multi-spectral sensor, with example data shown in ***Figure 14***. The field data are associated with a given date and compared to the DIST-ALERT from the same day as the field work. Specifically, the vegetation indicator layer is compared to the co-located 3 by 3 grid of vegetation cover derived from the field work collected during different seasons. A comparison of the vegetation indicator layer with respect to field reference maps will be reported, including RMSE and R^2.

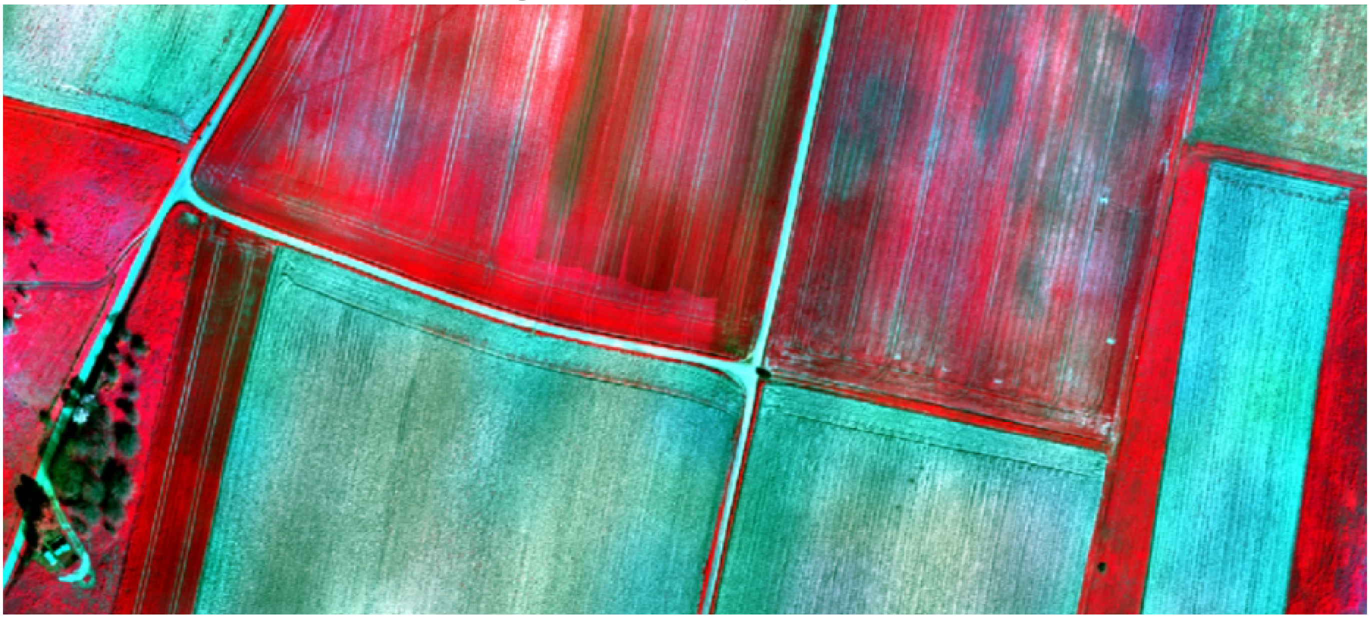

Figure 14: Example 4cm NIR-RED-GREEN imagery over an agricultural landscape using a WingtraONE Gen II drone with a Mica-Sense Red Edge-MX multi-spectral camera. Each sample site will characterized into yes/no vegetation cover, for both photosynthetically active and photosynthetically and non-photosynthetically active data.

5.2. Uncertainties

Confusion matrices of yes/no disturbance accuracies and associated uncertainties will be calculated from the global sample of reference data as compared to instantaneous DIST-ALERT and annual summary DIST-ANN data similar to the method outlined by (Ying et al., 2017). Overall accuracy for the disturbances ≥50% of DIST-ALERT is specified to exceed 80% and of DIST-ANN, 90% and both have an overall accuracy of 99%. For vegetation cover, we will report correlation measures for our opportunistically acquired drone-based field data.

5.3. Validation Errors

Errors will be assessed and assigned to categories. Errors of omission, for example disturbances unrelated to vegetation loss, will be quantified and their overall contribution to error calculated, justifying or not the use of a complementary/back-up spectral distance algorithm. Both omission and commission errors will be aggregated by climate domain/ecozone and disturbance type. In this way, users will know which applications may be more readily supported by the DIST products.

6. Algorithm Implementation

6.1. Algorithm Availability

github.com/gladumd/OPERA_DIST/

The implementation codes of the algorithm are open to the public on GitHub.

6.2. Input Data Access

https://search.earthdata.nasa.gov/search?q=HLS

The DIST product employs single-date HLS tiles, each processed for all valid land observations. HLS data access is through the Land Processes Distributed Active Archive Center (LPDAAC). Daily input volumes range from 0.5-2.0Tb and processed outputs are estimated to be 0.5Tb daily.

6.3. Output Data Access

https://search.earthdata.nasa.gov/search?q=OPERA%20HLS%20alert

Final product layers are available through the LPDAAC. Users can access the data product through NASA's Earthdata Search.

7. Contact Details

Matthew , Hansen

URL: https://glad.umd.edu/team/matthew-hansen
Contact mechanism: Email: mhansen@umd.edu
Role(s) related to this ATBD: Supervision, Writing – original draft, Writing – review & editing
Affiliation: University of Maryland, Department of Geographical Sciences, Global Land Analysis and Discovery (GLAD) laboratory

Amy , Pickens

URL: https://glad.umd.edu/team/amy-hudson-pickens
Contact mechanism: Email: ahudson2@umd.edu
Role(s) related to this ATBD: Writing – original draft, Writing – review & editing, Corresponding Author
Affiliation: University of Maryland, Department of Geographical Sciences, Global Land Analysis and Discovery (GLAD) laboratory

Zhen, Song

URL: https://glad.umd.edu/team/zhen-song
Contact mechanism: Email: zhensong@umd.edu
Role(s) related to this ATBD: Writing – review & editing, Validation
Affiliation: University of Maryland, Department of Geographical Sciences, Global Land Analysis and Discovery (GLAD) laboratory

8. References

Adams, J.,B., Sabol, D.,E., Kapos, V., Almeida Filho, R., Roberts, D.,A., Smith, M.,O. & Gillespie, A.,R. (1995). Classification of multispectral images based on fractions of endmembers: Application to land-cover change in the Brazilian Amazon. *Remote sensing of Environment, 52(2),* 137--154.

Carroll, M., Townshend, J., Hansen, M., DiMiceli, C., Sohlberg, R. & Wurster, K. (2010). MODIS vegetative cover conversion and vegetation continuous fields. 725--745.

Claverie, M., Ju, J., Masek, J.,G., Dungan, J.,L., Vermote, E.,F., Roger, J., Skakun, S.,V. et al. (2018). The Harmonized Landsat and Sentinel-2 surface reflectance data set. *Remote sensing of environment, 219,* 145--161.

Davies, D.,K., Ilavajhala, S., Wong, M.,M. & Justice, C.,O. (2008). Fire information for resource management system: archiving and distributing MODIS active fire data. *IEEE Transactions on Geoscience and Remote Sensing, 47(1),* 72--79.

DeFries, R., Hansen, M., Steininger, M., Dubayah, R., Sohlberg, R. & Townshend, J. (1997). Subpixel forest cover in central Africa from multisensor, multitemporal data. *Remote Sensing of Environment, 60(3),* 228--246.

Foley, J.,A., DeFries, R., Asner, G.,P., Barford, C., Bonan, G., Carpenter, S.,R., Chapin, F.,S. et al. (2005). Global Consequences of Land Use. *Science, 309(5734),* 570--574.

Gitelson, A.,A., Kaufman, Y.,J., Stark, R. & Rundquist, D. (2002). Novel algorithms for remote estimation of vegetation fraction. *Remote Sensing of Environment, 80(1),* 76--87. https://doi.org/10.1016/S0034-4257(01)00289-9

Hansen, M., DeFries, R., Townshend, J., Carroll, M., Dimiceli, C. & Sohlberg, R. (2003). Global percent tree cover at a spatial resolution of 500 meters: First results of the MODIS vegetation continuous fields algorithm. *Earth Interactions, 7(10),* 1--15.

Hansen, M., DeFries, R., Townshend, J., Sohlberg, R., Dimiceli, C. & Carroll, M. (2002). Towards an operational MODIS continuous field of percent tree cover algorithm: examples using AVHRR and MODIS data. *Remote Sensing of Environment*, *83(1-2)*, 303--319.

Hansen, M., Egorov, A., Potapov, P., Stehman, S., Tyukavina, A., Turubanova, S., Roy, D.,P. et al. (2014). Monitoring conterminous United States (CONUS) land cover change with web-enabled Landsat data (WELD). *Remote sensing of Environment*, *140*, 466--484.

Hansen, M.,C., Krylov, A., Tyukavina, A., Potapov, P.,V., Turubanova, S., Zutta, B., Ifo, S. et al. (2016). Humid tropical forest disturbance alerts using Landsat data. *Environmental Research Letters*, *11(3)*, 034008.

Hansen, M.,C., Potapov, P.,V., Moore, R., Hancher, M., Turubanova, S.,A., Tyukavina, A., Thau, D. et al. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, *342(6160)*, 850--853.

Jiang, Z., Huete, A.,R., Chen, J., Chen, Y., Li, J., Yan, G. & Zhang, X. (2006). Analysis of NDVI and scaled difference vegetation index retrievals of vegetation fraction. *Remote Sensing of Environment*, *101(3)*, 366--378. https://doi.org/10.1016/j.rse.2006.01.003

Kates, R., Turner II, B. & Clark, W. (1990). The Great Transformation in The Earth as Transformed by Human Action. 1-17.

Lenton, T.,M., Held, H., Kriegler, E., Hall, J.,W., Lucht, W., Rahmstorf, S. & Schellnhuber, H.,J. (2008). Tipping elements in the Earth's climate system. *Proceedings of the national Academy of Sciences*, *105(6)*, 1786--1793.

Li, J. & Roy, D.,P. (2017). A global analysis of Sentinel-2A, Sentinel-2B and Landsat-8 data revisit intervals and implications for terrestrial monitoring. *Remote Sensing*, *9(9)*, 902.

Mildrexler, D.,J., Zhao, M. & Running, S.,W. (2009). Testing a MODIS global disturbance index across North America. *Remote Sensing of Environment*, *113(10)*, 2103--2117.

Moffette, F., Alix-Garcia, J., Shea, K. & Pickens, A.,H. (2021). The impact of near-real-time deforestation alerts across the tropics. *Nature Climate Change*, *11(2)*, 172--178.

Nepstad, D., McGrath, D., Stickler, C., Alencar, A., Azevedo, A., Swette, B., Bezerra, T. et al. (2014). Slowing Amazon deforestation through public policy and interventions in beef and soy supply chains. *science*, *344(6188)*, 1118--1123.

Roberts, D.,A., Gardner, M., Church, R., Ustin, S., Scheer, G. & Green, R. (1998). Mapping chaparral in the Santa Monica Mountains using multiple endmember spectral mixture models. *Remote sensing of environment, 65(3),* 267--279.

Ross, K., Brown, M., Verdin, J.,P. & Underwood, L. (2009). Review of FEWS NET biophysical monitoring requirements. *Environmental Research Letters, 4(2),* 024009.

Settle, J. & Drake, N. (1993). Linear mixing and the estimation of ground cover proportions. *International Journal of Remote Sensing, 14(6),* 1159--1177.

Shimabukuro, Y.,E., Santos, J.,a., Formaggio, A., Duarte, V., Rudorff, B., Achard, F. & Hansen, M. (2012). The Brazilian Amazon monitoring program: PRODES and DETER projects. *Global forest monitoring from earth observation*, 153--169.

Shukla, P.,R., Skeg, J., Buendia, E.,C., Masson-Delmotte, V., Portner, H., Roberts, D., Zhai, P. et al. (2019). Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems.

Slough, T., Kopas, J. & Urpelainen, J. (2021). Satellite-based deforestation alerts with training and incentives for patrolling facilitate community monitoring in the Peruvian Amazon. *Proceedings of the National Academy of Sciences, 118(29),* e2015171118.

Song, X., Hansen, M.,C., Stehman, S.,V., Potapov, P.,V., Tyukavina, A., Vermote, E.,F. & Townshend, J.,R. (2018). Global land change from 1982 to 2016. *Nature, 560(7720),* 639--643.

Tucker, C.,J. & Nicholson, S.,E. (1999). Variations in the Size of the Sahara Desert from 1980 to 1997. *Ambio, 28(7),* 587--591.

Ying, Q., Hansen, M.,C., Potapov, P.,V., Tyukavina, A., Wang, L., Stehman, S.,V., Moore, R. et al. (2017). Global bare ground gain from 2000 to 2012 using Landsat imagery. *Remote Sensing of Environment, 194,* 161--176.

Ying, Q., Hansen, M.,C., Sun, L., Wang, L. & Steininger, M. (2019). Satellite-detected gain in built-up area as a leading economic indicator. *Environmental Research Letters, 14(11),* 114015.

Zhu, Z. & Evans, D.,L. (1994). US forest types and predicted percent forest cover from AVHRR data. *PE & RS- Photogrammetric Engineering & Remote Sensing, 60(5)*, 525--531.

MODIS VCF ATBD.