Department of the Interior
U.S. Geological Survey

Annual
National Land Cover Database (NLCD)
Collection 1
Science Product User Guide

Version 1.0

October 2024

Annual
National Land Cover Database (NLCD)
Collection 1
Science Product User Guide

October 2024

Document Owner:

JESSLYN BROWN

Digitally signed by JESSLYN BROWN
Date: 2024.10.15 13:21:24 -05'00'

Jesslyn Brown Date
Land Cover Next (LCNext) Project Manager
U.S. Geological Survey

Approved By:

KAREN ZANTER

Digitally signed by KAREN ZANTER
Date: 2024.10.15 14:54:09 -05'00'

Karen Zanter Date
Land Satellites Data System (LSDS) CCB Chair
U.S. Geological Survey

EROS
Sioux Falls, South Dakota

Executive Summary

This document describes the relevant characteristics of the Annual NLCD Collection 1 Science Products to facilitate their use in the land cover remote sensing community.

The U.S. Geological Survey's (USGS) Land Cover program has leveraged methodologies from legacy land cover projects – National Land Cover Database (NLCD) and Land Change Monitoring, Assessment, and Projection (LCMAP) – together with modern innovations in geospatial deep learning technologies to create the next generation of land cover and land change information. The resulting Annual NLCD product suite includes six annual products that represent U.S. land cover and surface change characteristics:

1. Land Cover
2. Land Cover Change
3. Land Cover Confidence
4. Fractional Impervious Surface
5. Impervious Descriptor
6. Spectral Change Day of Year

These land cover science algorithms harness the remotely sensed Landsat data record to provide land surface characteristics to scientists, resource managers, and decision-makers. Annual NLCD uses a modernized, integrated approach to map, monitor, synthesize, and understand the complexities of land use, cover, and condition change. With this first release – Annual NLCD Collection 1.0 – the product suite is available for the Conterminous U.S. for 1985 – 2023.

Basic foundational elements of Annual NLCD Collection 1 include:

- Landsat Collection 2 U.S. Analysis Ready Data (ARD)
- Land surface change and land cover data
- Independent reference data for validation and area estimation
- Scenario-driven projections of future land use and land cover extents and patterns
- Assessments focused on land change processes, characteristics, and consequences

This document is under the Land Satellites Data System (LSDS) Configuration Control Board (CCB) control. Please submit changes to this document, as well as supportive material justifying the proposed changes, via Change Request (CR) to the Configuration Management Tool.

Document History

Document Number	Document Version	Publication Date	Change Number
LSDS-2103	Version 1.0	October 2024	CR 21793

Contents

List of Figures

List of Tables

Section 1 Introduction

1.1 Background

The need for improved understanding and management of land surface change requires increased understanding of the basic drivers of change, identification of potential consequences of change on human and natural systems, and greater insight into the impacts and feedback of climate change and other drivers. The geospatial community requires a new generation of monitoring data and information for a wide range of applications. More than ever before, land cover and land change products need to span larger geographic extents, over longer time periods, at higher spatial resolutions, and provide more systematic and consistent information on change. To help meet these growing demands, the United States Geological Survey (USGS) developed the Annual NLCD Collection 1 Science Products.

Annual NLCD Collection 1 is a modern, integrated approach to mapping, monitoring, synthesizing, and understanding the complexities of land use, land cover, and conditional change, which leverages the Landsat satellite data record to provide for the needs of scientists, resource managers, and decision-makers.

1.2 Purpose and Scope

This user guide contains an overview of the current Annual NLCD Collection 1 approach, descriptions of the products and their characteristics, and other relevant information to facilitate use of Annual NLCD Collection 1 Science Products in the land change and land cover science community.

This document includes an overview of reference material regarding the current Annual NLCD Collection 1 Science Products and product information relevant to data users.

1.3 Document Organization

This document contains the following sections:

- Section 1 introduces Annual NLCD Collection 1 and provides an overview of the methods applied in producing Annual NLCD Collection 1 Science Products
- Section 2 provides product characteristics and descriptions
- Section 3 describes the availability of Annual NLCD Collection 1 Science Products via various distribution methods
- Section 4 describes methods used for the Algorithm Description
- Section 5 describes the known caveats and limitations of Annual NLCD Collection 1 Science Products
- Section 6 describes the independent reference data set and product validation along with error estimates
- Section 7 provides contact information for USGS Earth Resources Observation and Science (EROS) User Services
- The References Section contains bibliographic citations
- Appendix A provides the acronyms used in this document and their definitions

Section 2 Product Characteristics

2.1 Product Descriptions

Annual NLCD provides a product suite of six geospatial raster products for each year within the product release time frame. The time frame is dependent on the mapping region and the release version, with subsequent releases updating and extending the product series to additional years. The product suite is described in the following sections.

2.1.1 Land Cover

The Annual NLCD land cover product provides a categorical sixteen-class land cover classification system based on a modified Anderson Level II (Anderson et al., 1976), as seen in Table 2-1. The land cover product represents the predominant surface state within the mapping year with respect to broad categories of artificial or natural surface cover. The categories trace their history to the original Anderson land use and land cover classification system, which was designed as a compromise among the need for compatibility with existing classification systems across U.S. federal agencies, separability using primarily remote sensing data, and logical, hierarchical relationships among classes. Land cover Red, Green, Blue (RGB) color values are shown in Table 2-2.

Class/Value	Description
Water	
11	**Open Water** - areas of open water, generally with less than 25% cover of vegetation or soil.
12	**Perennial Ice/Snow** - areas characterized by a perennial cover of ice and/or snow, generally greater than 25% of total cover.
Developed	
21	**Developed, Open Space** - areas with a mixture of some constructed materials, but mostly vegetation in the form of lawn grasses. Impervious surfaces account for less than 20% of total cover. These areas most commonly include large-lot single-family housing units, parks, golf courses, and vegetation planted in developed settings for recreation, erosion control, or aesthetic purposes.
22	**Developed, Low Intensity** - areas with a mixture of constructed materials and vegetation. Impervious surfaces account for 20% to 49% percent of total cover. These areas most commonly include single-family housing units.
23	**Developed, Medium Intensity** - areas with a mixture of constructed materials and vegetation. Impervious surfaces account for 50% to 79% of the total cover. These areas most commonly include single-family housing units.
24	**Developed, High Intensity** - highly developed areas where people reside or work in high numbers. Examples include apartment complexes, row houses and commercial/industrial. Impervious surfaces account for 80% to 100% of the total cover.

Class/Value	Description
Barren	
31	**Barren Land (Rock/Sand/Clay)** - areas of bedrock, desert pavement, scarps, talus, slides, volcanic material, glacial debris, sand dunes, strip mines, gravel pits and other accumulations of earthen material. Generally, vegetation accounts for less than 15% of total cover.
Forest	
41	**Deciduous Forest** - areas dominated by trees generally greater than 5 meters tall, and greater than 20% of total vegetation cover. More than 75% of the tree species shed foliage simultaneously in response to seasonal change.
42	**Evergreen Forest** - areas dominated by trees generally greater than 5 meters tall, and greater than 20% of total vegetation cover. More than 75% of the tree species maintain their leaves all year. Canopy is never without green foliage.
43	**Mixed Forest** - areas dominated by trees generally greater than 5 meters tall, and greater than 20% of total vegetation cover. Neither deciduous nor evergreen species are greater than 75% of total tree cover.
Shrubland	
52	**Shrub/Scrub** - areas dominated by shrubs; less than 5 meters tall with shrub canopy typically greater than 20% of total vegetation. This class includes true shrubs, young trees in an early successional stage or trees stunted from environmental conditions.
Herbaceous	
71	**Grassland/Herbaceous** - areas dominated by graminoid or herbaceous vegetation, generally greater than 80% of total vegetation. These areas are not subject to intensive management such as tilling but can be utilized for grazing.
Planted/Cultivated	
81	**Pasture/Hay** - areas of grasses, legumes, or grass-legume mixtures planted for livestock grazing or the production of seed or hay crops, typically on a perennial cycle. Pasture/hay vegetation accounts for greater than 20% of total vegetation.
82	**Cultivated Crops** - areas used to produce annual crops, such as corn, soybeans, vegetables, tobacco, and cotton, and perennial woody crops such as orchards and vineyards. Crop vegetation accounts for greater than 20% of total vegetation. This class also includes all land being actively tilled.
Wetlands	
90	**Woody Wetlands** - areas where forest or shrubland vegetation accounts for greater than 20% of vegetative cover and the soil or substrate is periodically saturated with or covered with water.

Class/Value	Description
95	**Emergent Herbaceous Wetlands** - areas where perennial herbaceous vegetation accounts for greater than 80% of vegetative cover and the soil or substrate is periodically saturated with or covered with water.

Table 2-1. Land Cover Legend for CONUS

Pixel Value	Land Cover Class	Color Table RGB Value
250	NoData	0, 0, 0
11	Open Water	70, 107, 159
12	Perennial Ice/Snow	209, 222, 248
21	Developed, Open Space	222, 197, 197
22	Developed, Low Intensity	217, 146, 130
23	Developed, Medium Intensity	235, 0, 0
24	Developed, High Intensity	171, 0, 0
31	Barren Land (Rock/Sand/Clay)	179, 172, 159
41	Deciduous Forest	104, 171, 95
42	Evergreen Forest	28, 95, 44
43	Mixed Forest	181, 197, 143
52	Shrub/Scrub	204, 184, 121
71	Grassland/Herbaceous	223, 223, 194
81	Pasture/Hay	220, 217, 57
82	Cultivated Crops	171, 108, 40
90	Woody Wetlands	184, 217, 235
95	Emergent Herbaceous Wetlands	108, 159, 184

Table 2-2. RGB Color Values for Land Cover

2.1.2 Land Cover Change

There are many ways to represent change across a map product series. The NLCD land cover change product represents annual land cover change between one product year and the next, with those changes represented in the latter year (e.g., differences in land cover between 1985 and 1986 are shown in the 1986 land cover change product). These differences are represented categorically in the product data by concatenating the before and after land cover class codes. For example, a pixel changing from Emergent Herbaceous Wetlands (Class 95) to Woody Wetlands (Class 90), would be represented with a pixel value of 9590. Areas of no change maintain the original land cover classification of the associated year. This product can be independently derived by the user from successive land cover maps but is provided in the product suite as a convenience. Land cover change RGB values are shown in Table 2-3.

Pixel Value	Land Cover Change Class	Color Table RGB Value
9999	NoData	0, 0, 0
11	Open Water	70, 107, 159
12	Perennial Ice/Snow	209, 222, 248

Pixel Value	Land Cover Change Class	Color Table RGB Value
21	Developed, Open Space	222, 197, 197
22	Developed, Low Intensity	217, 146, 130
23	Developed, Medium Intensity	235, 0, 0
24	Developed, High Intensity	171, 0, 0
31	Barren Land (Rock/Sand/Clay)	179, 172, 159
41	Deciduous Forest	104, 171, 95
42	Evergreen Forest	28, 95, 44
43	Mixed Forest	181, 197, 143
52	Shrub/Scrub	204, 184, 121
71	Grassland/Herbaceous	223, 223, 194
81	Pasture/Hay	220, 217, 57
82	Cultivated Crops	171, 108, 40
90	Woody Wetlands	184, 217, 235
95	Emergent Herbaceous Wetlands	108, 159, 184
AABB	Change is shown by a concatenation of previous and current class values	162, 1, 255

Table 2-3. RGB Color Values for Land Cover Change
(AA represents "from" land cover class value; BB represents "to" value)

2.1.3 Land Cover Confidence

NLCD land cover product generation strongly relies on supervised classification that is implemented with a series of deep learning models. The final result from the system is the output of an activation function that transforms values from the neural network into a discrete probability distribution across the output classes. The land cover confidence product provides the probability value for the final output land cover class. Because the process of map product creation incorporates a number of post-classification steps, this probability might not correspond to the maximum value across all classes. It is also important to note that the confidence value does not correspond to the absolute likelihood of the land cover being correct (i.e., these are uncalibrated probabilities) but is, of course, expected to be strongly correlated. The land cover confidence information is provided in Table 2-4.

Pixel Value	Land Cover Confidence	Color Table RGB Value
250	NoData	0, 0, 0
1	Start Value	255, 255, 255
100	Final Value	0, 0, 0

Table 2-4. RGB Color Values for Land Cover Confidence
(colors are represented by grey-scale gradient)

2.1.4 Fractional Impervious Surface

The fractional impervious surface product provides the percentage of a 30-meter pixel that is covered with artificial substrate or structures (pavement, concrete, rooftops, and other constructed materials) that are assumed to be impermeable to water. The impervious surface product provides this percent in a zero to 100 continuous value. These values provide the basis for every land cover pixel mapped as one of the four developed classes and informs the categorical developed land cover class by the thresholds provided in Table 2-5. The value 250 represents unmapped or a background value area as zero represents no mapped impervious surface on the landscape.

Pixel Value	Fractional Impervious Surface	Color Table RGB Value
250	NoData	0, 0, 0
1	Start Value	209, 209, 209
100	Final Value	158, 31, 235

Table 2-5. RGB Color Values for Fractional Impervious Surface (colors are represented by red-scale gradient)

2.1.5 Impervious Descriptor

The impervious descriptor product provides additional categorical information for developed areas. The product distinguishes between non-road ("urban") and road surfaces. It provides a map of road networks that is discernible throughout dense urban interiors and distinguishable from scattered structures and paved lots in outlying areas. Unlike versions of NLCD prior to Annual NLCD Collection 1, this is not a reporting layer for urban source information but is the direct result of a supervised classification algorithm. The impervious descriptor classes are provided in Table 2-6.

Pixel Value	Impervious Descriptor	Color Table RGB Value
250	NoData	0, 0, 0
0	Non-Urban	0, 0, 0
1	Roads	33, 113, 181
2	Urban	246, 236, 39

Table 2-6. RGB Color Values for Impervious Descriptor

2.1.6 Spectral Change Day of Year

The spectral change product provides information on the occurrence of substantial changes in spectral behavior through time. Spectral behavior refers to the directly measurable physical properties (i.e., surface reflectance in one or more wavelength bands) that are derived from Landsat remote sensing data. This product provides the day-of-year (DOY) on which any substantial deviation in surface reflectance was detected within the calendar year of the map product. Spectral changes represent abrupt non-phenological changes in the land surface that may or may not be related to land cover change. For example, a high intensity wildfire in a forested area produces both a substantial change in surface reflectance and a thematic class change within the NLCD land cover legend. Other changes, such as those produced by drought or precipitation, might produce a significant change in Landsat spectral reflectance that still occur within the same land cover class. Product sensitivity to the nature and frequency of spectral change is directly related to the behavior of the underlying change detection algorithm. See Section 4 for more information. A color palette for Spectral Change Day of Year has not been predefined. Values range from 1-366 with zero showing No Change and NoData represented by 9999.

2.2 Product Specifications

2.2.1 Image Data File Format and Structure

NLCD products are provided as Cloud-Optimized GeoTIFF (COG) files. Georeferenced Tagged Image File Format (GeoTIFF) is a broadly supported format for embedding georeferencing information within a TIFF image file. A COG file is a type of GeoTIFF that meets additional requirements for internal structure and efficient support of Hypertext Transfer Protocol (HTTP) range requests, which allows a client to request a subset of data rather than the full file. Formalized GeoTIFF and COG standards are maintained by the Open Geospatial Consortium (OGC, 2023).

Efficient data retrieval from a COG file is supported by several optimizations. First, the data are internally stored as rectangular tiles, which reduces the amount of data that must be read when retrieving a subset of the image data. This tiling structure and the supporting metadata that allows a retrieval of a geospatial region from a subset of these tiles constitute the essential enhancement that a COG provides over a (non-COG) GeoTIFF.

Additional optimizations within a COG file constitute recommended or optional features under the version 1.0 OGC COG standard. COG files, including NLCD product data, are often compressed with one of the common lossless data compression algorithms, which reduce the size of the data that must be stored and retrieved. Internal compression is not unique to COGs, nor required by the standard, but is often employed when dealing with large data sizes. Large COG and other GeoTIFF files commonly contain internal overviews (or pyramids) that provide image data at one or more stages of reduced resolution to enable efficient data visualization.

2.2.2 GeoTIFF Metadata Tags

GeoTIFF defines a set of Tagged Image File Format (TIFF) tags, which describe cartographic and geodetic information associated with geographic TIFF imagery. COG tags are inherited from the GeoTIFF format. GeoTIFF tags convey information about the image. The tags describe the image using information the GeoTIFF reader needs to control the appearance of the image on the user’s screen. The TIFF tags are embedded in the same file as the TIFF image. The GeoTIFF tags provide information on the image projection and corner points, which define the geographic location and extent of the image.

The spatial description of an image in GeoTIFF requires keys stored within the image files and accessible by GeoTIFF readers. Table 2-7 defines the keys necessary to support the Albers Equal Area (AEA) map projection used for Landsat U.S. ARD.

Valid Keys	Possible Values	Meaning
GTModelTypeGeoKey	ModelTypeProjected	Projection Coordinate System
GTRasterTypeGeoKey	RasterPixelsPoint	The coordinate is at the upper left corner of the pixel. This matches the Level-2 source U.S. ARD tiles.
GTCitationGeoKey	AEA WGS84	American Standard Code for Information Interchange (ASCII) reference to public documentation; Albers, Stereographic South Pole, and Universal Transverse Mercator (UTM) are accounted for.
GeographicTypeGeoKey	GCS_WGS_84	Geographic coordinate system used to map lat-long to a specific ellipsoid over the Earth
GeogCitationGeoKey	WGS 84	General citation and reference for all Geographic CS parameters; World Geodetic System (WGS) 84
GeogAngularUnitsGeoKey	Angular_Degree	Geocentric CS Linear units
GeogSemiMajorAxisGeoKey	6378137.0	Ellipsoid Semi-Major Axis
GeogInvFlatteningGeoKey	298.257223563	Inverse of Ellipsoid's flattening parameter
ProjectedCSTypeGeoKey	20000–32760	User-Defined projected coordinate system; European Petroleum Survey Group (EPSG) Projection Codes
ProjectionGeoKey	10000-19999	User-Defined; EPSG / Petrotechnical Open Software Corporation (POSC) Projection Codes (see the EPSG Geodetic Parameter Registry for values)
ProjCoordTransGeoKey	CT_AlbersEqualArea	Coordinate transformation method
ProjLinearUnitsGeoKey	Linear_Meter	Linear units used by this projection
ProjStdParallel1GeoKey	45.5	Latitude of primary Standard Parallel; Value in units of GeogAngularUnits

Valid Keys	Possible Values	Meaning
ProjStdParallel2GeoKey	29.5	Latitude of second Standard Parallel; Value in units of GeogAngularUnits
ProjNatOriginLongGeoKey	-96.0	Longitude of map-projection Natural origin; Value in units of GeogAngluarUnits
ProjNatOriginLatGeoKey	23.0	Latitude of map-projection Natural origin; Value in units of GeogAngularUnits
ProjFalseEastingGeoKey	0.0000000	Easting coordinate of the map projection Natural origin; Value entered in units of ProjLinearUnits
ProjFalseNorthingGeoKey	0.0000000	Northing coordinate of the map projection Natural origin; Value entered in units of ProjLinearUnits

Table 2-7. Albers GeoTIFF Key Description

2.2.3 Raster Values and Data Types

Raster image data are stored within the COG file using one of two common data types, as appropriate to the range of data. All products are single-band raster's with an identical geospatial extent and mapping footprint. Areas within the raster that are outside of the mapping area are assigned the NoData value. Details of data values and data types for each of the NLCD products are shown in Table 2-8.

Product	Data Type	Valid Value Range	NoData Value	Interpretation
Land Cover	UINT8	11–95	250	Table 2-1
Land Cover Change	UINT16	11–9590	9999	Table 2-2
Land Cover Confidence	UINT8	1–100	250	Low (1) to high (100) model confidence
Fractional Impervious Surface	UINT8	0–100	250	Percent area (%)
Impervious Descriptor	UINT8	0–2	250	Table 2-6
Spectral Change Day of Year	UINT16	0–366	9999	Day-of-year (DOY) of change (1–366), or no change (0)

Table 2-8. NLCD Product Data Types and Data Values

2.2.4 External Metadata

In addition to the metadata information embedded in the TIFF files, each NLCD product is also accompanied by an XML (eXtensible Markup Language) file that provides geospatial metadata compliant with the Federal Geographic Data Committee (FGDC) Content Standard for Digital Geospatial Metadata (CSDGM) metadata standard. These metadata provide geospatial, descriptive, data value, and citation information associated with the data product.

2.2.5 Collection and Version Strategy

Annual NLCD will adhere to a data release and update strategy that prioritizes time series continuity. The strategy includes the use of “Collection(s)” and “Version(s).” A “Collection” is a set of Annual NLCD products developed using consistent algorithms, processing procedures, and methods. The initial release of a Collection will include end-to-end, systematic production of Annual NLCD products. For example, the initial release of Collection 1.0 included new annual products for 1985 – 2023. Within each Collection, there will be annual updates to expand the time series with recent years (i.e., 2024, 2025, etc.). These annual updates will be referred to as “Versions” and will utilize the same root algorithms, processing procedures, and methods during production.

Through ongoing USGS research, development, testing, and user feedback, Annual NLCD will mature to the point that end-to-end reprocessing becomes appropriate and cost effective (i.e., 2-4 years). At this time, a new collection will be introduced, followed by its own annual update versions.

2.2.6 File Naming Convention

Annual NLCD mosaic products adhere to the following file naming convention:

Annual_NLCD_{PRODUCT}_{YYYY}_{REGION}_C{C}V{V}.tif

Annual_NLCD	Annual National Land Cover Database
PRODUCT	One of six land cover product (See Table 2-9)
YYYY	Map Year
REGION	Region of the U.S. (“CU” = CONUS, “AK” = Alaska, “HI” Hawaii)
C	Collection number (“1, 2, or 3”)
V	Version number (“0, 1, or 2”)

Product Name	Filename: {PRODUCT}
Land Cover	LndCov
Land Cover Change	LndChg
Land Cover Confidence	LndCnf
Fractional Impervious Surface	FctImp
Impervious Descriptor	ImpDsc
Spectral Change Day of Year	SpcChg

Table 2-9. Annual NLCD Product Name and Respective Filename Convention

Examples of this convention for map year 2001, the Conterminous U.S. for region (CU), and Collection 1.0, would be represented as follows:

Annual_NLCD_LndCov_2001_CU_C1V0.tif
Annual_NLCD_LndChg_2001_CU_C1V0.tif
Annual_NLCD_LndCnf_2001_CU_C1V0.tif
Annual_NLCD_FctImp_2001_CU_C1V0.tif
Annual_NLCD_ImpDsc_2001_CU_C1V0.tif
Annual_NLCD_SpcChg_2001_CU_C1V0.tif

The external metadata files associated with each product follow the exact same structure, with exception for the .xml extension; for example:

Annual_NLCD_LndCov_2001_CU_C1V0.xml

Tiled Annual NLCD products adhere to the following file naming convention:

Annual_NLCD_H{xx}V{yy}_{PRODUCT}_{YYYY}_{REGION}_C{C}V{V}.tif

Annual_NLCD	Annual National Land Cover Database
xx	Horizontal tile number
yy	Vertical tile number
PRODUCT	One of six land cover product (See Table 2-9)
YYYY	Map Year
REGION	Region of the U.S. (“CU” = CONUS, “AK” = Alaska, “HI” Hawaii)
C	Collection number (“1, 2, or 3”)
V	Version number (“0, 1, or 2”)

Examples of this convention for U.S. ARD tile H05V02, map year 2001, the Conterminous U.S. for region (CU), and Collection 1.0, would be represented as follows:

Annual_NLCD_H05V02_LndCov_2001_CU_C1V0.tif
Annual_NLCD_H05V02_LndChg_2001_CU_C1V0.tif
Annual_NLCD_H05V02_LndCnf_2001_CU_C1V0.tif
Annual_NLCD_H05V02_FctImp_2001_CU_C1V0.tif
Annual_NLCD_H05V02_ImpDsc_2001_CU_C1V0.tif
Annual_NLCD_H05V02_SpcChg_2001_CU_C1V0.tif

The external metadata files associated with each product follow the exact same structure:

Annual_NLCD_H05V02_LndCov_2001_CU_C1V0.xml

Section 3 Data Access

Data distribution methods are summarized in the following sections. All products are available to download at no cost through the USGS, other than Amazon Web Services (AWS), which has fees associated with it.

3.1 MRLC Direct Download

CONUS-wide mosaic data can be accessed and downloaded from the Multi-Resolution Land Characteristics (MRLC) Consortium data downloads page here https://www.mrlc.gov/data.

3.2 MRLC Web Viewer

The MRLC web viewer and tutorial information can be found here https://www.mrlc.gov/viewer/. Data sets will be listed in the Contents area on the left side. Data can be downloaded from the Data Download Tool on the right side or in the top middle section on the Viewer page.

3.3 EarthExplorer

The Annual NLCD Collection 1 Science Products can be accessed via EarthExplorer (EE) here https://earthexplorer.usgs.gov/.

3.4 AWS Cloud Access

Cloud access will be available via the usgs-landcover production bucket (AWS Simple Storage Service (S3) object store) on a requester pays basis. Details below:

3.4.1 S3 Cloud Storage Structure

The Annual NLCD data in the cloud is stored in a nested directory structure. The structure to the directory containing the assets for each product is as follows.

s3://usgs-landcover/annual-nlcd/c1/v0/[region[cu-ak-hi]]/tile/h{xx}v{yy}/, for example:

S3 Uniform Resource Identifier (URI): s3://usgs-landcover/annual-nlcd/c1/v0/cu/tile/h14v15/Annual_NLCD_H14V15_FctImp_1985_CU_C1V0.tif

HTTP Uniform Resource Locator (URL): https://usgs-landcover.s3.us-west-2.amazonaws.com/annual-nlcd/c1/v0/cu/tile/h14v15/Annual_NLCD_H14V15_FctImp_1985_CU_C1V0.tif

s3://usgs-landcover/annual-nlcd/c1/v0/[region[cu-ak-hi]]/mosaic/, for example:

S3 Uniform Resource Identifier (URI): s3://usgs-landcover/annual-nlcd/c1/v0/cu/mosaic/Annual_NLCD_FctImp_1985_CU_C1V0.tif

HTTP Uniform Resource Locator (URL): https://usgs-landcover.s3.us-west-2.amazonaws.com/annual-nlcd/c1/v0/cu/mosaic/Annual_NLCD_FctImp_1985_CU_C1V0.tif

3.5 ScienceBase

ScienceBase provides long-term access and preservation of the data and metadata here https://www.sciencebase.gov/catalog/item/655ceb8ad34ee4b6e05cc51a. This also follows USGS Fundamental Science Practices (FSP) and meets USGS standards. Metadata are funneled via ScienceBase to other areas, such as data.gov, the USGS Science Data Catalog (SDC), and GeoPlatform.

3.6 EVA Tool

The Enhanced Visualization and Analysis (EVA) tool provides users with detailed county statistics for any two Annual NLCD land cover dates to support quick and powerful change analyses. Access to the EVA Tool can be found here https://www.mrlc.gov/eva/.

Section 4 Annual NLCD Collection 1.x Algorithmic Descriptions

4.1 Collection 1.0 CONUS Baseline Algorithms

The Annual NLCD Collection 1.0 product suite for CONUS is created by an ensemble of algorithms, incorporating lessons learned and data sets from previous NLCD and LCMAP releases.

At a high level, the process has three main parts (see Figure 4-1):

1. Distillation of Landsat U.S. ARD into per-pixel time series segments and annual leaf-on/leaf-off information
2. Deep learning architecture that integrates spatial and temporal information to predict land cover
3. Post-classification that reduces error and produces final raster outputs

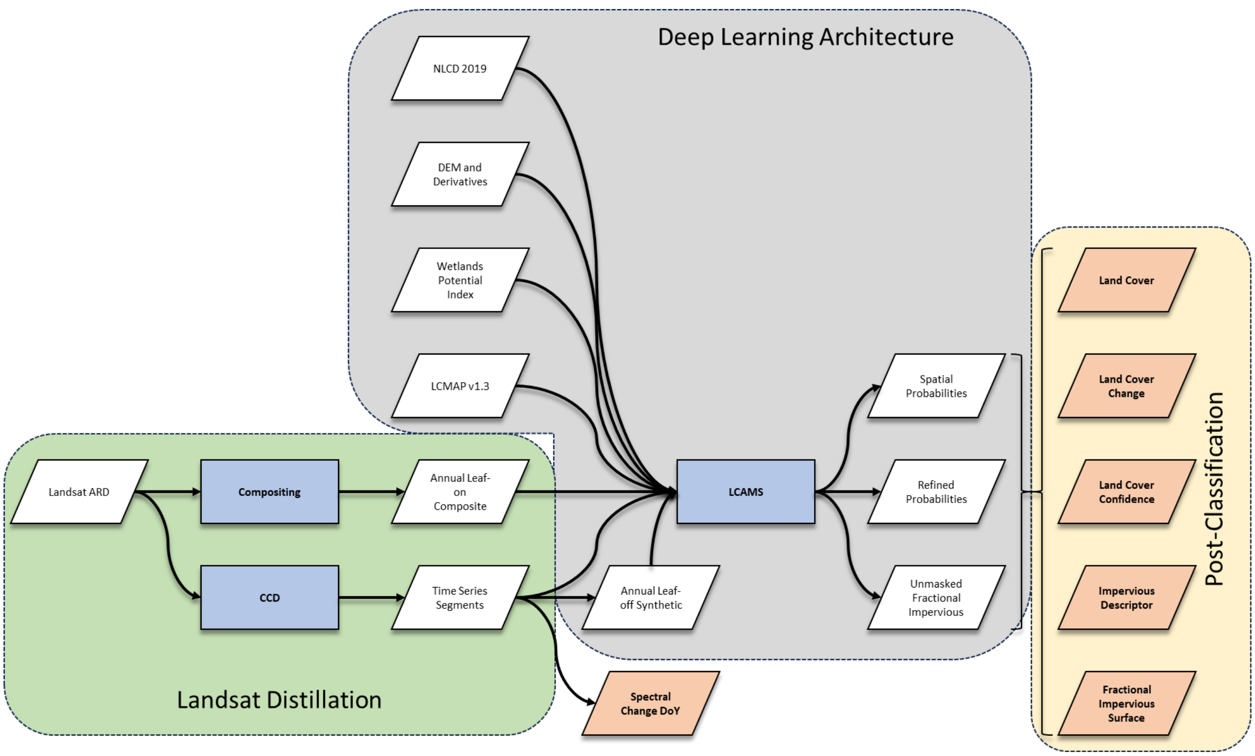

Figure 4-1. Overview of Annual NLCD Collection 1.0

4.1.1 Landsat U.S. ARD Distillation

4.1.1.1 Continuous Change Detection

The Continuous Change Detection (CCD) component is supported by two similar algorithms: PyCCD and JuliaCCD. Details of the PyCCD algorithm are provided in the LCMAP 1.3 documentation (U.S. Geological Survey, 2022a). JuliaCCD (Tollerud and others, 2023) was developed as a refinement to the original CCD methodology (Zhu and Woodcock, 2014), particularly for how change is identified, and to improve computational efficiency by using the Julia programming language. Both algorithms use all available Landsat surface reflectance, brightness temperature, and associated quality data to characterize periods of stability for a given pixel time series as a series of segments and determine when spectral responses deviate from previously established patterns to identify a spectral “break”. This is generally the result of an abrupt change (e.g., wildfire, logging, and mining) but can also result from a gradual change (e.g., forest growth, insect infestation, disease). When a break occurs, a new temporal segment or model is established for the subsequent data points. The fundamental input data for both algorithms are the same, with differences in terms of the Landsat collection used:

A complete history is gathered from Landsat U.S. ARD for each pixel location consisting of the following information:

- Landsat Level 2 Surface Reflectance (SR) for the Thematic Mapper (TM), Enhanced Thematic Mapper Plus (ETM+), and Operational Land Imager (OLI)
- Landsat Level 2 Brightness Temperature (BT) for TM, ETM+, and the Thermal Infrared Sensor (TIRS)
- Landsat Level 2 Pixel Quality Assessment for TM, ETM+, and OLI
- Observation dates

Similar to PyCCD, JuliaCCD is run as a per-pixel algorithm, and the fundamental output is the spectral characterizations (segments) of the input data. The output from JuliaCCD is transformed and stored as a series of Apache Parquet files matching the schema in Table 4-1.

Field Name	Data Type	Valid Range	Description
px	Integer	1-max chunk columns	Column offset from the upper-left of the processed chunk
py	Integer	1-max chunk rows	Row offset from the upper-left of the processed chunk
sday	String	“1982-01-01” thru “2023-12-31”	Start date of the segment in ISO-8601 format
eday	String	“1982-01-01” thru “2023-12-31”	End date of the segment in ISO-8601 format
bday	String	“1982-01-01” thru “2023-12-31”	Break date of the segment in ISO-8601 format

Field Name	Data Type	Valid Range	Description
curqa	integer	1,4,6,8,14,24,44,54	Describes the regression fitting procedure (See Table 4-2)
chprob	Boolean	True/False	Whether the segment's associated break date represents a spectral change
<band>int	float		Linear regression intercept value for the associated band
<band>slop	float		Linear regression slope value for the associated band
<band>cos1	float		Linear regression 1^{st} order cosine value for the associated band
<band>sin1	float		Linear regression 1^{st} order sine value for the associated band
<band>cos2	float		Linear regression 2^{nd} order cosine value for the associated band
<band>sin2	float		Linear regression 2^{nd} order sine value for the associated band
<band >cos3	float		Linear regression 3^{rd} order cosine value for the associated band
<band>sin3	float		Linear regression 3^{rd} order sine value for the associated band
<band>rmse	float		Root mean square error (RMSE) associated with the linear regression
<band>mag	float		Magnitude of spectral deviation from predicted values for the associated band

Table 4-1. Parquet Schema Used to Store JuliaCCD Outputs (where band is a two-letter shorthand for one of the 8 SR or BT inputs)

Curve Quality Value	Description
1	Simple linear regression with only an intercept and slope for segments with less than 12 observations
4	Normal fitting operations for segments with 12-17 observations that pass pixel quality filtering that encompass at least one year
6	Normal fitting operations for segments with 18-23 observations that pass pixel quality filtering that encompass at least one year
8	Normal fitting operations for segment with greater than 23 observations that pass pixel quality filtering that encompass at least one year

Curve Quality Value	Description
14	Referred to as “Start Fits”, it is a four coefficient fit used at the beginning of the time series when the normal fitting operation skipped observations due to spectral instability
24	Referred to as “End Fits”, it is a four coefficient fit used at the end of the time series after a break when there are insufficient data for normal fitting operations
44	Referred to as “Insufficient Clear”, it is a four coefficient fit used for the entire time series due to less than 25% of the observations passing pixel quality filtering
54	Referred to as “Persistent Snow”, it is a four coefficient fit used for the entire time series when greater than 75% of the observations are flagged as ice/snow in pixel quality

Table 4-2. JuliaCCD Segment Curve Quality Information

4.1.1.2 Annual Leaf-On Compositing

Two similar annual leaf-on compositing approaches were utilized to support the downstream deep learning architecture. For training purposes, the composites used in support of the NLCD 2019 Edition (Jin et al., 2023) were based on Landsat Collection 1 U.S. ARD, while an updated approach was developed for Landsat Collection 2 U.S. ARD. Figure 4-2 shows the flowchart of the new compositing approach using Landsat Collection 2 imagery.

The fundamental input data for both algorithms are the same, with differences in terms of the Landsat Collection used:

- Landsat Level 2 Surface Reflectance (SR) for the Thematic Mapper (TM), Enhanced Thematic Mapper Plus (ETM+), and Operational Land Imager (OLI)
- Landsat Level 2 Pixel Quality Assessment for TM, ETM+, and OLI

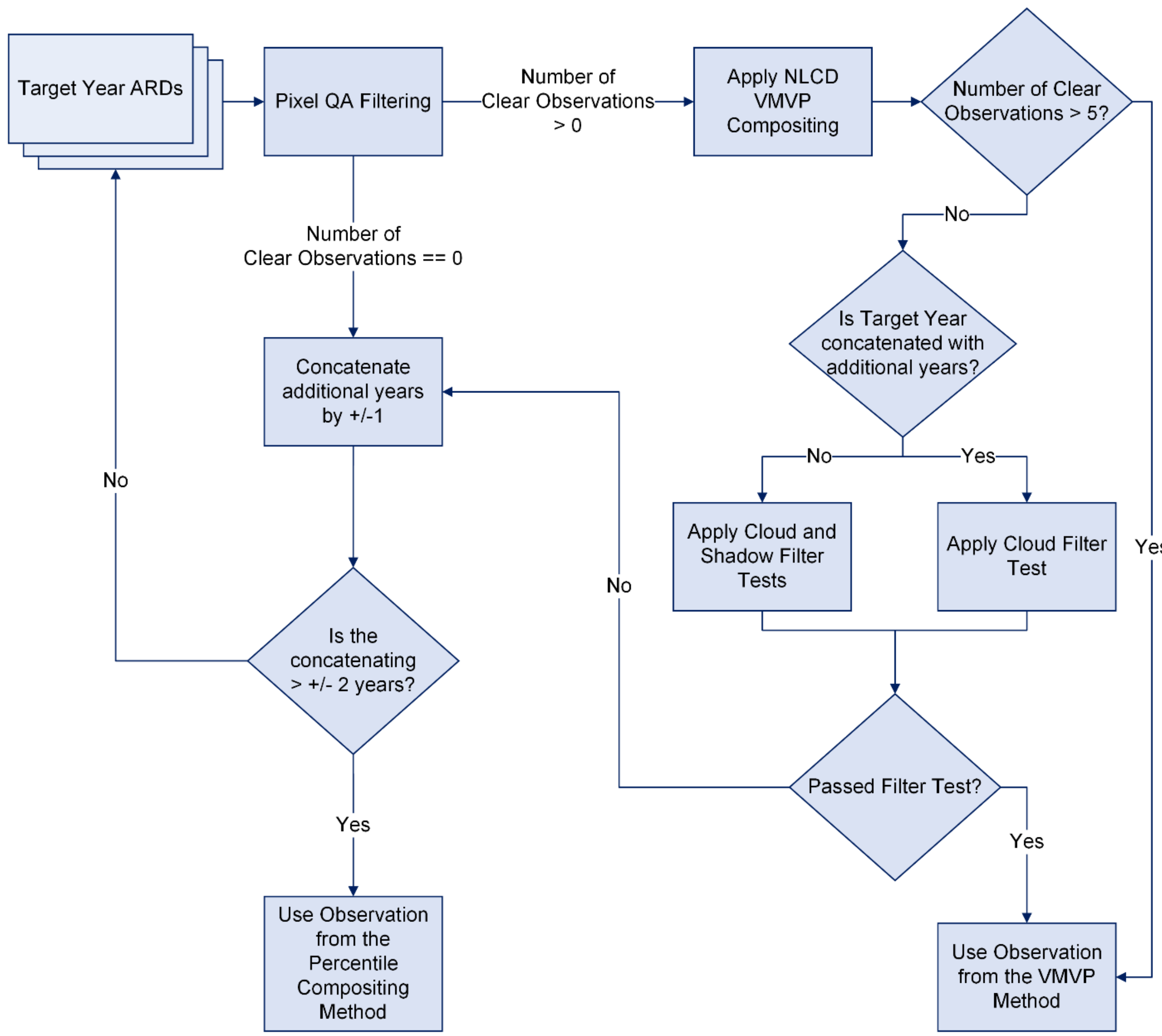

Figure 4-2. Overview of Landsat Collection 2 U.S. ARD Compositing

4.1.1.2.1 Key Concepts

<u>Virtual Median Value Point (VMVP)</u>

Representation of the “overall” median value for a single pixel, through time, across multiple bands. A Euclidean norm is taken from the median value of each band, through time, as described in Jin et al., 2023.

<u>Cloud Filter Test</u>

Uses near-infrared (NIR), short wavelength infrared (SWIR1) and visible bands (blue, green, and red) to determine the presence of clouds, as per Equation 4-1 and Equation 4-2.

$$Cloud = (Blue + Green + Red) > 1.5$$

Equation 4-1.

or

$$Cloud = (Blue + Green + Red) > 1.2 \ and \ \frac{NIR}{SWIR1} > 1$$

Equation 4-2.

<u>Shadow Filter Test</u>
Uses infrared and the visible red bands to determine the presence of shadows, as per Equation 4-3 and Equation 4-4.

$$Shadow = (NIR + SWIR1 + SWIR2) < 1.2 \ and \ \frac{NIR}{Red} < 1.2 \ and \ \frac{SWIR1}{Red} < 1.3$$

Equation 4-3.

or

$$Shadow = (NIR + SWIR1 + SWIR2) < 1.0 \ and \ \frac{NIR}{Red} < 1.6 \ and \ \frac{SWIR1}{Red} < 1.3$$

Equation 4-4.

<u>Percentile-Based Compositing</u>
Percentile-based compositing sums the three visible bands (blue, green, and red) for each observation. Observations with their sum within the lower percentile (20%) and upper percentile (30%) are considered valid.

4.1.1.2.2 Landsat Collection 2 U.S. ARD Workflow

The updated compositing algorithm used for processing Landsat Collection 2 U.S. ARD:

1. Identify observations and the associated bands (blue, green, red, NIR, SWIR1, SWIR2, QA_PIXEL) for the target year, and intra-year range (1 May thru 30 September) for a given tile
    a. Filter out Landsat 7 observations unless the target year is 2012 or before 2003 due to issues related to Scan Line Corrector (SLC)-off artifacts
2. Organize the pixels into spectral “rods”, containing the observations and supporting bands
3. Observations within a rod are filtered by their associated QA_PIXEL value, removing those flagged as fill, clouds, cirrus, clouds, cloud shadow, cloud dilation, and snow
4. If the number of remaining observations is greater than 5:
    a. Calculate the **VMVP** and select the observation that is closest in value
5. If the number of clear observations is greater than 0, and less than or equal to 5:
    a. If additional observations have been concatenated (see step 6):
        i. Calculate the **VMVP** and apply the additional **cloud filter test**
    b. Else, calculate the **VMVP** and apply the additional **cloud and shadow filter tests**

6. If there are pixels that are missing a valid value due to QA_PIXEL filtering or the additional **cloud and shadow filter tests**:
    a. Identify addition observations that are ±1 year from the target year and within the intra-year range and concatenate them with the current observation stack
    b. Repeat steps 4 and 5 for the identified pixels
    c. Repeat this process until the defined ±2 year threshold is reached
7. If there are still pixels missing after the concatenation of addition observations:
    a. Apply **percentile-based compositing** and use those observations for the VMVP selection process

4.1.2 Deep Learning Architecture

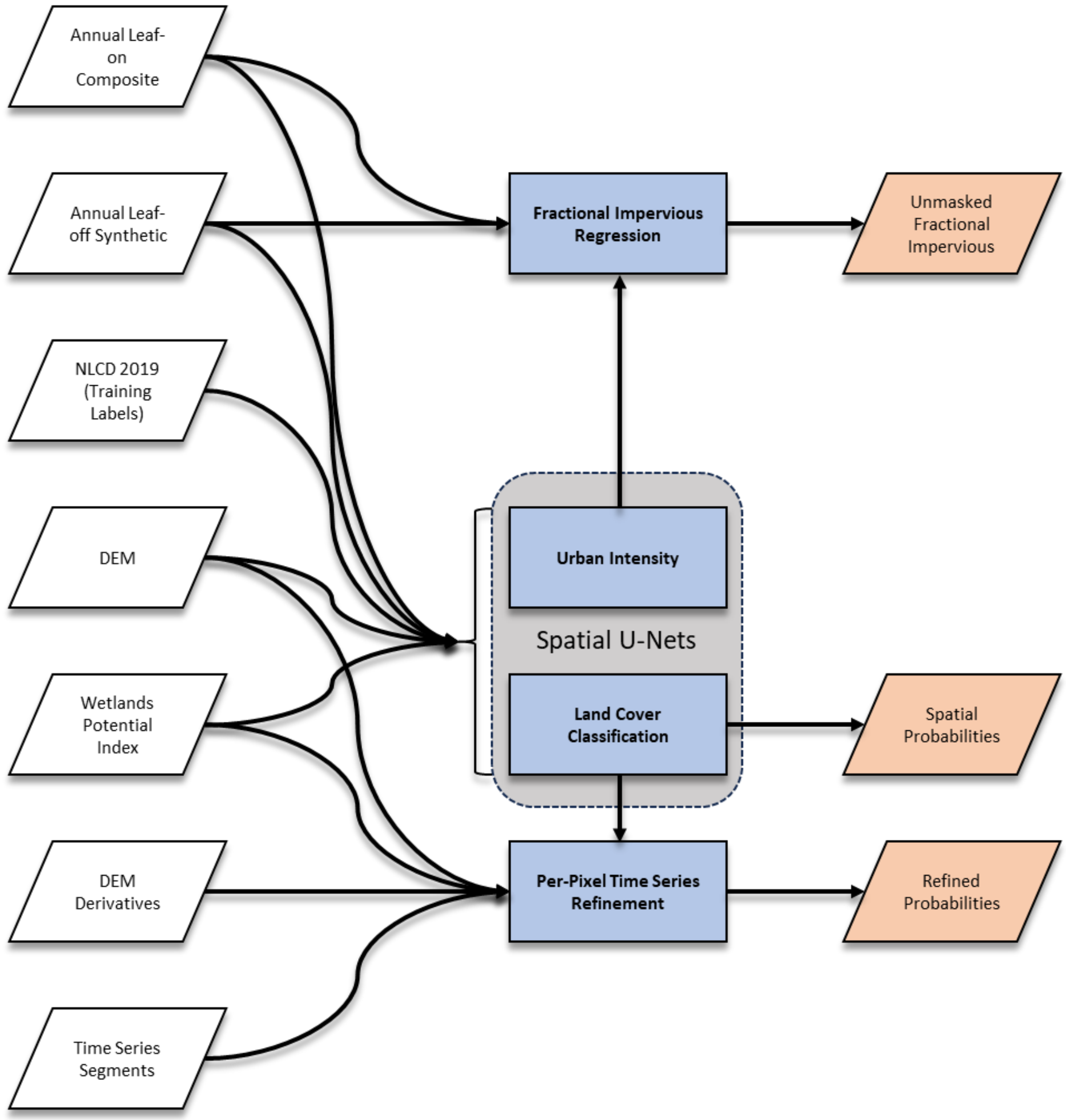

Figure 4-3. Overview of Deep Learning Approach

Annual NLCD Collection 1.0 relies on a multi-stage deep learning architecture referred to as the Land Cover Artificial Mapping System (LCAMS) to generate thematic land cover, fractional impervious cover, and other information that supports the product suite (see Figure 4-3). This system relies on a series of neural networks (see Table 4-3) that are trained using labels derived from a modified version of the NLCD 2019 Edition product suite.

Model	Associated Step(s)	Total Parameters	Loss Function
U-net	Land Cover	9,842,958	Focal Jaccard
Residual Attention U-net	Land Cover/Urban Intensity	26,275,232	Focal Jaccard
Residual Attention U-net	Land Cover/Urban Intensity	26,275,232	Focal Dice
Classification Transformer / Multi-layer Perceptron (MLP)	Time Series Refinement	1,382,528	Focal
Regression Transformer	Fractional Impervious	205,360	MSE

Table 4-3. Neural Networks Utilized by LCAMS

4.1.2.1 Input Data

These data sets were leveraged as-is from their sources to support the various deep learning steps:

- National Land Cover Database (NLCD) 2019 Edition Science Products
- CCD based time series segments
    - Land Change Monitoring, Assessment, and Projection (LCMAP) v1.3 PyCCD segments for training
    - Landsat Collection 2 based JuliaCCD Band First Probability (BFP) segments for prediction
- Digital Elevation Model (DEM) (U.S. Geological Survey (2022b))
- DEM Aspect
- DEM Slope
- DEM Topographic Position Index
- Wetland Potential Index (WPI) agreement layer (see Table 4-4)
- Leaf-on Composites
    - Landsat Collection 1 U.S. Analysis Ready Data (ARD)
    - Landsat Collection 2 U.S. Analysis Ready Data (ARD)
- Leaf-off Synthetics
    - Synthetics are created by predicting surface reflectance values for November 15th of each year off the regression models from the associated time series segments

- Supported by LCMAP v1.3 segments for training and JuliaCCD BFP segments for prediction

Value	Agreement
2	Hydric soil from the gridded Soil Survey Geographic Database (gSSURGO) of the Natural Resources Conservation Service (USDA NRCS, 2023)
3	National Wetlands Inventory (NWI; USFWS, 2023)
4	NLCD 2011 Edition (USGS, 2014)
5	gSSURGO + NWI
6	gSSURGO + NLCD
7	NWI + NLCD
8	gSSURGO + NWI + NLCD

Table 4-4. Agreement Values Used for Data Sets that Comprise WPI

4.1.2.2 Ensemble Land Cover Classification

The principal labels used for land cover classification are the expanded legend in the NLCD 2021 (ed.) Science Products, collapsing the urban/impervious-related labels with those cross-walked from the NLCD Impervious Descriptor layers as shown in Table 4-5. The outputs of predictions generated are annual sequences of prediction scores for sixteen thematic land cover classes. Note that these classes differ from the sixteen-class NLCD land cover legend in that they do not contain the subclasses of developed intensity (only “urban” and “roads” instead of open space developed and low, medium, and high intensity developed). In addition, they contain the NLCD 2021 (ed.) Science Product forest transition classes (herbaceous-forest and shrub-forest).

The approach consists of 3 stages, with predictions from stages 2 and 3 comprising the outputs used in post-classification processing. The 3 stages are as follows:

1. Stage 1 consists of training three U-Net variants
    a. Total spatial extent of CONUS is gridded into 256x256 spatial chips
    b. Spatial chips are split evenly into two sets
        i. Both sets further divided into 80/20 train/validation split
    c. Pair the first set of spatial chips with NLCD years 2001/2011, and pair the second set of chips with NLCD years 2016/2019
        i. Each chip set gets used twice, once with each associated year within the pair
    d. Each of the 3 U-Net variants are trained on the recombined set
2. Stage 2 takes the CONUS trained models and applies regional fine-tuning to produce the Spatial Probabilities data set
    a. Same basic approach as stage 1, except the spatial extents have been defined for 4 separate regions (see Figure 4-4), where each region has its own set of spatial chips and models
    b. Previously trained CONUS models are used as weights

    c. The Spatial Probabilities data set is derived from the ensembled average of the three models predictions, for each region
3. Stage 3 brings in the CCD segment information and uses an ensembled transformer and multi-layer perceptron (MLP) approach to produce the Refined Probabilities data set
    a. Organize CONUS spatial extent into 20 regions (see Figure 4-5), each with their own modeling
    b. Transformer inputs:
        i. Append two additional convolutional layers
        ii. Annual prediction scores from the Spatial Probabilities data set oriented into a time series of 5x5 pixel chips, with the center pixel representing the pixel of interest
    c. First MLP inputs:
        i. CCD non-intercept harmonic coefficients for six spectral bands and one thermal band
        ii. CCD RMSE for six spectral bands and one thermal band
        iii. Static layers: DEM, DEM derivatives, WPI
    d. Second MLP inputs:
        i. Outputs from the transformer and first MLP
    e. The Refined Probabilities data set is derived from the second MLP predictions

NLCD Impervious Descriptor Class	NLCD Impervious Descriptor Value	LCAMS Legend
Primary Road	20	Road
Secondary Road	21	Road
Tertiary Road	22	Road
Thinned Road	23	Road
Non-road non-energy impervious	24	Urban
Microsoft Buildings	25	Urban
LCMAP Impervious	26	Urban
Wind Turbines	27	Urban
Well Pads	28	Urban
Other energy production	29	Urban
Railroads	32	Urban
Solar Installations	33	Urban

Table 4-5. Crosswalk of NLCD Developed Subclasses to LCAMS Legend Using Impervious Descriptor Values

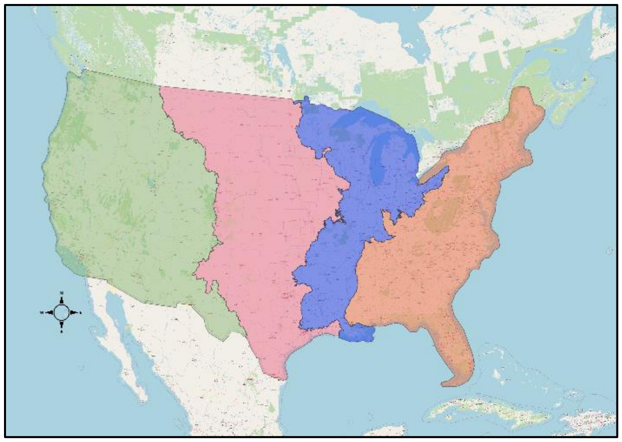

Figure 4-4. Regional Extents of U-Net Land Cover Classification Models

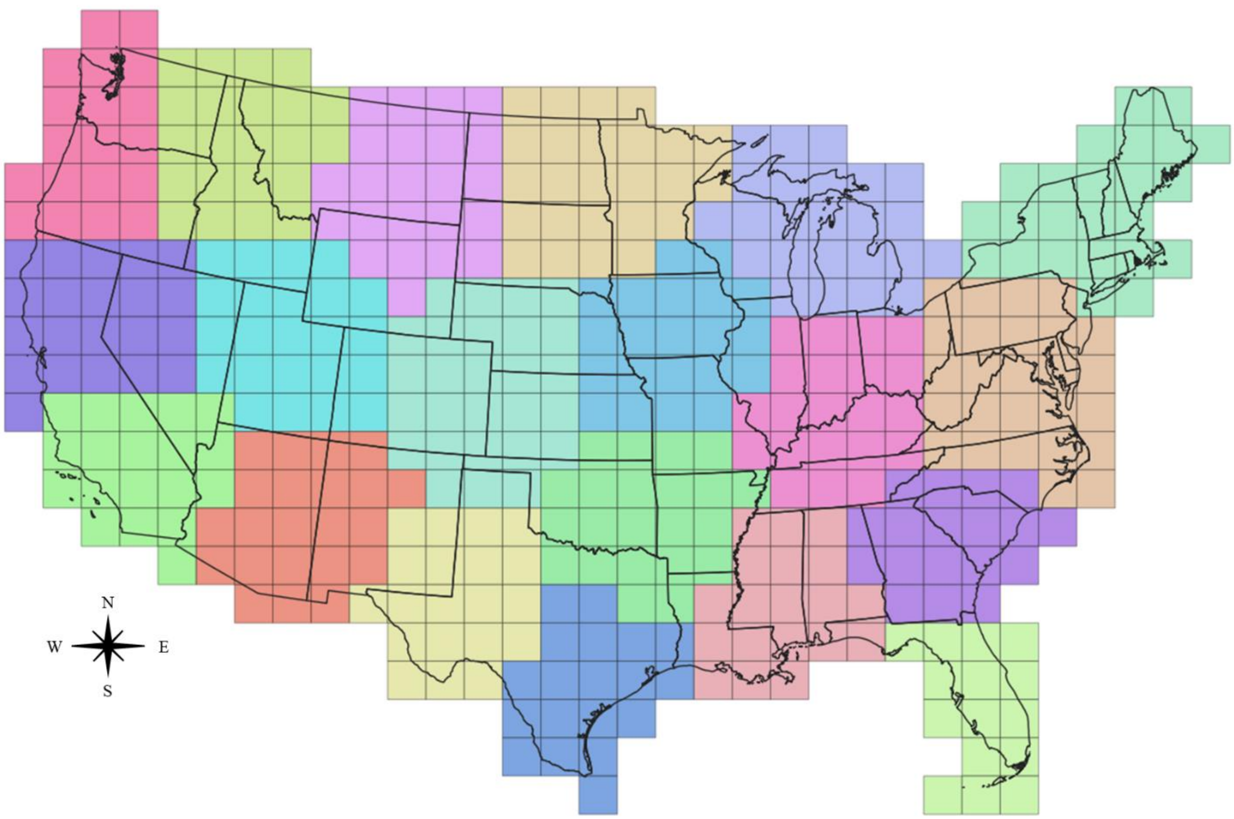

Figure 4-5. Regional Extents of Transformer/MLP Land Cover Models Overlaid with [Landsat U.S. ARD Tile Grid]

4.1.2.3 Impervious Intensity Classification

The LCAMS system includes a parallel branch of processing flow to predict the fractional impervious cover for each urban or road pixel. The resulting Unmasked Fractional Impervious data set contains annual 0 to 100 percent regression values.

1. Stage 1 consists of training 2 U-Net variants (see Table 4-3)
    a. Uses the same population of 256x256 spatial chips as Stage 1 of the Ensemble Land Cover Classification (see Section 4.1.2.2)
    b. Target labels are NLCD 2021 Edition Science Product developed classes, with non-developed classes cross-walked to open space developed
    c. Predictions are an average between the two models
2. Stage 2 uses a regression transformer to calculate fractional impervious
    a. Start with the same 256x256 spatial chips, subsampled to those chips with at least 40% of the pixels having > 0% impervious
    b. Training points are stratified by intensity values
        i. 66 million training, and 16 million test points
    c. The Unmasked Fractional Impervious data set is derived from these predictions

4.1.3 Land Cover Post-Classification

The final land cover determination is made by taking in the spatial probabilities, refined probabilities, time series segments, and unmasked fractional impervious information through a series of steps designed to reduce error associated with any one data set. The LCAMS classes (Table 4-6) are then cross-walked to the final NLCD legend, with the thresholded unmasked fractional impervious serving as inputs for the urban classifications.

The process is broken into two distinct flows depending on the class structure of the prediction time series. Any prediction time series containing only cultivated crops, pasture/hay, grassland/herbaceous, or shrub/scrub is processed utilizing the Spatial and Temporal Integrated Probability-based Post-processing (STIPP) methodology, with other land cover sequences using time series segment-based steps.

The following additional steps are applied:

1. If there are missing class values in the time series, then fill the missing values based on the previous or following class value
    a. If the previous or following class value is water, or the spatial predictions is water, then set the class value to the spatial prediction
    b. Else, use segment break day information to decide whether to use the previous or following class value to fill with, similar to LCMAP Rule-Based Assignment (LCMAP ADD Section 4.2.6.3)
2. If there are ice/snow class values within the time series
    a. If the complete time series shows perennial ice/snow, then keep it
    b. Else, use the mode of the non-ice/snow class values to replace the ice/snow values
3. Crosswalk the LCAMS forest transitional classes to the associated NLCD classes (see Table 4-6)

4. Replace the urban and road LCAMS classes based on thresholded values from the Unmasked Fractional Impervious data set (see Table 4-6)
5. Fractional impervious is derived by using the urban and road LCAMS classes as a mask for the Unmasked Fractional Impervious data set

LCAMS Class	Impervious Fractional Cover	NLCD Class
Open Water	N/A	Open Water
Perennial Ice / Snow	N/A	Perennial Ice / Snow
Roads / Urban	1-19	Developed, Open Space
Roads / Urban	20-49	Developed, Low Intensity
Roads / Urban	50-79	Developed, Medium Intensity
Roads / Urban	80-100	Developed, High Intensity
Barren Land	N/A	Barren Land
Deciduous Forest	N/A	Deciduous Forest
Evergreen Forest	N/A	Evergreen Forest
Mixed Forest	N/A	Mixed Forest
Shrub / Scrub	N/A	Shrub / Scrub
Grassland / Herbaceous	N/A	Grassland / Herbaceous
Pasture / Hay	N/A	Pasture / Hay
Cultivated Crops	N/A	Cultivated Crops
Woody Wetlands	N/A	Woody Wetlands
Emergent Herbaceous Wetlands	N/A	Emergent Herbaceous Wetlands
Forest Transitional Shrub	N/A	Shrub / Scrub
Forest Transitional Grassland	N/A	Grassland / Herbaceous

Table 4-6. Cross-walked Values from LCAMS Classification to Final Land Cover Classes

4.1.3.1 Spatial and Temporal Integrated Probability-Based Post-processing

The STIPP method is applied to time series predictions that contain only cultivated crops, pasture/hay, grassland/herbaceous, or shrub/scrub. This method integrates the Spatial and Refined Probabilities to reduce temporal error and be more spatially cohesive.

1. If the spatial predictions are the same value for the entire time series, then use these as the class values
2. Else, combine predicted land transitions from the spatial and refined predictions
    a. Sum the weighted spatial and refined probabilities
    b. For each identified transition point, look at the summed probabilities pre- and post-transition
        i. If the classes would remain the same, remove the transition

ii. Else, evaluate the difference in pre- and post-transition probabilities of the majority classes before and after which should be different from each other. Two relative difference indices and one absolute difference index are calculated to make the decision. One relative index is the difference between the mean pre- and post-probabilities. The other relative index is the difference of the last pre-transition year and the first post-transition year. Both relative indices are calculated as the value of abs(pre-prob – post-prob) / max(pre-prob, post-prob). The absolute index is calculated as the absolute difference in mean probabilities between both majority classes for post-transition.
    1. If both relative difference indices of both majority classes are <=0.25, then remove the transition
    2. If the relative mean difference index of either majority class is <=0.25, and the absolute index is <=0.1, then remove the transition

4.1.3.2 Segment-Based Post-Classification

4.1.3.2.1 Key Concepts

Vegetative Growth and Decline

For situations in which the forest transition classes were not directly predicted, additional vegetative growth and decline logic is applied based on the LCMAP Secondary Analysis approach (LCMAP ADD Section 4.2.6.2). If the annual predictions favor grassland/herbaceous or shrub/scrub in the first year of the segment and one of the three forest classes (deciduous, evergreen, or mixed) in the final year of the segment, *and* the segment meets a spectral change criterion consistent with growth, it is treated according to the rules of vegetative growth. Conversely, if the annual predictions favor one of the forest classes in the first year of the segment and either grassland/herbaceous or shrub/scrub in the last year of the segment, *and* the segment meets a spectral change criterion consistent with decline, it is treated according to the rules of vegetative decline.

The spectral criterion that governs this determination is defined based on $\hat{p}_\ell$ of Equation 4-5. A normalized band ratio between the near-infrared and SWIR1 bands is defined as follows:

$$BR(t) = ((p\hat{\ }\_\ell\ (NIR,t) - p\hat{\ }\_\ell\ (SWIR1,t))/(p\hat{\ }\_\ell\ (NIR,t) + p\hat{\ }\_\ell\ (SWIR1,t)\ ))$$

Equation 4-5.

Then, for a segment that spans from ordinal date t_{start} (the segment start date) to ordinal date t_{end} (the segment end date), the spectral criteria for vegetative growth and decline are as follows (per Equation 4-6 and Equation 4-7):

$$BR(t_{end}) - BR(t_{start}) > 0.05, \text{vegetative growth}$$

Equation 4-6.

$$BR(t_{end}) - BR(t_{start}) < -0.05, \text{vegetative decline}$$

Equation 4-7.

Urban Omission

If the annual prediction is cultivated crops, pasture/hay, grassland/herbaceous, or shrub/scrub for the first year of the segment and they are urban/roads in the final year of the segment, the segment is inferred to represent urban growth. Spatial probabilities were found to sufficiently capture these change events and were relied upon for change event timing. The time series data of vegetative growth often can be well-approximated by Equation 4.5 and that of urban growth defined by the LCAMS predictions.

4.1.3.2.2 Workflow

The start of the workflow uses class predictions from the annual refined probabilities as a baseline for the annual class values and works through the time series based on CCD segment partitioning, integrating information from the spatial probabilities, spatial predictions, and DEM.

1. Associate annual spatial and refined probabilities, along with their associated class predictions based on the max probability to a segment based on the start and end dates for the segment
    a. If the contained refined probabilities predict an LCAMS forest transitional class
        i. Partition the segment based on the first and last instances that forest transitional grass or forest transitional shrub are called from the refined probabilities, resulting in up to 3 partitions
        ii. Ensure that forest transitional shrub is not going to forest transitional grass at any point during the forest transitional partition
            1. If it does, change the forest transitional grass to forest transitional shrub after that point
        iii. If the start of the partition is not the same as the start of the segment, then set the class values to the mode for that section of the segment
        iv. If the end of the partition is not the same as the end of the segment, then set the class values to the mode for that section of the segment
        v. If the spatial predictions contain LCAMS forest transitional classes, then find where the spatial predictions first call a forest transitional class

            1. If the spatial predictions predicted a forest transitional class before the refined predictions, then set the values between them to the same as the refined predictions
   b. Else, if the refined predictions match **vegetative growth**
      i. Partition the segment where the refined predictions first indicate a shrub/scrub, and one of the three tree classes, resulting in up to 3 partitions assuming the first part of the segment is called grass/herbaceous
         1. All class values from the start of the segment to when shrub/scrub is first called are set to grass/herbaceous
         2. All class values from where shrub/scrub is first called to when a tree class is first called are set to shrub/scrub
         3. All class values from where a tree class is first called to the end of the segment are called tree
   c. Else, if the refined predictions match **vegetative decline**
      i. Use the inverse logic of vegetative growth
   d. Else, if the refined predictions match **urban omission**
      i. Partition the segment where the refined predictions first indicate urban/road
         1. All class values from the start of the segment to when urban/road is first called are set to the mode of the classes
         2. All class values from where urban or road is called to the end of the segment are set to the mode of urban and road occurrences
   e. Else, assign the annual class predictions for the temporal period that the segment covers based on the per-class average of the refined probabilities, similar to LCMAP Initial Classifier (LCMAP ADD Section 4.2.6.1)
2. Where spatial predictions are open water, set class values to open water
3. Where refined predictions are open water, and spatial predictions are not
   a. Assign a class value based on the per-class average of the non-water spatial probabilities for the segment
4. If the pixel location has an elevation > 800, a slope > 10, and the annual segment class values include water
   a. Set the class values for the segment based on the mode of the non-water class values, and the second predicted class value from the spatial probabilities for the water class values

Section 5 Caveats and Limitations

The Annual NLCD Collection 1 Science Products contain known caveats and limitations. Overall considerations regarding the full data set or the NLCD approach are described below.

Land Cover

- Pixelated developed, barren, and other out-of-place land cover calls over water, particularly Lake Superior, are attributed to input issues related to artifacts in the leaf-on/off imagery used.
- Linear artifacts or "blockiness" can be seen in some geographic areas, particularly the desert Southwest where the new spatial AI/ML approach had difficultly differentiating between shrub/scrub and grassland/herbaceous.

Fractional Impervious Surface

- Linear regression predictions produced values outside of the accepted 0-100 value range and were not properly truncated. This can cause values to be greater than 100 and, in some cases, cause a buffer underflow for the UINT8 data type.

Land Cover Confidence

- Linear regression predictions and the nature of the ensemble approach can give values greater than 100, which were not properly truncated.
- The wrong class can be referenced for the associated confidence value due to modeling on an expanded set of classes which get cross-walked to the final land cover calls. This often results in a much lower confidence value than intended.

Section 6 Accuracy Assessment

Each NLCD product release is followed by an accuracy assessment that is derived from an independent reference data set. NLCD produces and publishes reference data that are based on trained human interpretation of aerial photography and satellite imagery. These data are used to validate land cover and land cover change. The release of Annual NLCD Collection 1, the first NLCD product suite produced annually across the historical Landsat archive, increased the requirements for reference data collection. The sample design and response design were adapted accordingly to collect reference information for each NLCD product year.

The Annual NLCD Collection 1 sample design was developed as a two-phase collection by a team of image interpreters as follows: an initial base sample set across CONUS containing 5,000 simple random sample plots, followed by another collection of up to 5,000 stratified random sample plots. The strata used to define the sample frame for the stratified samples are based on Annual NLCD Collection 1.0 land cover and land cover change products. The fundamental sample units are 30×30 m spatial extents that align with the pixel grid of the NLCD map product suite. Coastal waters and the Great Lakes were excluded from the sample frame, as has been done in previous NLCD sample designs (Wickham et al., 2023). Every year in the time series at each spatial location is part of the sample and will be interpreted; in other words, no sampling is done across the temporal domain. As such, the full sample can be described as a stratified one-stage cluster sample.

Interpretation of the sample follows interpretation protocols and guidelines that provide pragmatic specificity to the class legend outlined in Table 2-1. Interpreters collect attributes that support the assignment of a primary class label and, where applicable, an alternate class label that reflects a “fuzzy set” conceptualization of land cover categories (Gopal & Woodcock, 1994). Accurate and efficient interpretation is enabled by a variety of Geographic Information System (GIS) software tools that support visualization of high-resolution satellite and aerial photography, as well as additional ancillary data sets. The TimeSync software tool (originally created by Cohen et al., 2010) allows the display of Landsat pixel time series and Landsat image “chips” centered on the sample unit. TimeSync also provides an integrated recording form used by interpreters to record attributes and track completion of assigned sample units.

Completion and quality assurance checks of the reference interpretations across the full sample built upon the methods described in Pengra et al., 2020. This was followed by a pairing of the reference data with the land cover and land cover change map values at the sample locations. This data set is used to generate statistical estimates of common accuracy metrics (c.f. Wickham et al., 2023; Stehman et al., 2021).

The Annual NLCD Collection 1 Reference and Validation Data for 1985-2023 will be available after collection is completed.

Section 7 User Services

Annual NLCD Collection 1 Science Products and associated interfaces are supported by USGS User Services staff at the USGS Earth Resources Observation and Science (EROS) Center. Questions or comments regarding Annual NLCD Collection 1 Science Products or interfaces are welcome. Email can be sent to USGS User Services with the topic indicated in the subject line.

USGS User Services
605-594-6151
1-800-252-4547
custserv@usgs.gov

User support is available Monday through Friday from 8:00 a.m. – 4:00 p.m. Central Time. Inquiries received outside of those hours are addressed the next business day.

References

Anderson, J. R., Hardy, E. E., Roach, J. T., and Witmer, R. E. (1976). A land use and land cover classification system for use with remote sensor data. U.S. Geological Survey Professional Paper 964. https://doi.org/10.3133/pp964

Cohen, W. B., Yang, Z., & Kennedy, R. (2010). Detecting trends in forest disturbance and recovery using yearly Landsat time series: 2. TimeSync—Tools for calibration and validation. *Remote Sensing of Environment*, *114*(12), 2911-2924. https://doi.org/10.1016/j.rse.2010.07.010

Fuller, R. M., Smith, G. M., & Devereux, B. J. (2003). The characterization and measurement of land cover change through remote sensing: problems in operational applications? International Journal of Applied Earth Observation and Geoinformation, 4(3), 243–253. https://doi.org/10.1016/S0303-2434(03)00004-7

Gopal, S., & Woodcock, C. (1994). Theory and methods for accuracy assessment of thematic maps using fuzzy sets. *Photogrammetric Engineering and Remote Sensing*, *60*(2). https://www.asprs.org/wp-content/uploads/pers/1994journal/feb/1994_feb_181-188.pdf

Hoen, B.D., Diffendorfer, J.E., Rand, J.T., Kramer, L.A., Garrity, C.P., and Hunt, H.E., (2018). United States Wind Turbine Database v6.1 (November 28, 2023): U.S. Geological Survey, American Clean Power Association, and Lawrence Berkeley National Laboratory data release, https://doi.org/10.5066/F7TX3DN0.

Jin, S., Dewitz, J., Danielson, P., Granneman, B., Costello, C., Smith, K., & Zhu, Z. (2023). National Land Cover Database 2019: A new strategy for creating clean leaf-on and leaf-off Landsat composite images. *Journal of Remote Sensing*, *3*, 0022. https://doi.org/10.34133/remotesensing.0022

Kruitwagen, L., Story, K. T., Friedrich, J., Byers, L., Skillman, S., & Hepburn, C. (2021). A global inventory of photovoltaic solar energy generating units. *Nature*, *598*(7882), 604–610. https://doi.org/10.1038/s41586-021-03957-7; data set: https://doi.org/10.5281/zenodo.5005868

Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., Glocker, B., & Rueckert, D. (2018). Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999. https://doi.org/10.48550/arXiv.1804.03999

Open Geospatial Consortium (2023). OGC Cloud Optimized GeoTIFF Standard. OGC 21-026, version 1.0. Joan Maso, editor. https://docs.ogc.org/is/21-026/21-026.html

Pengra, B. W., Stehman, S. V., Horton, J. A., Dockter, D. J., Schroeder, T. A., Yang, Z., Cohen, W. B., Healey, S. P. & Loveland, T. R. (2020). Quality control and assessment of interpreter consistency of annual land cover reference data in an operational national monitoring program. *Remote Sensing of Environment*, *238*, 111261. https://doi.org/10.1016/j.rse.2019.111261

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference*, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18 (pp. 234–241). https://arxiv.org/pdf/1505.04597

Stehman, S. V., Pengra, B. W., Horton, J. A., & Wellington, D. F. (2021). Validation of the U.S. Geological Survey's Land Change Monitoring, Assessment and Projection (LCMAP) Collection 1.0 annual land cover products 1985–2017. *Remote Sensing of Environment*, *265*, 112646. https://doi.org/10.1016/j.rse.2021.112646

Tollerud, H. J., Zhu, Z., Smith, K., Wellington, D. F., Hussain, R. A., & Viola, D. (2023). Toward consistent change detection across irregular remote sensing time series observations. *Remote Sensing of Environment*, *285*, 113372. https://doi.org/10.1016/j.rse.2022.113372

U.S. Department of Agriculture Natural Resources Conservation Service (2023). Gridded Soil Survey Geographic (gSSURGO) Database for the Conterminous United States. Available online at https://gdg.sc.egov.usda.gov/.

U.S. Department of Transportation (2023). The North American Rail Network (NARN) Rail Lines data set. National Transportation Atlas Database (NTAD) of the Bureau of Transportation Statistics (BTS). https://doi.org/10.21949/1528950

U.S. Fish and Wildlife Service (2023). The USFWS National Wetlands Inventory. https://www.fws.gov/wetlands

U.S. Geological Survey (2014). National Land Cover Database (NLCD) 2011 Land Cover Conterminous United States: U.S. Geological Survey data release, https://doi.org/10.5066/P97S2IID.

U.S. Geological Survey (2020). U.S. Landsat Collection 1 (C1) Analysis Ready Data (ARD) Data Format Control Book (DFCB). LSDS-1873, version 7.0. https://www.usgs.gov/media/files/landsat-collection-1-us-analysis-ready-data-dfcb

U.S. Geological Survey (2022a). Land Change Monitoring, Assessment, and Projection (LCMAP) Collection 1.3 Continuous Change Detection and Classification (CCDC) Algorithm Description Document (ADD) LSDS-2345, version 1.0. https://www.usgs.gov/media/files/lcmap-collection-13-ccdc-add

U.S. Geological Survey (2022b). LANDFIRE 2020 Elevation (ELEV), Aspect (ASP), and Slope Degrees (SlpD), CONUS. LF 2.2.0. https://landfire.gov/

U.S. Geological Survey (2023a). National Land Cover Database (NLCD) 2021 Products: U.S. Geological Survey data release, https://doi.org/10.5066/P9JZ7AO3.

U.S. Geological Survey (2023b). Landsat Collection 2 U.S. Analysis Ready Data Data Format Control Book. LSDS-1435, version 4.0. https://www.usgs.gov/media/files/landsat-collection-2-us-analysis-ready-data-dfcb

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5998–6008. https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

Wickham, J., Stehman, S. V., Sorenson, D. G., Gass, L., & Dewitz, J. A. (2023). Thematic accuracy assessment of the NLCD 2019 land cover for the conterminous United States. *GIScience & Remote Sensing*, *60*(1), 2181143. https://doi.org/10.1080/15481603.2023.2181143

Zhu, Z., & Woodcock, C. E. (2014). Continuous change detection and classification of land cover using all available Landsat data. *Remote sensing of Environment*, *144*, 152–171. https://doi.org/10.1016/j.rse.2014.01.011

Appendix A Acronyms

ADD	Algorithm Description Document
AEA	Albers Equal Area
ARD	Analysis Ready Data
ASCII	American Standard Code for Information Interchange
AWS	Amazon Web Services
BFP	Band First Probability
BT	Brightness Temperature
CCB	Configuration Control Board
CCD	Continuous Change Detection
CCDC	Continuous Change Detection & Classification
COG	Cloud-Optimized GeoTIFF
CONUS	Conterminous United States
CR	Change Request
CS	Coordinate System
CSDGM	FGDC Content Standard for Digital Geospatial Metadata
DEM	Digital Elevation Model
DFCB	Data Format Control Book
DOI	Digital Object Identifier
DOY	Day of Year
EE	EarthExplorer
EPSG	European Petroleum Survey Group
EROS	Earth Resources Observation and Science
ETM+	Enhanced Thematic Mapper Plus
EVA	Enhanced Visualization and Analysis Tool
FGDC	Federal Geographic Data Committee
FSP	USGS Fundamental Science Practices
GIS	Geographic Information System
gSSURGO	gridded Soil Survey Geographic Database
GeoTIFF	Georeferenced Tagged Image File Format
HTTP	Hypertext Transfer Protocol
LCAMS	Land Cover Artificial Mapping System
LCMAP	Land Change Monitoring, Assessment and Projection
LCNext	Land Cover Next
LSDS	Land Satellites Data System
MLP	Multi-layer Perceptron
MRLC	Multi-Resolution Land Characteristics (MRLC) Consortium
NIR	Near-Infrared
NLCD	National Land Cover Database
NRCS	Natural Resources Conservation Service

NWI	National Wetlands Inventory
OGC	Open Geospatial Consortium
OLI	Operational Land Imager
POSC	Petrotechnical Open Software Corporation
PyCCD	Python Continuous Change Detection
QA_PIXEL	Pixel Quality Assessment Band
RGB	Red, Green, Blue Color Values
RMSE	Root Mean Square Error
S3	Amazon Simple Storage Service
SDC	USGS Science Data Catalog
SLC	Scan Line Corrector
SPUG	Science Product User Guide
SR	Surface Reflectance
STIPP	Spatial and Temporal Integrated Probability-based Post-processing
SWIR1	Short Wavelength Infrared
.tif	Georeferenced Tagged Image File Format – file extension
TIFF	Tagged Image File Format
TIRS	Thermal Infrared Sensor
TM	Thematic Mapper
UINT	Unsigned Integer
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
USDA	U.S. Department of Agriculture
USFWS	U.S. Fish and Wildlife Service
USGS	U.S. Geological Survey
UTM	Universal Transverse Mercator
VMVP	Virtual Median Value Point
WGS	World Geodetic System
WPI	Wetland Potential Index
.xml	eXtensible Markup Language – file extension
XML	eXtensible Markup Language