Add Health
The National Longitudinal Study of Adolescent to Adult Health

Add Health as a Resource for the Science of the Exposome:

National Land Cover Database (NLCD) Neighborhood Land Cover Measures

Report prepared by

Andrea N. Goodwin

James D. Stewart

Robert A. Hummer

Eric A. Whitsel

UNC CAROLINA POPULATION CENTER

CAROLINA POPULATION CENTER | CAROLINA SQUARE - SUITE 210 | 123 WEST FRANKLIN STREET | CHAPEL HILL, NC 27516

https://doi.org/10.17615/k4y8-hy03

This user guide is one in a set of user guides focusing on the built, environmental, and natural features of geopositioned/geocoded Add Health respondent locations over Waves I-VI. Collectively, they describe exposomic measures in the following three domains:

Built Domain	**Environmental Domain**	**Natural Domain**
Commuting Area	Ambient Air	Altitude
Land Use	Indoor Air	Meteorology
Roadway Proximity/Density	Noise	Green space
	Waterborne Lead	Blue space
	Nighttime Light Pollution	
	Solar Irradiation	

Under the Built Domain, this particular user guide summarizes the rationale for the construction and assignment of land use. It also documents how the land use source data were acquired, as well as the protocol for quality controlling their assignment and classification across waves. Whenever possible, construction, assignment, and classification were harmonized to ensure temporal comparability, although important inter-wave differences exist and are grey-highlighted herein.

Acknowledgement

Data for Wave VI of Add Health was supported by two cooperative agreements from the National Institute on Aging (1U01AG071448, principal investigator Robert A. Hummer, and 1U01AG071450, principal investigators Robert A. Hummer and Allison E. Aiello) and a special supplement (U01-AG071450-02S1, principal investigators Robert A. Hummer, Allison E. Aiello, and Eric A. Whitsel) to the University of North Carolina at Chapel Hill. Co-funding for Wave VI was provided by the Eunice Kennedy Shriver National Institute of Child Health and Human Development, the National Institute on Minority Health and Health Disparities, the National Institute on Drug Abuse, the NIH Office of Behavioral and Social Science Research, and the NIH Office of Disease Prevention. Data from Waves I-V of Add Health are from the Add Health Program Project, grant P01 HD31921 (Kathleen Mullan Harris) from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Add Health was originally designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill. Add Health is currently directed by Robert A. Hummer; it was previously directed by Kathleen Mullan Harris (2004-2021) and J. Richard Udry (1994-2004). Information on obtaining Add Health data is available on the project website (https://addhealth.cpc.unc.edu).

Citation for User Guide

Goodwin AN, Stewart JD, Hummer RA, Whitsel EA. Add Health as a Resource for the Science of the Exposome**:** *Land Use.* Chapel Hill, NC: Carolina Population Center, University of North Carolina at Chapel Hill. *Available from: https://doi.org/10.17615/k4y8-hy03*

.

Table of Contents

1. Introduction

The National Longitudinal Study of Adolescent to Adult Health (Add Health) is a nationally representative sample of U.S. adolescents who were in grades 7-12 during the 1994-1995 school year. Using a complex, school-based cluster-sampling frame, researchers selected high school and feeder school pairs from 80 communities across the United States and drew a sex- and grade-stratified random sample of 20,745 adolescents for inclusion in the study. This sample has been followed from adolescence into early midlife across six waves of data collection to date, with the most recent wave of data collection (Wave VI) taking place between 2022 and 2025 when respondents were ages 39 to 49.

Over the years, Add Health has collected a wealth of information from respondents and their parents about demographic characteristics, familial structures, social relationships, health behaviors, cognition, physical and mental health status, medication usage, and health care access. Add Health also has collected anthropometric, cardiovascular, metabolic, renal, hepatic, inflammatory/immune, infectious, neurodegenerative, and multi-omic biomarkers from respondents. In addition, Add Health has merged multilevel contextual data about the economic, school, neighborhood, policy, and environmental contexts in which the respondents are embedded to the core survey and biological data at each wave. The Add Health dataset thereby provides researchers with rich opportunities to explore the causes and consequences of health status across multiple contextual domains as individuals age across the life course.

This user guide is one in a series documenting the latest contextual and environmental data assembled under the exposome supplement introduced in the preceding acknowledgment. Collectively, the supplemental data and documentation enable researchers to examine a broader array of built, environmental, and natural exposures linked to accurately geopositioned/geocoded Add Health respondent residences from Wave I through Wave VI. Because Wave VI data are not ready for geocoding or dissemination at present, this user guide and the associated data are focused on Wave I-V linkages. The Add Health Team will update this data set and user guide when Wave VI data are available for dissemination.

2. General Overview

The land use measures include the land areas (in meters squared) surrounding geocoded respondent residences that are classified as developed, forested, etc. The data file including them is based on data from the National Land Cover Database (NLCD). The rationale for and utility of acquiring the land use measures is described below.

2.1 Rationale

Since its inception, Add Health has continued amassing and disseminating contextual data files across multiple levels of geography, thus resulting in an increasingly comprehensive and diverse set of contextual measures in a nationally representative study spanning adolescence to mid adulthood. In general, these data have been provided to establish infrastructure for research addressing the role of diverse exposures across multiple levels and across the life course in the etiology and disparities of our most pressing health issues. The data collectively position Add Health as a central resource for scientists to more effectively operationalize and study the exposome and its consequences for population health across the life course, with particular attention to disparities across population subgroups.

2.2 Utility

The land use data described herein expand the contextual data available to Add Health researchers, enhancing their capacity to examine the social, environmental, and biological dimensions of the exposome and how they contribute to U.S. population health and disparities. The land use data may be valuable to researchers who study built and natural features of neighborhood environments[1] (including exposures to development, green space, blue space, etc.), insect-borne diseases[2,3], and disparities in health[4]. Additionally, land use measures may also enhance research centered on disparities in access to health-promoting resources.

3. Processing Details

To provide measures of land use in relation to Add Health respondents' residential locations over time, NLCD class information was extracted and summarized based on temporal proximity within "neighborhoods", operationalized as five different circular buffers centered on respondents' available geocoded addresses. Buffer radii were 100, 200, 300, 400 and 500 meters. Since temporal availability of NLCD source data varied by geographic region, assignment of source data to respondents was simplified by separating respondents into four groups according to the geographic regions that applied to their full address histories: Lower 48 states only, Alaska and Lower 48; Hawaii and Lower 48, and Alaska and Hawaii. This greatly facilitated assignment of source data to respondents. Temporal proximity details are described in **Table 1**.

Table 1. Temporal Assignment Crosswalk by Geographic Region based on Full Address Histories

Respondent Location Date Range	NLCD Source Year Lower 48 Only	NLCD Source Year Alaska-Lower 48		NLCD Source Year Hawaii-Lower 48		NLCD Source Year Alaska-Hawaii	
		Lower 48	Alaska	Lower 48	Hawaii	Alaska	Hawaii
Through Dec 1996	1992	1992	2001	1992	2001	2001	2001
Jan 1997 – Dec 2002	2001	2001	2001	2001	2001	2001	2001
Jan 2003 – Jun 2003	2004	2004	2001	2004	2001	2001	2001
Jul 2003 – Jun 2005	2004	2004	2001	2004	2005	2001	2005
Jul 2005 – Dec 2006	2006	2006	2001	2006	2005	2001	2005
Jan 2007 – Jun 2007	2006	2006	2011	2006	2005	2011	2005
Jul 2007 – Dec 2007	2008	2008	2011	2008	2005	2011	2005
Jan 2008 – Dec 2009	2008	2008	2011	2008	2010	2011	2010
Jan 2010 – Jun 2012	2011	2011	2011	2011	2010	2011	2010
Jul 2012 – Dec 2013	2013	2013	2011	2013	2010	2011	2010
Jan 2014 – Dec 2014	2013	2013	2016	2013	2010	2016	2010
Jan 2015 – Dec 2017	2016	2016	2016	2016	2010	2016	2010
Jan 2018 – Jan 2019	2019	2019	2016	2019	2010	2016	2010

3.1 Acquisition of National Land Cover Data

Eight Landsat-based, 30-meter resolution NLCD sets for 2001-2019 were downloaded from the Multi-Resolution Land Characteristics Consortium (MRLC) website.[5] Each data set was stored as a very large raster (pixel-based) file in ERDAS Imagine (.img) format covering the entire contiguous 48 states of the U.S.

(contiguous U.S., CONUS, or lower 48), and was projected to the Albers Conical Equal Area projection using the World Geodetic System 1984 (WGS84) datum. Processing continued with acquisition of NLCD 1992 from the U.S. Geological Survey (USGS) through special request. NLCD 1992 data also were spatially referenced to the Albers Conical Equal Area projection, but using the North America Datum 1983 (NAD83). In addition to NLCD data sets gathered for the CONUS, NLCD Alaska data for available years 2001, 2011, and 2016 were downloaded. Like more recent NLCD data sets for the CONUS, NLCD Alaska data sets were projected to Albers Conical Equal Area using the WGS84 datum, albeit with a central meridian, standard parallels, and latitude of origin specific to Alaska.

Hawaii NLCD data for 2001 were available with spatial characteristics very similar to those for the other NLCD data sets. Hawaii data for 2005 and 2010, however, came from the Coastal Change Analysis Program (C-CAP), which for Hawaii provides land cover data for the NLCD database at the much higher spatial resolution of 2.4 meters and projected to Universal Transverse Mercator (UTM) Zone 4. For details on source data sets, see **Table 2**.

Table 2. NLCD Source Data Details

NLCD Year	Geography	File Size	Row Count	Column Count	Projection	Datum
1992	CONUS	13.94 GB	96,995	154,264	Albers Conical Equal Area (Central Meridian -96)	NAD83
2001 2004 2006 2008 2011 2013 2016 2019	CONUS	15.68 GB	104,424	161,190	Albers Conical Equal Area (Central Meridian -96)	WGS84
2001 2011 2016	Alaska	7.85 GB	67,844	124,236	Albers Conical Equal Area (Central Meridian -154)	WGS84
2001	Hawaii	233.92 MB	12,618	19,439	Albers Conical Equal Area (Central Meridian -157)	WGS84
2005*	Hawaii	365.69 MB	17,593	21,796	UTM Zone 4	NAD83
2005*	Hawaii	554.39 MB	21,129	27,513	UTM Zone 4	WGS84
2010*	Hawaii	649.97 MB	20,986	32,476	UTM Zone 4	WGS84
2010*	Hawaii	3.22 GB	62,941	54,903	UTM Zone 5	WGS84

* Hawaii data for 2005 and 2010 were available from the C-CAP program on an island-specific basis, and with different datums, which required harmonization during source data pre-processing.

3.2 NLCD Source Data Pre-processing

To minimize spatial distortion when calculating land cover areas within circular buffers centered on respondents' geocoded residential locations, all source data sets were subset and reprojected to the UTM coordinate system (see **Figure 1**), which uses the WGS84 datum. During this step, year-specific projection parameters had to be used for the source data sets, because NLCD 1992 data were stored using the NAD83 datum, whereas NLCD data for later years were stored primarily using the WGS84 datum. The offset between the NAD83 and WGS84 datums is generally less than a meter within the contiguous U.S., but taking datum into account when projecting to UTM eliminated an unnecessary source of spatial error. In addition,

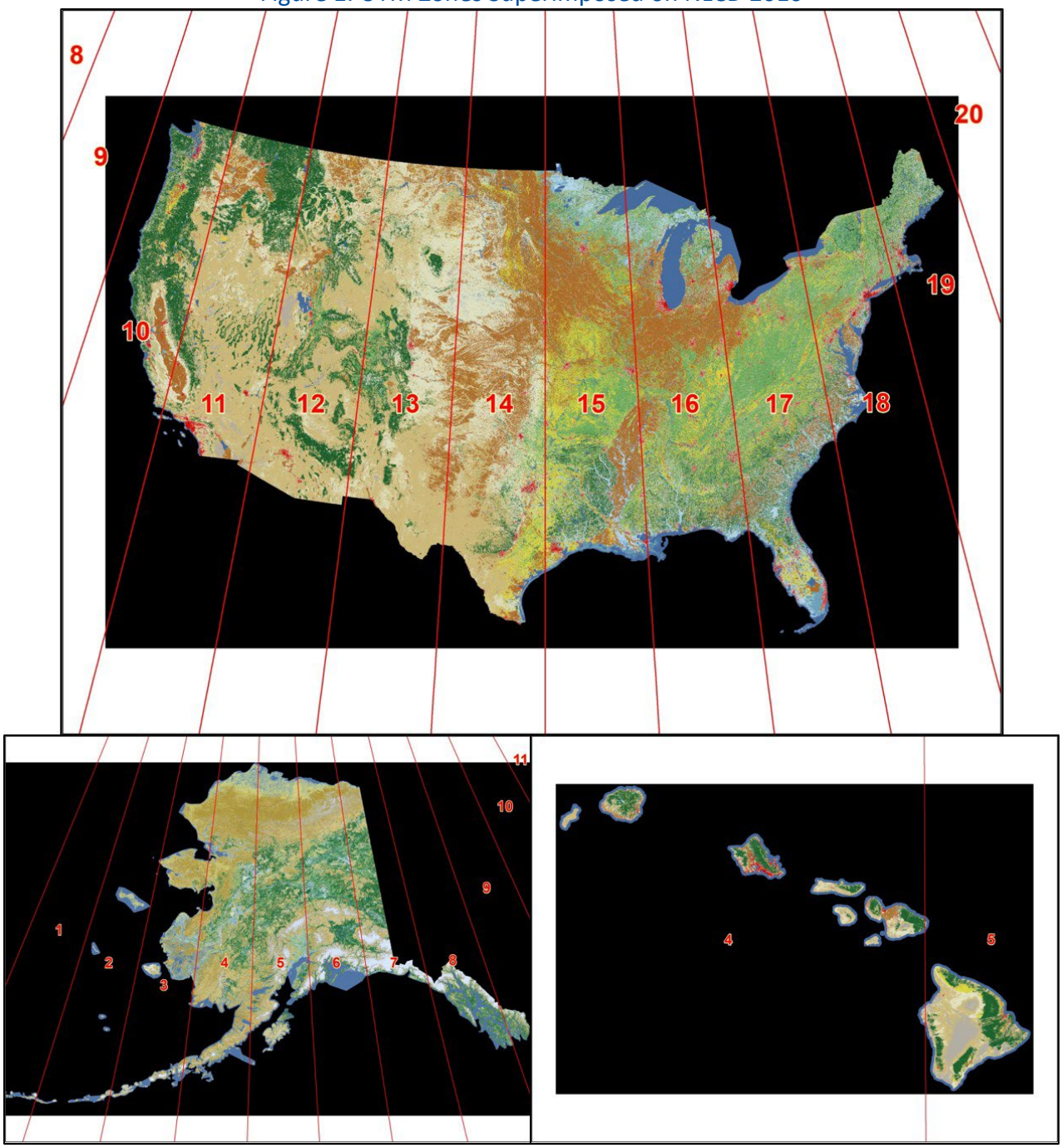
NLCD data for Alaska had different specifications than the CONUS for the Albers Conical Equal Area projection, and NLCD data for Hawaii were provided in a variety of projections and datums as well. Pre-processing of the NLCD data sets therefore addressed differences in source data projections and created consistent, harmonized data layers for analysis based on the UTM coordinate system.

With respect to Hawaii, C-CAP data for 2005 and 2010 were provided on an island-specific basis. To simplify processing, these island-specific files were first merged by UTM zone using the ArcGIS Desktop version 10.8.1 MosaicToNewRaster command (Management toolbox). Also, since the C-CAP program employs a classification scheme that provides more detail on wetlands, its numbering scheme differs from that used for the general NLCD program. Since the C-CAP program is the source of NLCD data for coastal areas, however, C-CAP classes translate directly to NLCD classes beginning in 2001. For consistency with the other NLCD data sets, therefore, the Hawaii land cover classes assigned by the C-CAP program for 2005 and 2010 were converted to NLCD land cover classes during pre-processing. To view the C-CAP-to-NLCD class conversion crosswalk, see **Table 8** in **Appendix II**.

To reduce time related to processing so many large raster images, the source data sets were projected to the UTM coordinate system using a set of complementary Python/ArcGIS scripts (Python 2.7.18 with ArcGIS Desktop 10.8.1) to run batches of 15 parallel sessions to process a total of 105 individual jobs (CONUS = nine NLCD source years x 10 UTM zones = 90 jobs; Alaska = three NLCD source years x 3 UTM zone = 9 jobs; Hawaii = three source years x 2 UTM zones = 6 jobs). Note that some Hawaii source data sets were already projected correctly to the UTM coordinate system, but during pre-processing were standardized with respect to name and file storage location. To eliminate large negative color values in the NLCD 1992 source data, which were used for unclassified areas and which generated errors when processing the file in ArcGIS, a separate Python/ArcGIS script was run to add the correct color map (legend) and to update the land cover class values and corresponding descriptions stored in the NLCD 1992 attribute table.

Converting NAD83 source data to WGS84 in UTM for NLCD 1992 was performed using the geographic datum transformation method WGS_1984_(ITRF00)_To_NAD_1983, which is accurate to within 0.1 m within the CONUS and Alaska. Converting NAD83 source data to WGS84 for Hawaii leveraged a combination of the NAD_1983_To_HARN_Hawaii transformation, which is accurate to within 0.05 m, plus the WGS_1984_(ITRF00)_To_NAD_1983_HARN transformation, which is accurate to within 0.1 m for all states. Details regarding geographic transformations can be found elsewhere.[6]

Figure 1. UTM Zones Superimposed on NLCD 2010



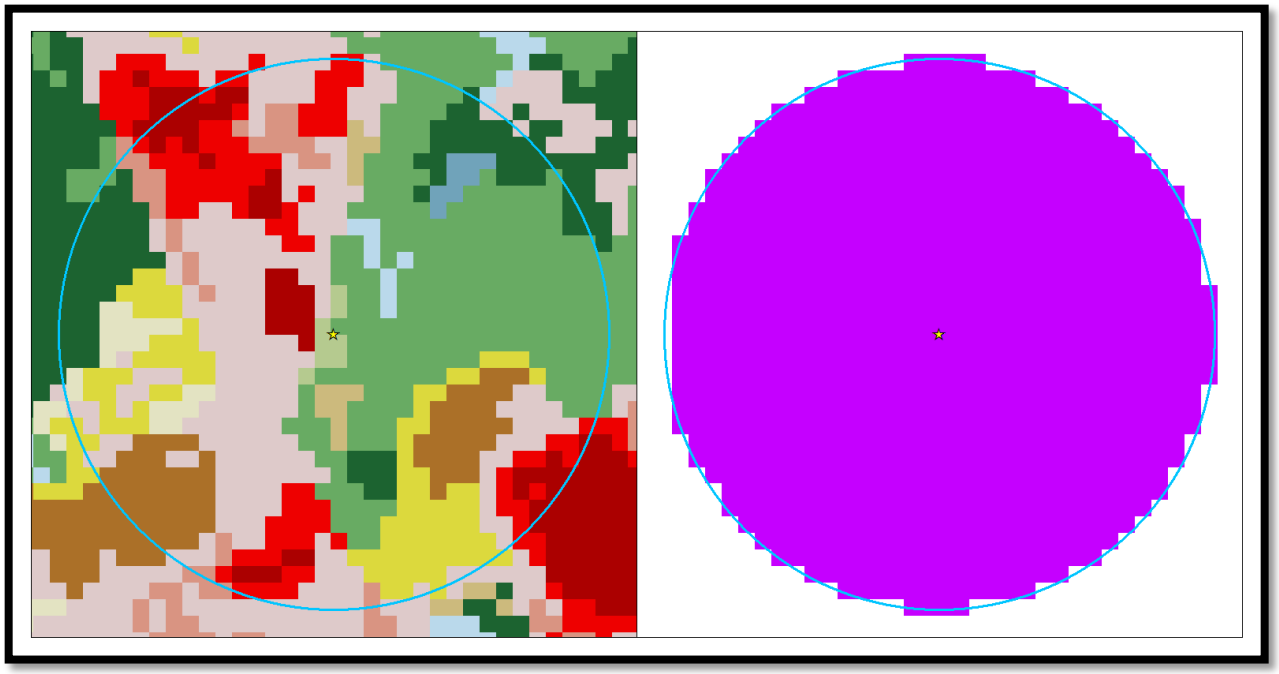
3.3 Respondent Locations Pre-processing

To ensure correct and consistent spatial alignment of respondent locations in relation to NLCD source data, respondent geocoded locations were also subset and projected to the UTM coordinate system.

3.4 Extraction of Neighborhood Land Cover Measures

Extracting NLCD raster data using a circular buffer, which selects pixels or grid cells based on center point coordinates (centroids), can lead to jagged edges that omit land cover information along the buffer boundary (see **Figure 2**). Also, the very large file sizes of the NLCD raster source data can easily overwhelm the processing capabilities of most computer systems. For these two key reasons, the NLCD data in each UTM zone were first subset using the ArcGIS ExtractByMask command (Spatial Analyst extension) to include only pixels with centroids that fell within 530-m buffers centered on respondent locations.

Figure 2. Raster Data Extracted using Circular Buffer

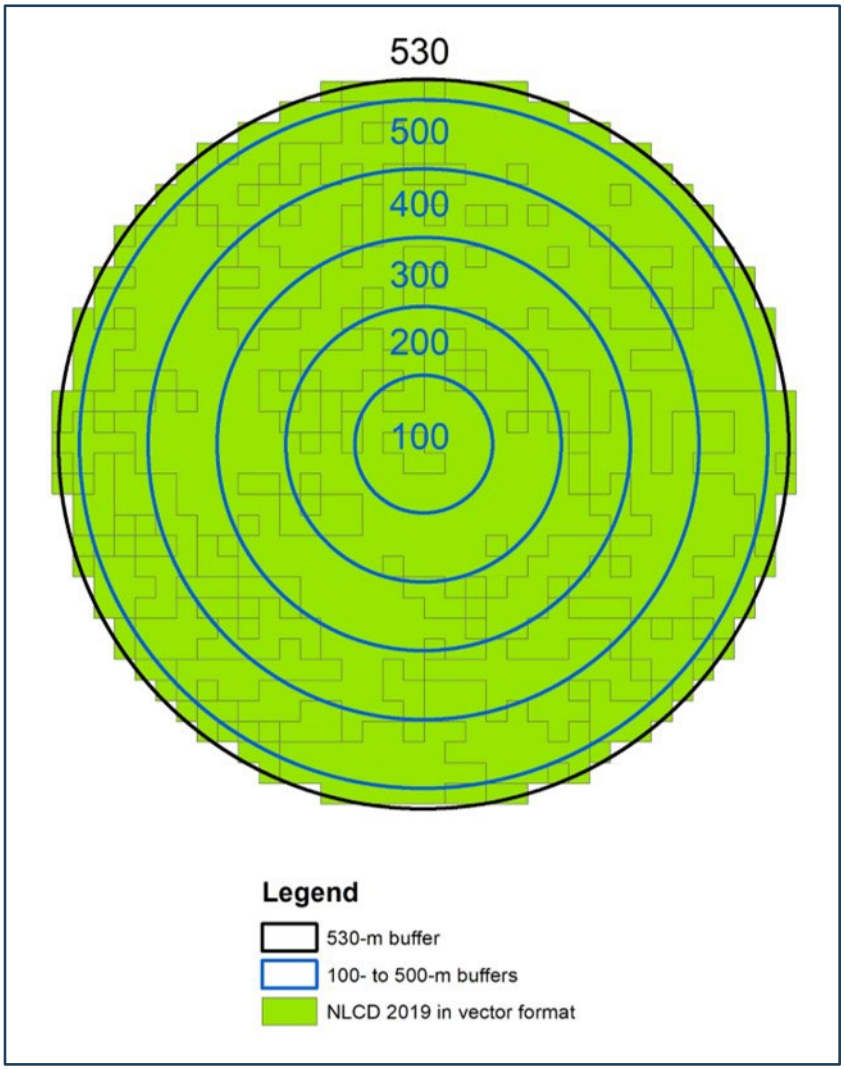


The choice of 530 meters was based on the geometry of the NLCD 30-m by 30-m pixels, which have a hypotenuse of length 42.43 meters. The midpoint along the hypotenuse, which coincides with the centroid of the pixel, is located around 21 meters, so padding the 500-m buffer radius by an extra 30 meters ensured the selection of at least one pixel in all directions outside the 500-m radius, thereby eliminating gaps or shortfalls in pixel selection (see **Figure 3**).

To reduce file sizes and to maximize accuracy in extracting land cover area by class from the NLCD subsets by UTM zone, the raster subsets were first converted to vector format (simple points, lines, and polygons) using the ArcGIS RasterToPolygon command (Conversion toolbox). This obviated the need to work with raster pixels, which can require significant storage space and are spatially referenced by just their corner and center coordinates. This allowed the extraction of smaller file sizes and more precisely defined polygon intersections with the circular boundaries of 100- to 500-m buffers. To allow extraction of vectorized NLCD data for all five buffer radii simultaneously, the circular buffers centered on respondent locations were constructed as multi-ring buffers (see **Figure 3**), using the ArcGIS MultipleRingBuffer command (Analysis toolbox).

Once converted to vector format, NLCD pixel-based data within 530 meters of respondent locations were extracted as efficiently and precisely as possible within 100- to 500-m buffers using ArcGIS vector intersection commands. To avoid running out of memory when processing densely overlapping multi-ring buffers for geographic regions with many study respondents in close proximity to one another, land cover extraction within the multi-ring buffers was accomplished using the ArcGIS TabulateIntersection command (Analysis toolbox). This command is more capable of dealing with densely overlapping multi-ring polygons by virtue of the fact that it does not create graphical output, but rather summarizes results as a table. To ensure that computing resources were not exceeded, processing was limited to respondent record counts of 250 per individual processing job (see **Table 3**).

Figure 3. Raster Data Extracted using Expanded Multi-ring Circular Buffer



Using the ArcGIS TabulateIntersection command, NLCD land cover area by class within each of the 100- to 500-m buffers around respondent locations was summarized into job-specific dBASE format tables (.dbf). The DBF tables were imported into SAS for construction of the final data set, which included summarizing total area for each respondent by summing individual class areas by buffer radius.

3.5 Parallel Processing

Pre-processing and extraction of data from the very large NLCD raster data sets were computer-intensive and time-consuming. Based on the need to process the data within the Windows operating system, the work could not have been accomplished within a reasonable amount of time without parallel processing. As a result, the processing pipeline was constructed to take advantage of 16 available cores and 48 GB of RAM. Holding one core in reserve for overall control, NLCD data were pre-processed and extracted in batches of 15 simultaneous jobs on targeted cores. Limiting batch sizes for respondents' multi-ring buffers to 250 to avoid overwhelming the system required up to hundreds of jobs per source year. Leveraging multiple cores reduced overall processing time from an estimated two-plus weeks to under 19 hours. As a result, parallel processing was an indispensable tool in the development of the NLCD data set.

Table 3. Parallel Processing Details

UTM Zone	A. Count of Source Data Years	B. Total Record Count (Unique Geocodes x Five Buffers)	C. Jobs Required to Process Total Record Count (B / 250 Rounded Up)	D. Total Job Count (A x C)	F. Batch Count (D / 15 Cores Rounded Up)
4	3	4,580	19	57	
5	6	100	1	6	
6	3	95	1	3	
8	3	10	1	3	
10	9	27,805	112	1,008	
11	9	47,705	191	1,719	
12	9	5,370	22	198	
13	9	10,010	41	369	
14	9	28,265	114	1,026	
15	9	42,070	169	1,521	
16	9	67,915	272	2,448	
17	9	97,725	391	3,519	
18	9	46,035	185	1,665	
19	9	8,440	34	306	
Total		386,125		13,848	924

3.6 Quality Control Checks

Quality control was infused into the entire development process through vetting of output from every software command used by first running the command against a small sample of data before full execution. Beyond step-by-step validation of inputs and outputs during data set development, post-processing quality control checks included verifying the accuracy and consistency of total area ranges by buffer radius and geographic region, verifying the accuracy of respondent-level minimum and maximum class areas by radius and region, and verifying accuracy of a random sample of respondent-level results by radius and region.

3.6.1 Verification of Total Area Ranges by Buffer Radius and Geographic Region

Verification of total areas by buffer radius focused on the generation of univariate statistics and histograms, which varied nominally across buffer radii. Because total areas were calculated as the sums of class areas, and total areas closely approximated the expected areas (i.e., πr^2), concern that there were significant differences in area calculations by UTM zone was eliminated. To be sure, total areas were calculated by UTM zone as well, which reinforced the conclusion that there was no distinct spatial pattern to the differences in area calculation beyond the slightly larger difference as buffer radius increased. For total area summary statistics by buffer radius, see **Table 4**.

Table 4: Total Area Summary Statistics

Buffer Radius (m)	Minimum Total Area (m^2)	Maximum Total Area (m^2)	Difference (m^2)
100	31,415.2	31,416.2	1.0
200	125,662.5	125,663.8	1.3
300	282,741.6	282,743.4	1.8
400	502,652.3	502,655.3	3.0
500	785,395.3	785,398.2	2.9

3.6.2 Verification of Respondent-Level Minimum and Maximum Class Areas by Radius and Region

Verification of class area values started with the generation of univariate statistics by land cover class to identify minimum and maximum values. From the full collection of 140 summary tables organized by land cover class and buffer radius (28 possible classes x five radii), a handful of minimum and maximum values were selected to represent each of the primary geographic regions (Lower 48, Alaska, and Hawaii) for manual validation. Manual validation involved interactive examination of source data sets and intermediate data products within ArcGIS, including detailed inspection of both the raster and vector versions of NLCD source data in relation to the multi-ring buffers for the respondent for which the minimum or maximum value was calculated. Inspection involved verification of area values by re-calculating them based on physical measurements in the ArcGIS graphical interface as well as interactive calculation of area values using the Summarize function within the ArcGIS feature class attribute table fields. All checks were successful.

In a similar manner, SAS and ArcGIS were also used to verify the accuracy of replacement codes for respondents with missing coordinates (-9990) and missing data for a particular source year (-9992). All checks were successful.

3.6.3 Verification of Random Sample of Respondent-Level Results by Radius and Region

To verify a random sample of respondent-level results, the SAS proc surveyselect function was used to generate a random sample of 10 respondents for each of the primary geographic regions. Based on the highly time-consuming nature of manual verifications using the ArcGIS graphical interface, a few of the 10

respondents for each region were selected to obtain a varied, representative sample of different radii and land cover classes. As with the verification of respondent-level minimum and maximum class areas, manual validation involved interactive examination of source data sets and intermediate data products within ArcGIS, including detailed inspection of both the raster and vector versions of NLCD source data in relation to respondents' multi-ring buffers. Inspection involved verification of area values by re-calculating them based on physical measurements in the ArcGIS graphical interface as well as interactive calculation of area values using the Summarize function within the ArcGIS feature class attribute table fields. All checks were successful.

4. Missing codes

When a respondent's geographic coordinates were missing for a particular residential address date range, values for NLCD land cover class area and total area variables (see **Table 3**) were set to -9990 (missing coordinates). When a land cover class was valid for the NLCD source year for a particular date range, but was absent from a respondent's buffer, it received a variable value of 0. When a land cover class did not apply to the NLCD source year for a particular date range, it received a variable replacement value of -9992 (missing data). The replacement code of -9992 was also used for both NLCD land cover class area and total area variables if there were no NLCD source data available for the time period of interest. See **Table 7** in **Appendix II** for a relevant comparison of land cover classes by NLCD source year.

5. Usage Notes

5.1 Temporal Comparability

According to the MRLC, "*The National Land Cover Database (NLCD) provides nationwide data on land cover and land cover change at a 30 m resolution…based on a modified Anderson Level II classification system.*"[7] The first NLCD data set was published in 1992, but based on methodology changes since its release, the NLCD 1992 data set has been archived and is available only upon request. Beginning in 2001, NLCD data sets were transitioned from land use classes to spectrally derived classes and harmonized to facilitate multi-temporal analyses. As a result, NLCD data from 2001 and later are directly comparable, whereas comparison to 1992 data requires the construction of a crosswalk. For this reason, the USGS, the lead developer of the NLCD, has issued a caveat emptor (buyer beware) warning about comparing NLCD data from 1992 to NLCD data from later years.

5.2 Geographic Heterogeneity

NLCD data are primarily available for the contiguous United States ("lower 48"), although the geographic coverage does include Alaska and Hawaii on a less frequent basis. The Alaska NLCD data begin in 2001, and follow the same classification scheme as that for the lower 48, albeit with a few Alaska-only classes. The Hawaii NLCD data for 2001 are consistent with those for Alaska and the lower 48, although the Hawaii data for 2005 and 2010 are from the C-CAP, and had to be converted to NLCD classes. For a full description of

land cover classes by NLCD source year, see **Appendix I**. For crosswalks of land cover classes by NLCD source year, see **Appendix II**.

5.3 Proper Denominators

Because total areas represent the sum of all individual land cover classes present within a given buffer radius for each respondent, they should be used as the denominators when calculating buffer-specific proportions covered by specific land cover classes for a period of interest.

6. Data File

6.1 Structure

The land use data file is provided as a multiple-records-per-respondent long file comprised of 149 variables linked to 199,821 observations. The data file including these observations is based on national land cover data for respondent years ranging from 1992 to 2019. Consistent with other Add Health data, the 20,745 Add Health Wave I respondents are identified by a masked respondent identifier (AID) at every time period during follow-up as presented by the date from (RMELNDCOVDFR) and date to (RMELNDCOVDTO) variables establishing the start and end of each period. Please consult the accompanying codebook for additional details.

6.2 Contents

The land use data file includes the variables below, which are described in the corresponding codebook documentation that also contains frequencies. For each land cover class, there are five area variables (asterisked below), each one corresponding to one of five buffer radii (R = 1-5), where R × 100 = the buffer radius in meters.

Variable Name	**Variable Description**
AID	Add Health Respondent ID
RMELNDCOVDFR	Date From
RMELNDCOVDTO	Date To
RMELNDCOVYR	National Land Cover Data (NLCD) Source Year
RMELNDCOV001-5*	Open Water [NLCD Class 11] (sq m); Radius R00 m
RMELNDCOV006-10*	Perennial Ice/Snow [NLCD Class 12] (sq m); Radius R00 m
RMELNDCOV011-15*	1992: Low Intensity Residential [NLCD Class 21]. Post-1992: Developed, Open Space [NLCD Class 21] (sq m); Radius R00 m
RMELNDCOV016-20*	High Intensity Residential [NLCD Class 22]. Post-1992: Developed, Low Intensity [NLCD Class 22] (sq m); Radius R00 m
RMELNDCOV021-25*	1992: Commercial/Industrial/Transportation [NLCD Class 23]. Post-1992: Developed, Medium Intensity [NLCD Class 23] (sq m); Radius R00 m
RMELNDCOV026-30*	1992: N/A. Post-1992: Developed, High Intensity [NLCD Class 24] (sq m); Radius R00 m

RMELNDCOV031-35*	1992: Bare Rock/Sand/Clay [NLCD Class 31)]. Post-1992: Barren Land (Rock/Sand/Clay) [NLCD Class 31] (sq m); Radius R00 m
RMELNDCOV036-40*	1992: Quarries/Strip Mines/Gravel Pits [NLCD Class 32]. Post-1992: N/A (sq m); Radius R00 m
RMELNDCOV041-45*	1992: Transitional [NLCD Class 33]. Post-1992: N/A (sq m); Radius R00 m
RMELNDCOV046-50*	Deciduous Forest [NLCD Class 41] (sq m); Radius R00 m
RMELNDCOV051-55*	Evergreen Forest [NLCD Class 42] (sq m); Radius R00 m
RMELNDCOV056-60*	Mixed Forest [NLCD Class 43] (sq m); Radius R00 m
RMELNDCOV061-65*	1992: Shrubland [NLCD Class 51]. Post-1992: Dwarf Scrub (Alaska Only) [NLCD Class 51] (sq m); Radius R00 m
RMELNDCOV066-70*	1992: N/A. Post-1992: Scrub/Shrub [NLCD Class 52] (sq m); Radius R00 m
RMELNDCOV071-75*	1992: Orchards/Vineyards/Other [NLCD Class 61]. Post-1992: N/A (sq m); Radius R00 m
RMELNDCOV076-80*	Grassland/Herbaceous [NLCD Class 71] (sq m); Radius R00 m
RMELNDCOV081-85*	1992: N/A. Post-1992: Sedge Herbaceous (Alaska Only) [NLCD Class 72] (sq m); Radius R00 m
RMELNDCOV086-90*	1992: N/A. Post-1992: Lichens (Alaska Only) [NLCD Class 73] (sq m); Radius R00 m
RMELNDCOV091-95*	1992: N/A. Post-1992: Moss (Alaska Only) [NLCD Class 74] (sq m); Radius R00 m
RMELNDCOV096-100*	Pasture/Hay [NLCD Class 81] (sq m); Radius R00 m
RMELNDCOV101-105*	1992: Row Crops [NLCD Class 82]. Post-1992: Cultivated Crops [NLCD Class 82] (sq m); Radius R00 m
RMELNDCOV106-110*	1992: Small Grains [NLCD Class 83]. Post-1992: N/A (sq m); Radius R00 m
RMELNDCOV111-115*	1992: Fallow [NLCD Class 84]. Post-1992: N/A (sq m); Radius R00 m
RMELNDCOV116-120*	1992: Urban/Recreational Grasses [NLCD Class 85]. Post-1992: N/A (sq m); Radius R00 m
RMELNDCOV121-125*	1992: N/A. Post-1992: Woody Wetlands [NLCD Class 90] (sq m); Radius R00 m
RMELNDCOV126-130*	1992: Woody Wetlands [NLCD Class 91]. Post-1992: N/A (sq m); Radius R00 m
RMELNDCOV131-135*	1992: Emergent Herbaceous Wetlands [NLCD Class 92]. Post-1992: N/A (sq m); Radius R00 m
RMELNDCOV136-140*	1992: N/A. Post-1992: Emergent Herbaceous Wetlands [NLCD Class 95] (sq m); Radius R00 m
RMELNDCOV141-145*	Total Area (Sum) of All Land Cover Classes (sq m); Radius R00 m

7. References

1. Song Y, Gordon-Larsen P, Popkin B. A national-level analysis of neighborhood form metrics. Landsc Urban Plan. 2013 Aug 1;116:73-85. doi: 10.1016/j.landurbplan.2013.04.002. PMID: 23888091; PMCID: PMC3718082.
2. Reiskind MH, Styers DM, Hayes I, Richards SL, Doyle MS, Reed EM, Hollingsworth B, Byrd BD. Short-Term, Large-Area Survey of Container *Aedes spp.* (Diptera: Culicidae): Presence and Abundance is Associated with Fine-scale Landscape Factors in North Carolina, USA. Environ Health Insights. 2020 Sep 21;14:1178630220952806. doi: 10.1177/1178630220952806. PMID: 33013159; PMCID: PMC7513404.
3. Messier KP, Jackson LE, White JL, Hilborn ED. Landscape risk factors for Lyme disease in the eastern broadleaf forest province of the Hudson River valley and the effect of explanatory data classification resolution. Spat Spatiotemporal Epidemiol. 2015 Jan;12:9-17. doi: 10.1016/j.sste.2014.10.002. Epub 2014 Oct 22. PMID: 25779905.
4. Tsai WL, McHale MR, Jennings V, Marquet O, Hipp JA, Leung YF, Floyd MF. Relationships between Characteristics of Urban Green Land Cover and Mental Health in U.S. Metropolitan Areas. Int J Environ Res Public Health. 2018 Feb 14;15(2):340. doi: 10.3390/ijerph15020340. PMID: 29443932; PMCID: PMC5858409.
5. Multi-Resolution Land Characteristics (MRLC) Consortium. https://www.mrlc.gov/
6. Buckley, Aileen. About geographic transformations and how to choose the right one. ArcGIS Blog - Esri, Inc. Redlands, CA. 2009 May 6. https://www.esri.com/arcgis-blog/products/product/mapping/about-geographic-transformations-and-how-to-choose-the-right-one/.
7. Anderson, James R., et al., 1976. A land use and land cover classification system for use with remote sensor data, Professional Paper 964, A revision of the land use classification system in Circular 671, https://doi.org/10.3133/pp964.
8. Multi-Resolution Land Characteristics (MRLC) Consortium. "National Land Cover Database Class Legend and Description." https://www.mrlc.gov/sites/default/files/NLCDclasses.pdf
9. National Oceanic and Atmospheric Administration (NOAA) "Coastal Change Analysis Program (C-CAP) Regional Land Cover- Frequent Questions." https://coast.noaa.gov/data/digitalcoast/pdf/ccap-faq-regional.pdf.

Appendix I: National Land Cover Database Class Legends

Table 5: NLCD 1992 Legend [5]

Class\ Value	Classification Description
Water	*areas of open water or permanent ice/snow cover.*
11	**Open Water** - areas of open water, generally with less than 25% cover of vegetation/land cover.
12	**Perennial Ice/Snow** - areas characterized by year-long surface cover of ice and/or snow.
Developed	*areas characterized by a high percentage (30 % or greater) of constructed materials (e.g. asphalt, concrete, buildings, etc.).*
21	**Low Intensity Residential** - areas with a mixture of constructed materials and vegetation. Constructed materials account for 30% to 80% of the cover. Vegetation may account for 20% to 70 % of the cover. These areas most commonly include single-family housing units. Population densities will be lower than in high intensity residential areas.
22	**High Intensity Residential** - areas highly developed where people reside in high numbers. Examples include apartment complexes and row houses. Vegetation accounts for less than 20% of the cover. Constructed materials account for 80% to100% of the cover.
23	**Commercial/Industrial/Transportation** – areas of infrastructure (e.g. roads, railroads, etc.) and all highly developed areas not classified as High Intensity Residential
Barren	*areas characterized by bare rock, gravel, sand, silt, clay, or other earthen material, with little or no "green" vegetation present regardless of its inherent ability to support life. Vegetation, if present, is more widely spaced and scrubby than that in the green vegetated categories; lichen cover may be extensive.*
31	**Bare Rock/Sand/Clay** - perennially barren areas of bedrock, desert pavement, scarps, talus, slides, volcanic material, glacial debris, beaches, and other accumulations of earthen material.
32	**Quarries/Strip Mines/Gravel Pits** - areas of extractive mining activities with significant surface expression.
33	**Transitional** - areas of sparse vegetative cover (less than 25% of cover) that are dynamically changing from one land cover to another, often because of land use activities. Examples include forest clear cuts, a transition phase between forest and agricultural land, the temporary clearing of vegetation, and changes due to natural causes (e.g. fire, flood, etc.).
Forest	*areas characterized by tree cover (natural or semi-natural woody vegetation, generally greater than 6 meters tall); tree canopy accounts for 25% to 100% of the cover.*
41	**Deciduous Forest** - areas dominated by trees where 75% or more of the tree species shed foliage simultaneously in response to seasonal change.
42	**Evergreen Forest** - areas dominated by trees where 75% or more of the tree species maintain their leaves all year. Canopy is never without green foliage.

Class\ Value	Classification Description
43	**Mixed Forest** - areas dominated by trees where neither deciduous nor evergreen species represent more than 75% of the cover present.
Shrubland	*areas characterized by natural or semi-natural woody vegetation with aerial stems, generally less than 6 meters tall, with individuals or clumps not touching to interlocking. Both evergreen and deciduous species of true shrubs, young trees, and trees or shrubs that are small or stunted because of environmental conditions are included.*
51	**Shrubland** - areas dominated by shrubs; shrub canopy accounts for 25 to 100% of the cover. Shrub cover is generally greater than 25% when tree cover is less than 25%. Shrub cover may be less than 25% in cases when the cover of other life forms (e.g. herbaceous or tree) is less than 25% and shrubs cover exceeds the cover of the other life forms.
Non-natural woody	*areas dominated by non-natural woody vegetation; non-natural woody vegetative canopy accounts for 25% to 100% of the cover. The non-natural woody classification is subject to the availability of sufficient ancillary data to differentiate non-natural woody vegetation from natural woody vegetation.*
61	**Orchards/Vineyards/Other** - orchards, vineyards, and other areas planted or maintained for the production of fruits, nuts, berries, or ornamentals.
Herbaceous Upland	*upland areas characterized by natural or semi-natural herbaceous vegetation; herbaceous vegetation accounts for 75% to 100% of the cover.*
71	**Grasslands/Herbaceous** - areas dominated by upland grasses and forbs. In rare cases, herbaceous cover is less than 25%, but exceeds the combined cover of the woody species present. These areas are not subject to intensive management, but they are often utilized for grazing.
Planted/Cultivated	*areas characterized by herbaceous vegetation that has been planted or is intensively managed for the production of food, feed, or fiber; or is maintained in developed settings for specific purposes. Herbaceous vegetation accounts for 75% to 100% of the cover.*
81	**Pasture/Hay** - areas of grasses, legumes, or grass-legume mixtures planted for livestock grazing or the production of seed or hay crops.
82	**Row Crops** - areas used for the production of crops, such as corn, soybeans, vegetables, tobacco, and cotton.
83	**Small Grains** - areas used for the production of graminoid crops such as wheat, barley, oats, and rice.
84	**Fallow** - areas used for the production of crops that do not exhibit visible vegetation as a result of being tilled in a management practice that incorporates prescribed alternation between cropping and tillage.
85	**Urban/Recreational Grasses** - vegetation (primarily grasses) planted in developed settings for recreation, erosion control, or aesthetic purposes. Examples include parks, lawns, golf courses, airport grasses, and industrial site grasses.
Wetlands	*areas where the soil or substrate is periodically saturated with or covered with water as defined by Cowardin et al., (1979).*
91	**Woody Wetlands** - areas where forest or shrubland vegetation accounts for 25% to 100 % of the cover and the soil or substrate is periodically saturated with or covered with water.

Class\ Value	Classification Description
92	**Emergent Herbaceous Wetlands** - areas where perennial herbaceous vegetation accounts for 75% to 100% of the cover and the soil or substrate is periodically saturated with or covered with water.

Table 6: NLCD 2001-2019 Legend [8]

Class\ Value	Classification 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.
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	

Class\ Value	Classification Description
51	**Dwarf Scrub**** - Alaska only areas dominated by shrubs less than 20 centimeters tall with shrub canopy typically greater than 20% of total vegetation. This type is often co-associated with grasses, sedges, herbs, and non-vascular vegetation.
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 gramanoid 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.
72	**Sedge/Herbaceous**** - Alaska only areas dominated by sedges and forbs, generally greater than 80% of total vegetation. This type can occur with significant other grasses or other grass like plants, and includes sedge tundra, and sedge tussock tundra.
73	**Lichens**** - Alaska only areas dominated by fruticose or foliose lichens generally greater than 80% of total vegetation.
74	**Moss**** - Alaska only areas dominated by mosses, generally greater than 80% of total vegetation.
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 for the production of annual crops, such as corn, soybeans, vegetables, tobacco, and cotton, and also 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.
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.
**** Alaska only.**	

Appendix II: National Land Cover Database Class Crosswalk

Table 7: Crosswalk of 1992 to 2001-2019 NLCD Land Cover Classes

Aggregate Classes of Interest	NLCD 1992	NLCD 2001, 2004, 2006, 2008, 2011, 2013, 2016, 2019 C-CAP 2005 and 2010*
Trees	41 – Deciduous Forest 42 – Evergreen Forest 43 – Mixed Forest	41 – Deciduous Forest 42 – Evergreen Forest 43 – Mixed Forest
Vegetation	51 – Shrubland 61 – Orchards/ Vineyards/ Other 71 – Grassland/ Herbaceous 81 – Pasture/ Hay 82 – Row Crops 83 – Small Grains 84 – Fallow	51 – Dwarf Scrub† 52 – Scrub/ Shrub 71 – Grassland/ Herbaceous 72 – Sedge Herbaceous† 73 – Lichens† 74 – Moss† 81 – Pasture/Hay 82 – Cultivated Crops
Wetlands	91 – Woody Wetlands 92 – Emergent Herbaceous Wetlands	90 – Woody Wetlands 95 – Emergent Herbaceous Wetlands
Developed	21 – Low Intensity Residential 22 – High Intensity Residential 23 – Commercial/ Industrial/ Transportation 85 – Urban/ Recreational Grasses	22 – Developed, Low Intensity 23 – Developed, Medium Intensity 24 – Developed, High Intensity 21 – Developed, Open Space
Water	11 – Open Water	11 – Open Water
Other (Snow and Rock)	12 – Perennial Ice/ Snow 31 – Bare Rock/ Sand/ Clay 32 – Quarries/ Strip Mines/ Gravel Pits 33 – Transitional	12 – Perennial Ice/Snow 31 – Barren Land (Rock/ Sand/ Clay)
*** After conversion to NLCD classes (see Table 8). † Alaska only.**		

Table 8: Crosswalk of C-CAP to 2001-2019 NLCD Land Cover Classes [9]

Anderson Level 1 Category	NLCD Category	C-CAP Category
Urban or Built-up Land (1)	Developed, High Intensity (24) Developed, Medium Intensity (23) Developed, Low Intensity (22) Developed, Open Space (21)	High Intensity Developed (2) Medium Intensity Developed (3) Low Intensity Developed (4) Open Space Developed (5)
Agricultural Land (2)	Cultivated Crops (82) Pasture/Hay (81)	Cultivated Land (6) Pasture/Hay (7)
Rangeland (3)	Grassland / Herbaceous (71) Scrub / Shrub (52)	Grassland (8) Scrub Shrub (12)
Forest (4)	Deciduous Forest (41) Evergreen Forest (42) Mixed Forest (43)	Deciduous Forest (9) Evergreen Forest (10) Mixed Forest (11)
Wetlands (6)	Woody Wetlands (90) Emerging Herbaceous Wetlands (95)	Palustrine Forested Wetlands (13) Palustrine Scrub Shrub Wetlands (14) Estuarine Forested Wetlands (15) Estuarine Scrub Shrub Wetlands (16) Palustrine Emergent Wetlands (17) Estuarine Emergent Wetlands (18)
Open Water (5)	Open Water (11)	Open Water (21) Palustrine Aquatic Bed (22) Estuarine Aquatic Bed (23)
Barren Land (7)	Barren Land (31)	Unconsolidated Shore (19) Barren Land (20)
Tundra (8)		Tundra (24)
Perennial Ice/Snow (9)	Perennial Ice/Snow (12)	Perennial Ice/Snow (25)