2021 Harvest Area Remote Sensing-Based Spectral Regeneration

Metadata & Technical Documentation

Alberta Biodiversity Monitoring Institute, Geospatial Centre

Version 1.0 | June, 2024

ABMI ALBERTA BIODIVERSITY MONITORING INSTITUTE

Table of Contents

1. Overview

1.1. Summary

The Alberta Biodiversity Monitoring Institute's (ABMI's) 2021 Harvest Area Spectral Regeneration dataset provides a remote sensing-based characterization of vegetation regrowth in relevant harvest area polygons contained within the ABMI's Human Footprint Inventory (HFI) [1]. Its intent is to represent the status and trends of post-harvest regeneration as seen through changes in spectral signals detected from the Earth's land surface.

1.2. Description

This dataset is provided in the form of a vector layer containing harvest area polygons from the ABMI's HFI 2021 [1], the original attributes that accompany the latter dataset, and an additional series of attributes containing metrics and other information related to remote sensing-based spectral regeneration. According to the ABMI HFI dataset a harvest area is an area "where forestry operations have occurred (clear-cut, selective harvest, salvage logging, etc." [1], and spectral regeneration is here defined as the level and rate at which the spectral signature of the land surface, as measured through a spectral vegetation index, has returned to its pre-disturbance level. Such metrics exist for the roughly 73,000 (29%) of the harvest areas identified over Alberta by the HFI 2021, for which such information could be reliably and confidently extracted (i.e., they were harvested within the appropriate time period, and minimal noise or other interference is present in their spectral signals). The harvest areas possessing spectral regeneration data are distributed widely across Alberta and we believe they present a good representation of the various landscapes and regions of the province. Spectral regeneration data are generated using a multi-decadal time series of Landsat Earth Observation satellite imagery, covering the province of Alberta for the years 1984 through 2021. The imagery is processed and analyzed using the Google's online Earth Engine platform. Further details can be found in Hird et al. [2].

1.3. Credits

This dataset was developed and generated by the ABMI's Geospatial Centre.

1.4. Citation

Suggested citation for this data product:

> *Alberta Biodiversity Monitoring Institute. 2024. Harvest Area Spectral Regeneration* 2021 (Version 1.0). Shapefile. Edmonton, Alberta, Canada. Last modified May 30, 2024.

Suggested citation for this documentation:

> *Alberta Biodiversity Monitoring Institute. 2024. "2021 Remotely Sensed Harvest Area Spectral Regeneration – Metadata and Technical Documentation, Version 1.0." Edmonton, Alberta, Canada.*

1.5. List of Updates

Updates to the Data

Date	Version	Description
May 30, 2024	1.0	First published

Updates to the Documentation

Date	Version	Section(s)	Description
May 30, 2024	1.0	All	First published

1.6. Contact Information

If you have questions or concerns about the data, please contact:
Geospatial Centre
Alberta Biodiversity Monitoring Institute
CW 405 Biological Sciences Centre
University of Alberta
Edmonton, Alberta, Canada, T6G 2E9
Email: abmiinfo@ualberta.ca

1.7. Keywords

Alberta, harvest area, forest spectral regeneration, remote sensing, Landsat, time series analysis, LandTrendr, cloud computing, Google Earth Engine

2. Use Limitations

This product was developed and produced using freely-available, open-source Landsat 5, 7 and 8 satellite imagery, and the ABMI Human Footprint Inventory (2021). The 2021 Harvest Area Remote Sensing-Based Spectral Regeneration dataset may be freely used provided it is cited properly (see the Citation section above).

2.1. Open-Sourced Data

This dataset contains data originating from open sources, which has subsequently been enhanced through computer analysis processing. The Open Sourced Data may be reproduced in whole or in part and in any form for educational, data collection or non-profit purposes without special permission from the ABMI provided acknowledgement of the source is made. No use of the Open Sourced Data may be made for resale without prior permission in writing from the ABMI. By accessing the Open Sourced Data, you agree to indemnify and hold harmless the ABMI and the ABMI's subsidiaries, affiliates, related parties, officers, directors, employees, agents, independent contractors, advertisers, partners, co-branders, and Open Sourced Data sources from any and all actions, proceedings, claims, demands, liabilities, losses, damages, and expenses which may be brought against or suffered by the ABMI or which it may sustain, pay or incur, arising or resulting from your violation of this clause. The Open Sourced Data is provided on an "As Is" and "As Available" basis and the ABMI does not guarantee that the Open Sourced Data will be suitable for your purposes or requirements. The ABMI further states that the Open Sourced Data is subject to change, and the ABMI gives no guarantee that the content is complete, accurate, error or virus free, or up to date. The ABMI disclaims all warranties, conditions, and other terms of any kind, whether express or implied, whether in contract, tort (including liability for negligence) or otherwise, including, but not limited to any implied term of satisfactory quality, fitness for a particular purpose, and any standard of reasonable care and skill.

2.2. Exclusive ABMI-Sourced Data

This dataset contains data created by the ABMI through active visual interpretation and computer processing. The ABMI Sourced Data may be reproduced in whole or in part and in any form for educational, data collection or non-profit purposes without special permission from the ABMI provided acknowledgement of the source is made. No use of the ABMI Sourced Data may be made for resale without prior permission in writing from the ABMI. By accessing the ABMI Sourced Data, you agree to indemnify and hold harmless the ABMI and the ABMI's subsidiaries, affiliates, related parties, officers, directors, employees, agents, independent contractors, advertisers, partners, and co-branders, from any and all actions, proceedings, claims, demands, liabilities, losses, damages, and expenses which may be brought against or suffered by the ABMI or which it may sustain, pay or incur, arising or resulting from your violation of this clause. The ABMI Sourced Data is provided on an "As Is" and "As Available" basis and the ABMI does not guarantee that the ABMI Sourced Data will be suitable for your purposes or requirements. The ABMI further states that the ABMI Sourced Data is subject to change, and the ABMI gives no guarantee that the content is complete, accurate, error or virus free, or up to date. The ABMI disclaims all warranties, conditions, and other terms of any kind, whether express or implied, whether in contract, tort (including liability for negligence) or otherwise, including, but not limited to any implied term of satisfactory quality, fitness for a particular purpose, and any standard of reasonable care and skill.

2.3. Use of Multiple Data Versions

We strongly recommend caution when considering combining the use of multiple versions of the Harvest Area Spectral Regeneration dataset together (e.g., 2019, 2020 and/or 2021 together). Variations in the input polygons and in the spectral metrics between the final versions result from a variety of sources, and do not necessarily reflect changes in the vegetative community on the land surface. We recommend that users rely on the most recent version available, and do not recommend using data from multiple versions together in order to examine year to year changes. Please see section 5.2 for more details.

3. Data Specifications

3.1. Processing Environment

The Google Earth Engine online code editor [3]; ESRI ArcGIS Pro 3.2.1; R 4.2.0; RStudio Version 2023.12.0, Build 369

3.2. Extents

Latitude and Longitude
West: -120
East: -110
South: 49
North: 60
Projection: Alberta Environment & Parks 10TM, NAD83, 'Forest' North: 6650732.874 metres (m)
South: 5425911.945 m
East: 850578.966 m
West: 179312.099 m

3.3. Spatial Reference

Projected Coordinate System: NAD_1983_10TM_AEP_Forest WKID: 3400; Authority: EPSG
Projection: Transverse Mercator False Easting: 500000.00000000
False Northing: 0.00000000
Central Meridian: -115.00000000
Scale Factor: 0.99920000 Latitude of Origin: 0.00000000 Linear Unit: Meter
Geographic Coordinate System: GCS_North_American_1983
Prime Meridian: Greenwich Angular Unit: Degree
Datum: D_North_American_1983 Spheroid: GRS_1980
Semi-major Axis: 6378137.0
Semi-minor Axis: 6356752.314140356
Inverse Flattening: 298.257222101

3.4. Data Format

These data are provided as an ESRI Shapefile (.shp), containing relevant ABMI 2021 HFI harvest area polygons and associated attributes, with the addition of a set of spectrally-based, regeneration-related attributes.

3.5. Lineage

This dataset is an updated version, to the year 2021, of the publicly available spectral regeneration dataset provided by the ABMI (previous versions: to 2019 and 2020) and accessible through the ABMI website (www.abmi.ca). The dataset may be further updated or replaced with a new version in future where relevant (e.g., when important changes, improvements, or additions are made to the data).

3.6. Attribute Fields

Table 1 summarizes the list of attributes/fields found in the ABMI 2021 Harvest Area Spectral Regeneration dataset (provided in ESRI's Shapefile format). For detailed information regarding the generation of the regeneration attributes, see the Methods section below. For detailed information on HFI feature attributes, refer to the HFI 2021 metadata documentation [1].

Table 1. List of attribute fields found within the 2021 Harvest Area Spectral Regeneration dataset. For further details, see Section 3 (Methods) below.

Field/Attribute	Possible Values	Description
HFI feature attributes		
OBJECTID	1 to 300000	Unique polygon object identification number; specific to the 2021 HFI dataset
SOURCE	See [1]	ABMI's data source for a harvest area polygon
HFI_ID	See [1]	An alpha-numeric identifier for this polygon feature
FEATURE_TY	See [1]	The type or category of human footprint feature in the ABMI HFI (e.g., harvest area)
YEAR	See [1]	Year integer value representing a feature's "year of origin", either brought in as part of a reference data
Modifier_Year	See [1]	Indicates the year in which a feature's type was modified from one feature type to another feature type within the same sublayer
Spectral regeneration metric attributes		
regnAnlyYN	Y (Yes), N (No)	Indicates whether this harvest area polygon was included in analyses (does not indicate whether spectral regeneration metrics were reliably or appropriately extracted)
regnMetsYN	Y, N	Indicates whether spectral regeneration metrics were reliably and appropriately extracted for this harvest area
AnlysID	1 to 300000	Unique identifier used in analyses of harvest area polygons, for all harvest area polygons analyzed
preNBR_m	-1000.0 to 1000.0	Pre-harvest normalized burn ratio (NBR) spectral vegetation index value (mean*). Scaled by 1000.
preNBR_s	-1000.0 to 1000.0	Pre-harvest spectral NBR value (standard deviation (sd)**)
hrvYr_m	1989.0 to 2013.0	Year in which the harvest event is detected (mean)
hrvYr_s	0 to 3.0	Year in which the harvest event is detected (sd)
lnDstb_m	1.0 to 24.0	Length of time (years) between the detected harvest event and the beginning of regeneration (mean)
lnDstb_sd	0 to 10.0	Length of time (years) between the detected harvest event and the beginning of regeneration (sd)

Field/Attribute	Possible Values	Description
regStYr_m	1990.0 to 2013.0	Year in which detectable post-harvest spectral regeneration begins (mean)
regStYr_s	0 to 10.0	Year in which detectable post-harvest spectral regeneration begins (sd)
nbrDstb_m	0 to 2000.0	Total detected change in NBR values detected at harvest event (mean)
nbrDstb_s	0 to 300.0	Total detected change in NBR values detected at harvest event (sd)
reg5yr_m	0 to 100.0+	Percent spectral regeneration 5 years after regeneration has begun (mean)
reg5yr_s	0 to 50.0	Percent spectral regeneration 5 years after regeneration has begun (sd)
y2reg80_m	0 to 30.0	Length (years) of time required to reach 80% spectral regeneration (mean)
y2reg80_s	0 to 15.0	Length (years) of time required to reach 80% spectral regeneration (sd)
reg2021_m	0 to 100.0+	Current (to 2021) level of percent spectral regeneration (mean)
reg2021_s	0 to 90.0	Current (to 2021) level of percent spectral regeneration (sd)
totPolyPix	>= 0	Total number of Landsat 30 m pixels representing the pre-processed (i.e., buffered, simplified) harvest area polygon
Data quality flag attributes		
perRelvPix	0 to 100.0	Percent of total intersecting pixels that were appropriate, relevant, and retained for use in metric calculations
perOutRng	0 to 100.0	Percent of intersecting pixels flagged for 'out of date range'
perNoRegn	0 to 100.0	Percent of intersecting pixels flagged for 'no regeneration detected'
perMltDstb	0 to 100.0	Percent of intersecting pixels flagged for 'multiple disturbances detected'
perNoHrv	0 to 100.0	Percent of intersecting pixels flagged for 'no harvest events detected'
Confidence score attributes ᵻ		
confSz	0 to 6	Confidence score based on size of harvest area polygon (greater confidence is given to larger harvest areas as they are represented by a larger sample of pixels)
confRelvPx	0 to 6	Confidence score based on percentage of representative pixels used in metric calculations (i.e., not flagged and removed from analyses)
confCntgPx	0 to 6	Confidence score based on the number of contiguous pixels used in metric calculations
confHrvYr	0 to 6	Confidence score based on within-polygon variability in the detected year of harvest event
confLnDstb	0 to 6	Confidence score based on within-polygon variability in the length of time between the detected harvest event and detected beginning of regeneration
confNBRchg	0 to 6	Confidence score based on within-polygon variability in NBR total disturbance values
confRegn	0 to 6	Confidence score based on within-polygon variability in current levels of percent spectral regeneration
confY2R80	0 to 6	Confidence score based on within-polygon variability in the years required to reach 80% spectral regeneration
conf5yReg	0 to 6	Confidence score based on within-polygon variability in 5-year post-harvest spectral regeneration
confTotSum	0 to 54	Overall confidence score; a sum of all calculated confidence scores

* Mean: mean of metric values from all relevant/appropriate pixels intersecting the harvest area polygon of interest

** Standard deviation: mean of metric values from all relevant/appropriate pixels intersecting the harvest area polygon of interest

ᵻ Confidence scores range from 0 (very low) to 6 (very high)

3.7. No Data Values

No Data or Null values are filled with a value of -9999 where a particular metric or attribute was not calculable. This is the case for those harvest area polygons for which metrics were calculated and are provided, or for those that were analyzed but not appropriate for reporting spectral regeneration. For instance, where all relevant pixels representing a harvest area contained the same year of harvest event, the standard deviation of these values is not calculable, and is given a value of -9999. As another example, where spectral signals did not reach 80% spectral regeneration, these metrics are given a no data value of -9999. Metric and related attributes for harvest areas that were not analyzed are given a value of zero.

4. Methods

The following provides a brief summary of the methods used to produce the harvest area spectral regeneration dataset. These are described further in the peer-reviewed paper published by Hird et al. [2]. The ABMI 2021 HFI harvest area polygons were pre-processed before being brought into the Google Earth Engine (GEE) analysis environment. They were first negatively buffered by 30 m (i.e., the outer 30 metres of each polygon was removed), so as to minimize edge effects resulting from any misalignment between the polygons themselves and the satellite imagery that has a 30 m spatial resolution.

These outlines were then simplified (with a maximum change tolerance of 15 m, or half the pixel width of the satellite imagery used in this workflow). This enabled efficient uploading of the polygon features into the analysis environment. Finally, those individual polygons < 900 m2 in size – the size of one Landsat image pixel (30 m x 30 m) – were removed. This pertained to both polygons that represented a single harvest area feature, and those that represented a disjointed piece of a larger harvest area feature (e.g., that resulted from negative buffering). Once processed, a total of 192,589 processed harvest area polygons from the HFI 2021 dataset were brought into the GEE environment for analysis, which equals 76.7% of the 250,954 harvest area features in the original dataset.

4.1. Landsat Data Processing

Figure 1 illustrates the workflow implemented to generate this dataset. The majority of the methodology is undertaken with the GEE online platform, using a customized script written with the help of the JavaScript-based GEE code editing application programming interface.

Tier 1 processed surface reflectance imagery from Landsat 5 Thematic Mapper, 7 Enhanced Thematic Mapper +, and 8 Operational Land Imager are first calibrated and masked for cloud and cloud shadow using provided quality flags (see the GEE Data Catalog for more information: https://developers.google.com/earth-engine/datasets). They are then integrated into a single time series stack of growing-season images (i.e., June through September) covering 1984 to 2021. This image stack is processed to produce yearly best pixel composites using per-pixel median compositing. Per-pixel time series of a calculated spectral vegetation index derived from this composited dataset form the foundation of the ABMI’s remotely-sensed characterization of post-harvest spectral regeneration.

Equation (1) shows the calculation used for the spectral vegetation index employed here - an index commonly referred to as the Normalized Burn Ratio or NBR - which has been shown in published research to work well for detecting and characterizing forest vegetation disturbances and regeneration.

$$NBR = \frac{\mathrm{NIR}_{\mathrm{reflectance}} - \mathrm{SWIR}_{\mathrm{reflectance}}}{\mathrm{NIR}_{\mathrm{reflectance}} + \mathrm{SWIR}_{\mathrm{reflectance}}} \quad (1)$$

where NIR is near infrared, and SWIR is shortwave infrared [4].

The equation results in unitless values ranging from -1 to 1 which are often scaled up by 1000 for data handling.per-pixel time series of annual, growing-season NBR values are processed using the LandTrendr algorithm – a temporal segmentation method designed to extract changes in surface vegetation conditions from time series of remotely-sensed spectral values using a series of linear trend segments fit to a time series [5]. The resulting segmented time series are then ready for extracting information related to detectable harvest events and spectral regeneration.

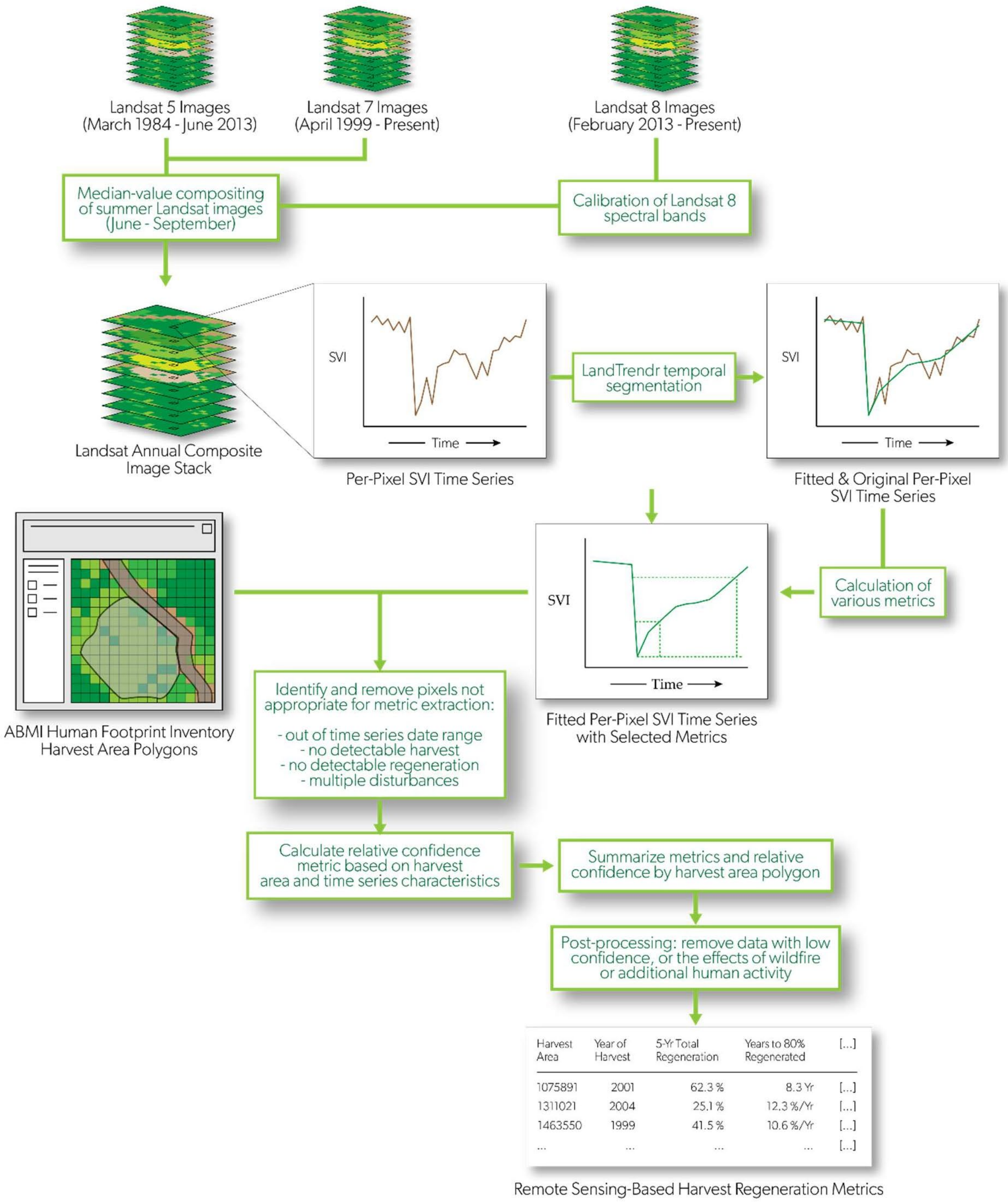

Figure 1. General workflow used for generating per-harvest area polygon metrics related to remotely- sensed spectral regeneration. SVI corresponds to *spectral vegetation index*, which in this work is the normalized burn ratio.

4.2. Characterizing Post-Harvest Spectral Regeneration

Table 1 in section 3.6 (Attribute Fields) describes the metrics contained within the harvest area spectral regeneration dataset. It should be noted that these were identified as most relevant from those tested during development, based on information contained in the scientific literature, as well as observations of the data. Alongside these metrics, related quality information extracted for each corresponding harvest area polygon are provided. Regeneration-specific metrics are generally given as percent spectral regeneration, which refers to the percent of the total drop in NBR values (occurring with the detected harvest event) that has been regained – i.e., the percent of the spectral signature that has returned. Figure 2 illustrates how these metrics are calculated.

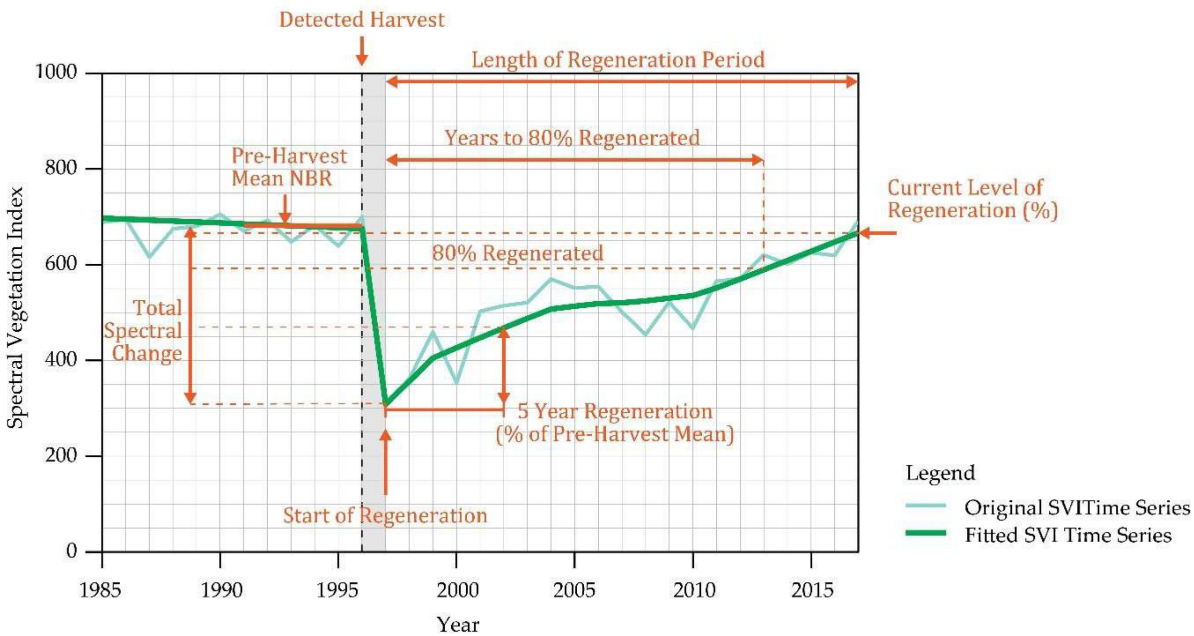

Figure 2. Graphic illustrating how various metrics and information related to post-harvest spectral regeneration are extracted from a per-pixel spectral vegetation index time series.

The pre-processed 2021 HFI harvest area polygons are brought into the GEE environment, and per-polygon metric statistical summaries (mean and standard deviation) are calculated for each harvest area. As indicated in Table 1, pixels for which extracting regeneration metrics is either infeasible or inappropriate are flagged and removed from further analysis. That is, these flagged pixels are not included in per-polygon summaries. The conditions under which a pixel is flagged and subsequently removed are detailed in Table 2.

Table 2. Description of conditions under which pixels are flagged for removal from further processing.

Flag	Description
No harvest	No inter-annual drop in NBR values beyond a certain magnitude is detected, indicating notable vegetation removal did not occur (e.g., pixels where retention during harvest was practiced)
Multiple disturbances	More than one inter-annual drop in NBR beyond a certain magnitude is detected, and separated by more than 3 years (indicating separate events)
Out of date range	A harvest was detected, but occurred too early or late in the time series for proper metric calculations; the current workflow requires data be available 5 years before harvest and 5 years after regeneration for calculations
No regeneration	A harvest or disturbance was detected, but NBR values do not increase post-disturbance, indicating no regeneration is occurring

4.3. Post-Processing

Harvest area polygons wherein > 50% of the pixels intersecting that polygon are flagged and removed, or where fewer than 9 pixels in total remain for metric calculations, are removed from the dataset.

These are judged to offer insufficiently reliable representations of spectral regeneration within the harvested area. The threshold of 9 pixels was used as it represents, under ideal conditions, a situation in which a block of pixels wherein the centre pixel is surrounded by other pixels representing the same spectral signatures. This will in theory help further minimize any remaining edge effects from areas adjacent to the harvest area polygon.

Those harvest areas overlapping other HFI human footprint features (e.g., mines, wellsites, cultivation), wherein this overlap constitutes more than 20% of their area, were also removed from the dataset so as to minimize the risk that spectral signals have been affected by other anthropogenic activities post-harvest. Visual inspection showed those harvest areas overlapped by other HFI features by less than 20%, were often overlapped by roads or wellsites – features we assume are not captured in our metric calculations due to the use of the previously-discussed flagging system.

The effects of wildfire on post-harvest spectral regeneration metrics were also minimized in post-processing. We removed harvest areas that were overlapped by wildfires in the Government of Alberta's most recent Wildfire Perimeter database (available at: https://wildfire.alberta.ca/resources/historical-data/spatial-wildfire-data.aspx), which had occurred within the 20 years prior to the detected harvest date or any time after the harvest date, and which occupied more than 10% of the harvest area feature's area. The 20-year threshold was chosen because in the majority of cases we observed spectral signals to have returned completely to pre-disturbance levels by this time after a single disturbance event (i.e., the spectral signal has generally saturated after 20 years).

Confidence scores designed to reflect various characteristics of per-harvest area polygon calculations were calculated and used to identify and either evaluate or further remove remaining polygon data that is judged to be of low confidence. Table 3 describes the set of confidence scores calculated for each harvest area polygon.

Table 3. Descriptions and thresholds used for confidence scores calculated for each harvest area polygon for which metrics were generated.

Confidence Factor	Description	Score	Confidence Level
Total Pixel count	<1 pixels	0	Very Low
	>=1 & <11 pixels	1	Low
	>=11 & <20 pixels	2	Medium-Low
	>=20 & <30 pixels	3	Medium
	>=30 < 100 pixels	4	Medium-High
	>=100 & <200 pixels	5	High
	>=200 pixels	6	Very High
Percent of pixels available for use in calculations	<50% available	0	Very Low
	>=50% & <60% available	1	Low
	>=60% & <70% available	2	Medium-Low
	>=70% & <80% available	3	Medium
	>=80% & <90% available	4	Medium-High
	>=90% & <100% available	5	High
	100% available	6	Very High
Number of contiguous (i.e., adjacent) pixels available for use in calculations	= 0 contiguous pixels	0	Very Low
	> 0 & <= 2 contiguous pixels	1	Low
	>2 & <=4 contiguous pixels	2	Medium-Low
	>4 & <=6 contiguous pixels	3	Medium
	>6 & <=8 contiguous pixels	4	Medium-High
	>8 & <=10 contiguous pixels	5	High
	>= 11 contiguous pixels	6	Very High
Within-polygon variability* in detected year of harvest	>=3 standard deviation (sd) of within-polygon year of detected harvest (YOH)	0	Very Low
	>=2.5 & <3 YOH sd	1	Low
	>=2 & <2.5 YOH sd	2	Medium-Low
	>=1.5 & <2 YOH sd	3	Medium
	>=1 & <1.5 YOH sd	4	Medium-High
	>=0.5 & <1 YOH sd	5	High
	<0.5 YOH sd	6	Very High
Within-polygon variability in length of time (years) between initial detected harvest and detected beginning of regeneration	>=3 standard deviation (sd) of within-polygon length of disturbance period (YOH)	0	Very Low
	>= 2 & <3 disturbance length sd	1	Low
	>=1.5 & <2 disturbance length sd	2	Medium-Low
	>=1 & <1.5 disturbance length sd	3	Medium
	>=0.75 & <1 disturbance length sd	4	Medium-High
	>=0.25 & <0.75 disturbance length sd	5	High
	<0.25 disturbance length sd	6	Very High
Within-polygon variability in total NBR spectral	>=200 standard deviation (sd) of total spectral	0	Very Low

Confidence Factor	Description	Score	Confidence Level
disturbance detected	>=150 & <200 total disturbance sd	1	Low
	>=125 & <150 total disturbance sd	2	Medium-Low
	>=115 & <125 total disturbance sd	3	Medium
	>=100 & <115 total disturbance sd	4	Medium-High
	>=75 & <100 total disturbance sd	5	High
	<75 total disturbance sd	6	Very High
Within-polygon variability in current levels of spectral regeneration	>=200 standard deviation (sd) of total spectral	0	Very Low
	>=150 & <200 total disturbance sd	1	Low
	>=125 & <150 total disturbance sd	2	Medium-Low
	>=115 & <125 total disturbance sd	3	Medium
	>=100 & <115 total disturbance sd	4	Medium-High
	>=75 & <100 total disturbance sd	5	High
	<75 total disturbance sd	6	Very High
Within-polygon variability in years required to reach 80% required to reach 80% spectral regeneration	>=3.5 standard deviation (sd) of years	0	Very Low
	>=3 & < 3.5 sd years to 80% regeneration	1	Low
	>=2.5 & < 3 sd years to 80% regeneration	2	Medium-Low
	>=2 & < 2.5 sd years to 80% regeneration	3	Medium
	>=1.5 & < 2 sd years to 80% regeneration	4	Medium-High
	>=0.5 & < 1.5 sd years to 80% regeneration	5	High
	< 0.5 sd years to 80% regeneration	6	Very High
Within-polygon variability in percent spectral regeneration at 5 years post-harvest	>= 30 standard deviation (sd) of percent spectral regeneration at 5 years	0	Very Low
	>=25 & < 30 sd 5-year regeneration	1	Low
	>=20 & < 25 sd 5-year regeneration	2	Medium-Low
	>=15 & < 20 sd 5-year regeneration	3	Medium
	>=10 & < 15 sd 5-year regeneration	4	Medium-High
	>=5 & < 10 sd 5-year regeneration	5	High
	< 5 sd 5-year regeneration	6	Very High

*Variability is evaluated using the within-polygon standard deviation of the metric in question

The confidence scores described in Table 3 were summed together into an overall confidence score for each harvest area polygon (up to a maximum score of 54). The total confidence score mean and standard deviation for the full, post-processed 2021 harvest area spectral regeneration dataset were calculated, and those polygons with a total confidence score three or more standard deviations below the mean were removed from the dataset. Those polygons with a harvest event year confidence score below 4 were also removed, as this suggests the pixels representing this polygon do not reflect a single, unified harvest event. This resulted in a total of 72,762 polygons within the 2021 Harvest Area Spectral Regeneration dataset that possess spectral regeneration metrics and information, which reflects 30.0% of the 250,954 harvest area polygons comprising this sublayer of the HFI.

5. Important Information for Data Usage

5.1. Comparing Current and Previous Datasets

As described in section 3.5 (Lineage) above, this dataset is the most current version in a series of Harvest Area Spectral Regeneration datasets produced by the ABMI. It is important to note that while the methods used to generate the current 2021 Harvest Area Spectral Regeneration dataset are the same as those used to generate the previous 2019 and 2020 versions, there are difference between the previous and current version of the dataset, including harvest areas where spectral regeneration metrics are available for a particular polygon in one of these versions but not the other. The following section assesses differences found between various versions of the Harvest Area Spectral Regeneration dataset, and describes likely causes.

Table 4 compares the different versions of the Harvest Area Spectral Regeneration datasets, listing the original numbers of HFI harvest area features, the numbers of features loaded into Google Earth Engine for analysis, and the numbers of features with spectral metrics in each version. A total of 55,350 harvest area features are both geometrically identical across the different existing datasets and possess spectral regeneration metrics in each of them. This represents 22.0% to 23.2% of the total number of harvest areas in the HFI product over the various years, and 76.1% to 95.7% of all harvest area polygons in the datasets that possess spectral metrics and information.

Table 4. Comparisons of valid metrics for 2019, 2020, and 2021 harvest area spectral regeneration layer based on preNBR_m and/or RegnMetsYN. We assume that a feature has valid metrics if it has preNBR_m value that is not 0, null, or -9999, and/or has a RegnMetsYN attribute value of "Y"..

Numerical Comparison	Dataset Year		
	2019	2020	2021
Total no. HFI harvest area features	238,102	244,294	250,954
No. features loaded into GEE for analysis	181,848	187,040	192,589
No. features with spectral regeneration metrics in the final dataset	57,844	70,517	72,762
No. features with metrics in just 2019 and 2020	56,418		
No. features with metrics in just 2020 and 2021		68,233	
No. features with metrics in 2019 and 2021	55,762		55,762
No. features with metrics in all three dataset years	55,350		

As is evident in Table 4, there are a number of harvest area features that possess these metrics in only some of the existing year's datasets Harvest Area Spectral Regeneration datasets. There are a number of likely reasons for this, which can include the following:

- Fluctuations found toward the end (most recent years) of the temporally-segmented Landsat NBR time series, which can reflect remaining atmospheric effects (e.g., particularly cloudy seasons) or an actual change in land surface conditions (e.g., further disturbance), can lead an area to be flagged as having multiple disturbances, and therefore disqualified from further analysis.

- Updates to the ABMI's HFI will result in the appearance of new harvest area features and the aging of existing features, spectral metrics each year. For instance, there are differences in the numbers of harvest area features between different HFI dataset, in the range of ~6,000 to 12,000In addition, a harvest event occurring in 2016 would have been too recent for spectral metrics to have been derived in the 2019 dataset, but would then be eligible for metrics to be extracted in 2021
- New wildfire perimeters and/or other human footprint features can remove harvest areas from further metric calculations related to spectral regeneration. For instance, a sudden drop in spectral signals in 2021 due to wildfire would be flagged as a second disturbance for an area where only one previous disturbance was detected in earlier spectral regeneration metrics. This would lead to its removal from further analyses in the more recent version of the dataset despite showing spectral regeneration metrics in the earlier version.

A comparison of spectral regeneration metric descriptive statistics across the three currently existing datasets (from 2019, 2020, and 2021) using those harvest areas for which metrics exist in all three years (n = 55,350), is provided in Table 5. There is an increase in average metric values from 2019 to 2021. We observe a steady increase on mean and median levels of percent spectral regeneration (reg2019, reg2020, reg2021) over the three years. This implies that overall the harvest areas in the province have been continuing to regenerate as expected, while consistency in metrics such as the preNBR across the three years (Table 5) highlight a level of reliability from one dataset to the next.

Table 5. Comparison of descriptive statistics of selected spectral metrics or attributes (mean values) from 2019 to 2021. Fields are only selected for comparison if *regnMetsYN = Y* for the 3 years. Kruskal-Wallis test results on median values for each triplet of metrics are summarized in the last column.

Variable	Data Year	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Reg2019_m	2019	0.9326	95.0	109.3	105.5	119.6	219.4
Reg2020_m	2020	13.85	100.6	112.4	110	122.1	226.3
Reg2021_m	2021	5.193	102.0	113.5	111.0	122.9	226.4
InDistb (mean)	2019	1	1	1.081	1.275	1.397	7.9
	2020	1	1.001	1.089	1.283	1.411	8.1
	2021	1	1.001	1.091	1.28	1.407	8.7
RegStYr (mean)	2019	1989	1996	2002	2002	2007	2014
	2020	1989	1996	2002	2002	2007	2015
	2021	1989	1996	2002	2002	2007	2015
nbrDstb (mean)	2019	205.8	425.9	505.2	504.8	582.5	926.1
	2020	202.7	426.5	506.1	505.5	583.3	927.8
	2021	203.7	426.8	506.1	505.5	583.2	925.4
	2019	274.9	616.5	651.8	645.1	680	767.5

Variable	Data Year	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
preNBR (mean)	2020	274.9	616.4	651.6	645.1	679.9	767.4
	2021	273.7	616.4	651.7	645.1	680	767.4
Y2R80 (mean)	2019	-9999	6	8.465	-78.844	11.177	28.98
	2020	-9999	6.297	8.578	-23.007	11.274	28.90
	2021	-9999	6.474	8.707	-14.554	11.394	30
reg5yr (mean)	2019	4.14	46.26	60.14	59.25	72.22	176.02
	2020	4.879	46.638	60.575	59.538	72.425	172.76
	2021	8.674	46.742	60.637	59.517	72.354	171.80
hrvYr (mean)	2019	1989	1996	2002	2002	2007	2014
	2020	1989	1996	2002	2002	2007	2014
	2021	1989	1996	2002	2002	2007	2015
totPolyPix	2019	9.008	38.315	85.87	150.959	180.92	6662.41
	2020	9.008	38.315	85.86	150.927	180.856	6661.62
	2021	9.008	38.311	85.86	150.921	180.856	6661.62
perRelvPix	2019	50.02	80.18	91.54	87.33	97.57	100
	2020	50.01	80.9	92.07	87.76	97.8	100
	2021	50	80.74	91.72	87.58	97.51	100

Variations across spectral metrics from 2019, 202, and 2021 are more evident in metric minimums and maximums, and emphasize that there can be notable variability both within and between these datasets. Certain metrics are expected see annual changes as the nature of the algorithms. This applies to the metrics such as absolute percent of spectral regeneration compared to the pre-harvest level will have changed slightly across years. There are a few other possible reasons for subtle and less subtle differences in metrics across the three years. For instance, from the time series stacks on GEE, we observed spatial differences in pixel flag distributions across the different datasets. As we progress from 2019 to the 2021 time series, newly annual median composites are added to the stack of time series, and can lead to one or more of the following:

- Each annual median composite (2019, 2020, and 2021) is subject to the changes in the cloud, cloud shadow, and snow masks as new data is added to the time series, which result in changes of metrics generation for certain pixels in GEE, and/or change in confidence scores used for the post-calculation filtering process in RStudio.
- LandTrendr uses only one value per year per pixel. Radiometric variabilities caused by annual phenological cycles may exist when an additional annual composite is added to the time series and cause slight shifts of spectral metrics.
- The output and vertices extracted from LandTrendr are influenced by the algorithm's parameter settings, such as number of segments allowed. When an additional annual composite is added into the time series, the segmentation process can yield different vertices that could cause inconsistencies in the resultant metric and flag calculations.

Finally, the Landsat time series data used to produce this spectral regeneration layer are all from

"Collection 1" of Landsat, as what has been adapted from LandTrendr guide [5]. However, beginning in July 2024, the GEE data catalog will fully migrate all of its Landsat Collection 1 to the radiometrically improved Collection 2 [6]. This means that future spectral regeneration-related data products will also use Landsat Collection 2. It's possible that using Collection 2 will improve our future generation of spectral metrics, but further inconsistencies comparing with the current year metrics are also to be expected.

5.2. Combined Use of Current and Previous Datasets

Caution must be exercised when combining the use of multiple versions (e.g., 2020, 2021) of the Harvest Area Spectral Regeneration dataset together. As described in section 5.1 above, there are variations in spectral regeneration metrics between datasets that can arise from a multitude of sources (e.g., atmospheric or phenological differences that impact NBR time series, later natural or anthropogenic disturbances, shifts in the HFI harvest area source geometry, etc.), and which do not reflect or represent actual vegetative community or structural changes on the land surface related to regeneration. In general, *we strongly recommend that users rely on the most recent version of the data available*, and only use previous versions if they are particularly interested in spectral regeneration metrics for a specific, relevant year (e.g., if a user's other datasets date to 2019 and it is important to remain temporally consistent, then use of the 2019 Harvest Area Spectral Regeneration dataset rather than the most recent version might be appropriate).

We do not recommend using multiple versions of this dataset from different years to assess or evaluate year-to-year changes in regeneration for Alberta harvest areas (e.g., comparing total spectral regeneration from 2019, to 2020 and 2021). Such comparisons will not produce meaningful or informative results that accurately or consistently depict on-the-ground changes related to vegetation regeneration. Rather, these data are meant for informing on spatial patterns and variations in satellite-based spectral regeneration across forest harvest areas (e.g., see Hird et al., 2021).

6. References

1. Alberta Biodiversity Monitoring Institute and Alberta Human Footprint Monitoring Program, 2024. ABMI Human Footprint Inventory (HFI): Wall-to-Wall Human Footprint Inventory Alberta, 2021. Edmonton, AB, Canada.

2. Hird, J.N., Kariyeva, J., McDermid, G.J., 2021, Satellite Time Series and Google Earth Engine Democratize the Process of Forest-Recovery Monitoring over Large Areas. Remote Sensing, 13, p. 4745. https://doi.org/10.3390/rs13234745

3. Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

4. Key, C.H., Benson, N.C., 2006. Landscape assessment: Sampling and analysis methods. USDA For. Serv. Gen. Tech. Rep. RMRS-GTR-164-CD 1–55. https://doi.org/10.1002/app.1994.070541203

5. Kennedy, R.E., Yang, Z., Gorelick, N., Braaten, J., Cavalcante, L., Cohen, W.B., Healey, S., 2018. Implementation of the LandTrendr algorithm on Google Earth Engine. Remote Sens. 10, 1–10. https://doi.org/10.3390/rs10050691

6. Google Earth Engine: Guides, 2023. Landsat Collection 1 to Collection 2 migration. https://developers.google.com/earth-engine/landsat_c1_to_c2