NEON USER GUIDE TO COARSE DOWNED WOOD LOG SURVEY (DP1.10010.001)

PREPARED BY	ORGANIZATION
Courtney Meier	SCI
Samuel Simkin	SCI
Katherine Jones	SCI

CHANGE RECORD

REVISION	DATE	DESCRIPTION OF CHANGE
A	12/15/2017	Initial Release
B	07/07/2020	Included general statement about usage of neonUtilities R package and statement about possible location changes. Updated taxonomy information. Included 'Sampling Design Changes' section. Updated 'Spatial Resolution and Extent' section. Made other clarifications within document.
C	04/11/2022	Added language in section 4 Taxonomy addressing RTE species obfuscation in the data. Updated section 5.3 Data Revision with latest information regarding data release. Updated information regarding the geoNEON package
C.1	12/08/2023	Added identification history section 4.2 to Taxonomy and updated Data Relationships section (3.7) and Analytical Facility Data Quality section (5.5) to reference identification history.
D	04/09/2024	Minor formatting updates
E	03/18/2025	Added calculations and Appendix sections that explain how to calculate CDW volume and mass. Added information about the new neonutilities Python package.
F	03/09/2026	Added new Figure 1 with generalized TOS site schematic, and added more detail to the 'Spatial Resolution' and 'Spatial Sampling Design' section to better illustrate how to calculate the location of tallied logs and to include links to information about plots that have moved. Updated the 'Design Change' section to document 2026 scope modifications, added infomation about automated data review to the 'Automated Data Processing Steps' section and made edits throughout the document to improve accuracy and clarity.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

1 DESCRIPTION

1.1 Purpose

This document provides an overview of the data included in this NEON Level 1 data product, the quality controlled product generated from raw Level 0 data, and associated metadata. In the NEON data products framework, the raw data collected in the field, for example the dimensions of logs, are considered the lowest level (Level 0). Raw data that have been quality checked via the steps detailed herein, as well as simple metrics that emerge from the raw data are considered Level 1 data products.

The text herein provides a discussion of measurement theory and implementation, data product provenance, quality assurance and control methods used, and approximations and/or assumptions made during L1 data creation.

1.2 Scope

This document describes the steps needed to generate the L1 data product, Coarse Downed Wood Log Survey - the tally and raw measurement of coarse downed wood particles ≥ 2 cm diameter and ≥ 1 m length - and associated metadata from input data. This document also provides details relevant to the publication of the data products via the NEON data portal, with additional detail available in the file, NEON Data Variables for Coarse Downed Wood Log Survey (DP1.10010.001) (AD[05]), provided in the download package for this data product.

This document describes the process for ingesting and performing automated quality assurance and control procedures on the data collected in the field pertaining to TOS Protocol and Procedure: Coarse Downed Wood (AD[07]). The raw data that are processed in this document are detailed in the file, NEON Raw Data Validation for Coarse Downed Wood Log Survey (DP0.10010.001) (AD[04]), provided in the download package for this data product. Please note that raw data products (denoted by 'DP0') may not always have the same numbers (e.g., '10010') as the corresponding L1 data product.

2 RELATED DOCUMENTS AND ACRONYMS

2.1 Associated Documents

AD[01]	NEON.DOC.000001	NEON Observatory Design (NOD) Requirements
AD[02]	NEON.DOC.000913	TOS Science Design for Spatial Sampling
AD[03]	NEON.DOC.002652	NEON Data Products Catalog
AD[04]	NEON.DP0.10010.001 _dataValidation.csv	NEON Raw Data Validation for Coarse Downed Wood Log Survey (DP0.10010.001)
AD[05]	NEON.DP1.10010.001 _variables.csv	NEON Data Variables for Coarse Downed Wood Log Survey (DP1.10010.001)
AD[06]	NEON.DOC.000914	TOS Science Design for Plant Biomass, Productivity, and Leaf Area Index
AD[07]	NEON.DOC.001711	TOS Protocol and Procedure: Coarse Downed Wood
AD[08]	NEON.DOC.000913	TOS Science Design for Spatial Sampling
AD[09]	NEON.DOC.000913	TOS Science Design for Plant Diversity
AD[10]	NEON.DOC.000008	NEON Acronym List
AD[11]	NEON.DOC.000243	NEON Glossary of Terms
AD[12]	NEON.DOC.000987	TOS Protocol and Procedure: Measurement of Vegetation Structure
AD[13]	NEON.DOC.001710	TOS Protocol and Procedure: Measurement of Litterfall and Fine Woody Debris
AD[14]	NEON.DOC.004825	NEON Algorithm Theoretical Basis Document: OS Generic Transitions
AD[15]	Available on NEON data portal	NEON's Ingest Conversion Language (NICL) specifications
AD[16]	NEON.DOC.005424	NEON Algorithm Theoretical Basis Document: OS Data Quality Control

2.2 Acronyms

Acronym	**Definition**
CDW	Coarse Downed Wood
DST	Decay class x size category x taxonID combination
LIDS	Line Intercept Distance Sampling

3 DATA PRODUCT DESCRIPTION

The Coarse Downed Wood Log Survey data product provides volume estimates of downed logs ≥ 2cm diameter across NEON sites. Tally surveys are completed according to the Line Intercept Distance Sampling method (LIDS; Affleck 2008). The LIDS sampling method tallies logs with probability proportional to volume and restricts the search for logs to a transect or group of radial transects at each sampling point (Affleck 2008, 2010). By using transects, detection errors in brushy or complex terrain are minimized compared to other techniques that require searching for logs over large areas (Jordan et al. 2004). The length of the transect is not fixed with the LIDS method; instead, the length of the transect increases for large-volume logs, ensuring that a representative sample of large logs is tallied across multiple field sites (Affleck 2008).

Qualifying CDW logs are divided into three different diameter size categories, as defined by Keller et al. (2004): 2-5 cm, 5-10 cm, and ≥ 10 cm diameter. To qualify for inclusion, logs or portions of logs within a given size category must also be ≥ 1 m in length (Harmon and Sexton 1996). Dead trees (i.e., snags) that have not yet fallen to an angle > 45° from vertical are accounted for via the Vegetation Structure protocol (AD[12]), and logs that are suspended in the air > 2 m above the ground at the transect intersection point are ignored. Woody particles with diameter < 2 cm at the transect intersection point are considered fine woody debris and are sampled according to the Litterfall and Fine Woody Debris protocol (AD[13]).

All logs are classified into unique 'decay class x size category x taxonID' (DST) combinations. Size category is based on diameter but is supplemented by additional physical characteristics such as log length that allow for calculation of frequency for each DST. Decay class is supplemented with data on the presence/absence of leaves and branches, % bark cover, etc. The distance and azimuth angle from the transect origin is also recorded for mapping and QC purposes.

3.1 Spatial Sampling Design

Within each NEON terrestrial site, Coarse downed wood survey sampling is organized at two different sampling scales: 1) Within Distributed plots established at the site scale; and 2) Within Tower plots established at the scale of the 'airshed' – i.e., the comparatively smaller land-surface area within the site that is the source of the flux data collected by the NEON TIS tower (Figure 1). At the site scale, the TOS allocates Distributed plots according to a spatially-balanced, stratified-random sampling design that is applied consistently across all NEON TOS sites (Barnett *et al.* 2019, Theobald *et al.* 2007). According to this design, plots are allocated across the landscape in proportion to the area of dominant National Land Cover Database (NLCD) cover types (Fry *et al.* 2011). Within the tower airshed, a spatially-balanced randomized sampling design is employed (Barnett *et al.* 2019).

Coarse downed wood surveys are completed at all terrestrial NEON sites that contain a sufficient number of qualifying logs. At sites where qualifying CDW logs are present, sampling occurs at each Tower plot (20-30 plots), and a subset of spatially balanced Distributed plots (20 plots maximum), for a maximum of 50 plots at any given site. At each plot, three LIDS transects radiate outward from the plot centroid, with 120° separating each transect (Figure 2). The azimuthal orientation of each group of transects is chosen randomly for each plot to minimize effects of topography, directional blowdown, logging management, etc., on selection of CDW logs across all sampling locations. See TOS Science Design for Plant Biomass, Productivity, and Leaf Area Index (AD[06]), TOS Protocol and Procedure: Coarse Downed Wood (AD[07]), TOS Science Design for Spatial Sampling (AD[08]) and TOS Science Design for Plant Diversity (AD[09]) for further details on sampling design, specific protocol steps, plot allocation, and species identification, respectively.

Sampling typically occurs in the same locations over the lifetime of the Observatory. However, sampling locations may become impractical to sample, due to disturbance or other local changes. When this occurs, the location and its location ID are retired or shifted to slightly different coordinates. Refer to the TOS plot location changes spreadsheet found in the "Terrestrial Observation System Sampling Locations" download on the spatial-data-maps page at neonscience.org (https://www.neonscience.org/data-samples/data/spatial-data-maps) for details about locations that have been retired or added since the operations phase started in 2019. The same download also includes the "versionedPoints" and "versionedSubplots" files, which document shifts in coordinates.

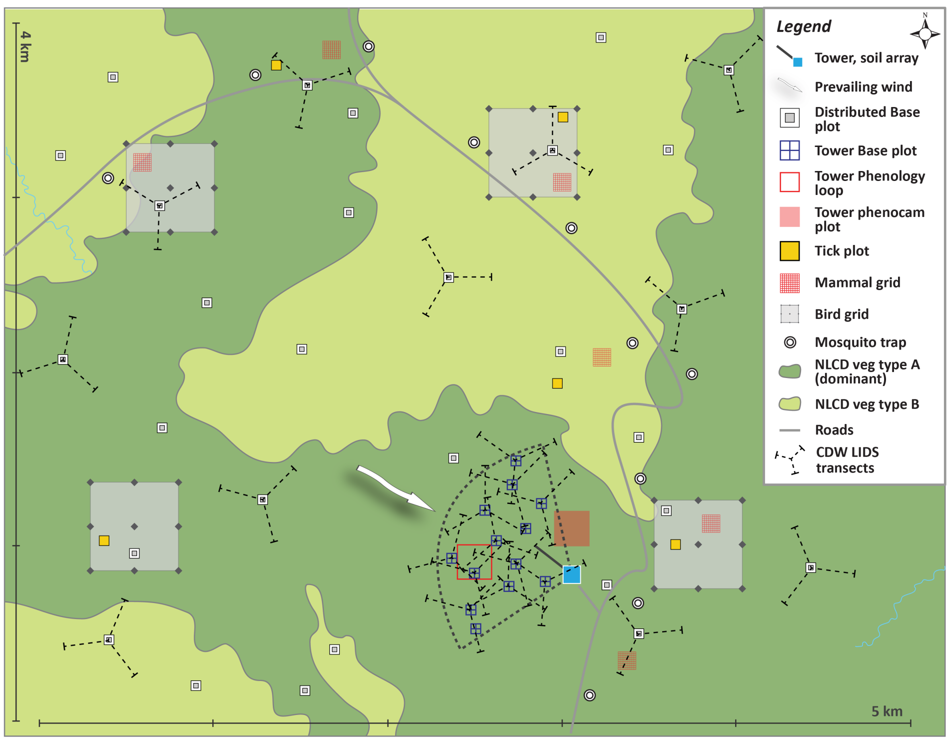

Figure 1: Generalized TOS sampling schematic, showing the placement of Distributed and Tower plots and superimposed CDW LIDS transects.

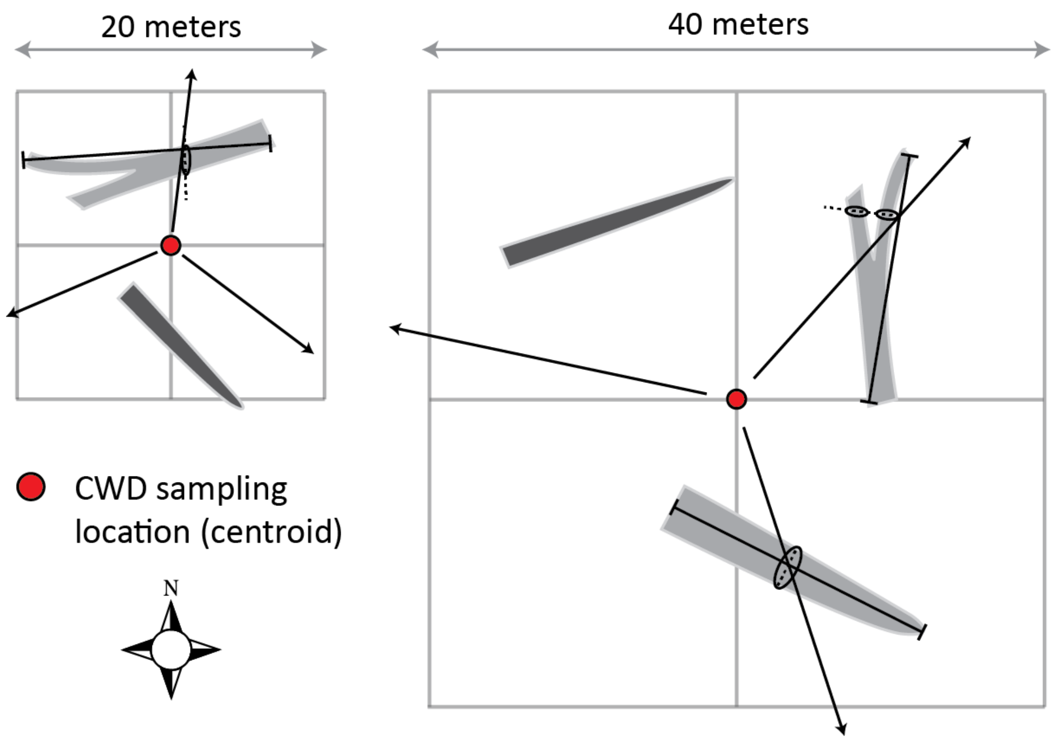

Figure 2: Illustration of three randomly oriented LIDS transects (arrow vectors) superimposed over a 20m x 20m plot (left), and a 40m x 40m plot (right). There is a 3 meter gap between the sampling reference point (i.e., the plot centroid) and the start of each transect, which minimizes the chance that CDW logs close to the plot centroid get double- or triple-counted. Gray shapes represent CDW logs, dashed lines show the location and orientation of cross-sectional area measurement(s) (ovals) for those logs that intersect a LIDS transect, and capped black lines imply the logs' central axes that are measured for length. Note that CDW logs selected for tally and/or bulk density sampling may lie outside of the plot boundary.

3.2 Temporal Sampling Design

Coarse downed wood tally bouts are completed in both Distributed plots and Tower plots on a 5-year sampling interval. At all sites, Distributed plot sampling and Tower plot sampling are scheduled in a staggered manner such that sites generate data either from all Distributed plots or all Tower plots every 2-3 years. Sampling bouts should take no longer than 3 months to complete. New coarse downed wood logs are typically produced during periods of high wind or rain events associated with seasonal storm activity. Sampling bouts at any given site are scheduled to occur after the expected storm season. For example, in the Southeast (D03) and Atlantic Neotropical domains (D04), CDW tally sampling occurs during either the winter or spring and outside of the summer/fall hurricane season. In contrast, in the Pacific Northwest (D16), most storm activity occurs primarily in the winter, and CDW tally sampling occurs in the summer/autumn seasons.

3.3 Sampling Design Changes

The NEON TOS has implemented sampling design changes affecting the 'Coarse downed wood log survey' data product over the course of data collection (Table 1). Such changes arise due to continual evaluation of the sampling design for best practices, efficiency, and Observatory health, both internally and in collaboration with external technical working groups. Changes to protocols and data products also occur when NEON improves data product usability and structure, and when optimization of sampling designs is necessary to ensure that allocation of sampling effort is poised to maximize returns to the scientific community.

Table 1: Summary of significant sampling design and data product changes for the NEON TOS 'Coarse downed wood log survey' data product.

Change Date	Affected Sites	Change Summary Description
2018-01-01	All sites	The 'Coarse downed wood log survey' sampling frequency was reduced from once every three years to once every five years as a cost-savings measure.
2026-02-09	All sites	Data collection for 'Coarse downed wood log survey' was suspended in 2026, leading to a 6-year interval between sampling events rather than the expected 5-year interval. Sampling originally scheduled for 2026 will be carried out in 2027, and the schedule originally intended for 2027 will be implemented in 2028 and so forth, leading to a consistent 6-year interval between sampling events across all sites in the schedule until 2032.

3.4 Variables Reported

All variables reported from the field or laboratory (L0 data) are listed in the file, NEON Raw Data Validation for Coarse Downed Wood Log Survey (DP0.10010.001) (AD[04]). All variables reported in the published data (L1 data) are also provided separately in the file, NEON Data Variables for Coarse Downed Wood Log Survey (DP1.10010.001) (AD[05]).

Field names have been standardized with Darwin Core terms (http://rs.tdwg.org/dwc/; accessed 16 February 2014), the Global Biodiversity Information Facility vocabularies (http://rs.gbif.org/vocabulary/gbif/; accessed 16 February 2014), and the VegCore data dictionary (https://projects.nceas.ucsb.edu/nceas/projects/bien/wiki/VegCore; accessed 16 February 2014), where applicable. NEON TOS spatial data employs the World Geodetic System 1984 (WGS84) for its fundamental reference datum and Geoid12A geoid model for its vertical reference surface. Latitudes and longitudes are denoted in decimal notation to six decimal places, with longitudes indicated as negative west of the Greenwich meridian.

Some variables described in this document may be for NEON internal use only and will not appear in downloaded data.

3.5 Spatial Resolution and Extent

The finest resolution at which spatial data are reported is the point location of the log along a transect.

The spatial hierarchy from finest to coarsest resolution is:

point (see below to calculate) → **plotID** → **siteID** → **domainID**

The basic spatial data included in the data downloaded include the latitude, longitude, and elevation of the centroid of the plot where sampling occurred + associated uncertainty due to GPS error and plot width. Shapefiles of all NEON Terrestrial Observation System sampling locations can be found on the NEON science webpage at https://www.neonscience.org/data-samples/data/spatial-data-maps.

To derive a more precise estimate of the location of each log reported in the 'cdw_fieldtally' table, download the `R` `geoNEON` package (https://github.com/NEONScience/NEON-geolocation) and use the `geoNEON::getLocTOS()` function with the following arguments:

- `data = df`, where 'df' is a user-supplied data frame containing NEON named location data, typically the cdw_fieldtally table as downloaded via the NEON API or the NEON Data Portal.
- `dataProd = "cdw_fieldtally"`
- `token = "NEON_TOKEN"`, where "NEON_TOKEN" is an optional token generated via a NEON user account. While tokens are not required, they provide enhanced download speeds.

Alternatively, use the steps below to perform the same calculation:

1. Extract 'namedLocation' values for the plot centroid from the 'namedLocation' field in the 'cdw_fieldtally' table. The 'namedLocation' is associated with the coordinates of the plot centroid in the NEON named location database.
   a. *Example namedLocation*: HARV_052.basePlot.cdw
2. Construct a URL for the namedLocation by concatenating the NEON API (http://data.neonscience.org/api) with the 'namedLocation' value, and use this URL to retrieve easting ('locationUtmEasting'), northing ('locationUtmNorthing'), coordinate uncertainty ('Value for Coordinate uncertainty'), and utm zone ('locationUtmZone') data as inputs to the next step. If the location data include more than one record, use the coordinates corresponding to the sampling date.
   a. *Example URL*: https://data.neonscience.org/api/v0/locations/HARV_052.basePlot.cdw?history=TRUE
3. Calculate the geolocation of each log according to the equation:

$$Easting = locationUtmEasting + d * \sin\theta \tag{1}$$

and

$$Northing = locationUtmNorthing + d * \cos\theta \tag{2}$$

where,

$$\theta = \frac{lidsAzimuth * \pi}{180} \quad (3)$$

$$locationUtmEasting = \quad (4)$$

the easting value associated with the plotID (plot centroid)

$$locationUtmNorthing = \quad (5)$$

the northing value associated with the plotID (plot centroid)

$$d = logDistance + 3 \quad (6)$$

Note: The ''logDistance'' is the distance along transect where the log was tallied, and a 3 m offset is applied because transects start 3 m from the plot centroid.

4. Increase 'coordinateUncertainty' by an appropriate amount to account for error introduced by navigating along the transect. Additional error may be introduced due to tape stretching and drift from a compass heading.

3.6 Temporal Resolution and Extent

The finest resolution at which temporal data are reported for a given log tally record is the range between **startDate** and **endDate** associated with when the transect was tallied.

date (date range a transect was tallied) → **eventID** (bout-level identifier)

3.7 Associated Data Streams

The Coarse Downed Wood Log Survey data product (DP1.10010.001) is tightly integrated with the Coarse Downed Wood Bulk Density Sampling data product (DP1.10014.001). Together, these two data products allow calculation of CDW dry mass per sizeCategory per decayClass per taxonID. Although these two data products are related, logs that are tagged during Log Survey are specifically avoided for Bulk Density Sampling.

The 'Woody vegetation structure' data product (DP1.10098.001) is also linked to the CDW Log Survey product due to the fact that individuals tagged and measured while living may appear in the CDW Log Survey data product once they have fallen to the ground. In the event that data in CDW Log Survey are collected from an individual that was previously part of one of these two datasets, the **vstTagID** field in the CDW dataset may be matched to the **tagID** field in the Woody and Non-Woody Perennial Vegetation Structure products.

3.8 Product Instances

At a given site, a maximum of 50 plots is selected for coarse downed wood survey once every five years. Each plot has 3 transects, with the total number of logs tallied along all 3 transects for a given plot ranging between 0-50. Thus, each site should generate no more than approximately 2,500 data product instances over a five year period.

3.9 Data Relationships

The protocol dictates that each transect is surveyed only once in a sampling year. If no logs ≥ 2cm diameter and ≥ 1m length are tallied, only a single record will be generated for the transect, and these records will have **targetTaxaPresent**= 'No', indicating that the transect was checked and no qualifying logs were discovered. For transects where **targetTaxaPresent**= 'Yes', each qualifying log along the transect results in a distinct record. Duplicates and/or missing data may exist where protocol and/or data entry aberrations have occurred; users should check data carefully for anomalies before joining the field tally table with other data products.

cdw_fieldtally.csv → One record expected per **plotID** + **lidsAzimuth** + **logID** combination per year.

cdw_tally_identificationHistory.csv → One or more records expected per identificationHistoryID. Records are only created when data corrections to taxonomic identifications are made. If errors in identification are detected through QAQC processes after data publication, then corrected taxonomy will be provided in the cdw_fieldtally table. The cdw_tally_identificationHistory table is populated with all prior names used for specimen(s) in the data product. When data are populated in the cdw_tally_identificationHistory table, **identificationHistoryID** is used as a linking variable between the cdw_tally_identificationHistory table and the other table where updates were made.

An **individualID** is generated for each tallied log with logMaxDiameter ≥ 10 cm, and the **individualID** is built from the **logID**. No **individualID** is generated for logs with logMaxDiameter < 10 cm. For logs < 10 cm diameter, **logID** values are incremented values that reset for each **plotID**, and are denoted with the letter "L" preceding an incremented value (e.g. "L005"), and the **individualID** field is null. Note that the CDW **individualID** cannot be matched with the **individualID** created for other NEON plant data products (described below).

Recently downed logs in the CDW Log Survey dataset that were previously measured while standing according to one or more protocols are identified in the CDW Log Survey dataset by the **vstTagID** field. To match **vstTagID** with the linked **individualID** reported in other NEON data products, an end-user must perform the following steps:

1. Flag **vstTagID** with leading zeroes until the total string length is 6 characters
2. Concatenate: "NEON.PLA.D##.SITE" + "." + "**vstTagID**", where:
   - D## = **domainID**
   - SITE = **siteID**
3. Match the resulting string (e.g., `NEON.PLA.D16.WREF.000047`) with **individualID** in any/all of the following associated data products:
   - Vegetation structure (DP1.10098.001)
   - Plant foliar traits (DP1.10026.001)

- Plant phenology observations (DP1.10055.001)

Data downloaded from the NEON Data Portal are provided in separate data files for each site and month requested. The R 'neonUtilities' package and the Python 'neonutilities' package both contain functions to merge these files across sites and months into a single file for each table. The R neonUtilities package is available from the Comprehensive R Archive Network (CRAN; https://cran.r-project.org/web/packages/neonUtilities/index.html) and can be installed using the install.packages() function in R. The Python neonutilities package is available on the Python Package Index (PyPi; https://pypi.org/project/neonutilities/) and can be installed using pip. For instructions on using the package in either language to merge NEON data files, see the Download and Explore NEON Data tutorial on the NEON website: https://www.neonscience.org/download-explore-neon-data.

3.10 Special Considerations

A subset of records for coarse downed wood particles < 10 cm diameter, measured between 2016-04-04 and 2017-06-19, were recorded without **decayClass**. These records have been flagged as a **dataQF**="decayClassNotRecorded". Affected records can be used to calulate volume of coarse downed wood, but do not contain sufficient information to calculate wood density or mass when combined with the Coarse Downed Wood Bulk Density product.

3.11 Calculations

Using the LIDS method (Affleck 2008, 2009 (p. 3, eq. 1), 2010), CDW volume is calculated at the plot scale as:

$$CDW volume(m3/ha) = F * n \tag{7}$$

where F is the site-specific volume factor (an approximate inverse to plot size), n is the number of qualifying logs tallied across all three transects originating from the centroid of the plot, and CDW volume density has the unit of m^3/hectare

Two examples:

1. HARV (volumeFactor = 5) has a plot with 11 qualifying logs. At this HARV plot the volume of coarse downed wood equals 55 m^3/hectare (volumeFactor of 5 x 11 logs).
2. ABBY (volumeFactor = 15) has a plot with 7 qualifying logs. At this ABBY plot the volume of coarse downed wood equals 105 m^3/hectare (volumeFactor of 15 x 7 logs).

CDW mass is calculated in the following steps:

1. In both the CDW log survey and CDW bulk density data products, concatenate 'decayClass', 'sizeCategory', and 'taxonID' to create a Decay x Size x Taxon (DST) identifier for all logs.
2. In the CDW bulk density data product, average 'bulkDensDisk' values by sampleID in the cdw_densitydisk table, join with the cdw_densitylog by sampleID, and then calculate the average bulkDensDisk for each DST, yielding the CDW density (mass/volume in units of g/cm^3).

3. In the CDW log survey data product, summarize the logCount for each eventID x plotID x DST combination in cdw_fieldTally and calculate the CDW volume (m3/ha) as indicated in equation 7 for each group.
4. Join the CDW density from step 2 with the CDW volume from step 3 using the concatenated DST field from step 1, yielding a single table with eventID, plotID, DST, CDW density, and CDW volume. For each eventID x plotID x DST combination, calculate CDW mass as follows:

$$CDWmass = CDWvolume(m3/ha)*CDWdensity(g/cm3)*1000000cm3/m3*1Mg/1000000g \quad (8)$$

which simplifies to

$$CDWmass = CDWvolume * CDWdensity \quad (9)$$

if final desired mass unit is Mg/ha

5. Finally, for plot-level CDW mass (m3/ha), sum all DST masses within each eventID x plotID combination. See Appendix for sample R code.

4 TAXONOMY

NEON manages taxonomic entries by maintaining a master taxonomy list based on the community standard, if one exists. Through the master taxonomy list, synonyms submitted in the data are converted to the appropriate name in use by the standard. The master taxonomy list also indicates the expected geographic distribution for each species by NEON domain and whether it is known to be introduced or native in that part of the range. Errors are generated if a species is reported at a location outside of its known range. If the record proves to be a reliable report, the master taxonomy table is updated to reflect the distribution change.

The full master taxonomy lists are available on the NEON Data Portal for browsing and download: https://data.neonscience.org/taxonomic-lists.

4.1 Plant taxonomy

The master taxonomy for plants is the USDA PLANTS Database (USDA, NRCS. 2014. https://plants.usda.gov). Taxon ID codes used to identify taxonomic concepts in the NEON master taxonomy list are alphanumeric codes, 4-6 characters in length based on the accepted scientific name. Each code is composed of the first two letters of the genus, followed by the first two letters of the species and first letter of the terminal infraspecific name (if applicable) then, if needed, a tiebreaking number to address duplicate codes. Genus and family symbols are the first five (genus) or six (family) letters of the name, plus tiebreaking number (if needed). Symbols were first used in the Soil Conservation Service's National List of Scientific Plant Names (NLSPN) and have been perpetuated in the PLANTS system. The portions of the PLANTS Database included in the NEON plant master taxonomy list includes native and naturalized plants present in NEON observatory sampling area including the Lower 48 U.S. States, Alaska, Hawaii, and Puerto Rico.

NEON plans to keep the taxonomy updated in accordance with USDA PLANTS Database starting in 2020 and annually thereafter.

Geographic ranges and native statuses used in this data product are also from the USDA PLANTS Database.

Prior to the 2022 data release, publication of species identifications were obfuscated to a higher taxonomic rank when the taxon was found to be listed as threatened, endangered, or sensitive at the state level where the observation was recorded. The state-level obfuscation routine was removed from the data publication process at all locations excluding sites located in D01 and D20, and data have been reprocessed to remove the obfuscation of state-listed taxa for all years. Federally listed threatened and endangered or sensitive species remain obfuscated at all sites and sensitive species remain redacted at National Park sites.

4.2 Identification History

Beginning in 2023, the cdw_tally_identificationHistory table was added to track any changes to taxonomic identifications that have been published in NEON data. Such taxonomic revisions may be necessary when errors are found in QAQC checks. Requests for taxonomic changes are reviewed by NEON science staff. Proposed changes are evaluated based on evidence in the form of photographs, existing samples, genetic data, consultation with taxonomic experts, or range maps. Upon approval, the existing record in cdw_fieldtally is updated with the new taxonomic information and a unique identifier is added to the identificationHistoryID field. A record with the same **identificationHistoryID** is created in the cdw_tally_identificationHistory table where the previous taxonomic information is archived along with the date the change was made.

Beginning in 2023, the ‘cdw_tally_identificationHistory’ table was added to track any changes to taxonomic identifications that have been published in this data product. Such taxonomic revisions may be necessary when errors are found in QAQC checks, or when evidence from genetic analysis of samples or re-analysis of archived samples indicate a revision is necessary. Requests for taxonomic changes are reviewed by NEON science staff. Proposed changes are evaluated based on evidence in the form of photographs, existing samples, genetic data, consultation with taxonomic experts, or range maps. Specific to the CDW log survey data product, approved taxonomic changes result in revision of the **taxonID** and **scientificName** fields in the ‘cdw_fieldtally’ table to reflect the updated taxonomy, and the **identificationHistoryID** field in this table is populated with a unique identifier. In addition, a record is added to the ‘cdw_tally_identificationHistory’ table with the same **identificationHistoryID** and the previous taxonomic information is recorded along with the date the change was made.

5 DATA QUALITY

5.1 Data Entry Constraint and Validation

Many quality control measures are implemented at the point of data entry within a mobile data entry application or web user interface (UI). For example, data formats are constrained and data values controlled through the provision of dropdown options, which reduces the number of processing steps necessary to prepare the raw data for publication. An additional set of constraints are implemented during the process of ingest into the NEON database. The product-specific data constraint and validation requirements built into data entry applications and database ingest are described in the document NEON Raw Data Validation for Coarse Downed Wood Log Survey (DP0.10010.001), provided with every download of this data product. Contained within this file is a field named 'entryValidationRulesForm', which describes syntactically the validation rules for each field built into the data entry application. Data entry constraints are described in Nicl syntax in the validation file provided with every data download, and the Nicl language is described in NEON's Ingest Conversion Language (NICL) specifications ([AD[15]). Data collected prior to 2017 were processed using a paper-based workflow that did not implement the full suite of quality control features associated with the interactive digital workflow.

5.2 Automated Data Processing Steps

Following data entry into a mobile application or web user interface, the steps used to process the data through to publication on the NEON Data Portal are detailed in the NEON Algorithm Theoretical Basis Document: OS Generic Transitions (AD[14]).

Published data are reviewed for completeness, timeliness, and validity using an internal set of tests and metrics, as detailed in the NEON Algorithm Theoretical Basis Document: OS Data Quality Control (AD[16]). These quality tests are used to guide process improvements, audits of analytical facilities, and data updates, but do not generate quality flags in published data.

5.3 Data Revision

All data are provisional until a numbered version is released. Annually, NEON releases a static version of all or almost all data products, annotated with digital object identifiers (DOIs). The first data Release was made in 2021. During the provisional period, QA/QC is an active process, as opposed to a discrete activity performed once, and records are updated on a rolling basis as a result of scheduled tests or feedback from data users. The Issue Log section of the data product landing page contains a history of major known errors and revisions.

5.4 Quality Flagging

The **dataQF** field in each data record is a quality flag for known errors applying to the record. Please see the *Special Considerations* section of this document for a list of known errors that may be present in the data, and below for an explanation of **dataQF** codes specific to this product.

Table 2: Descriptions of the dataQF codes for quality flagging

fieldName	**value**	**definition**
dataQF	legacyData	Data recorded using a paper-based workflow that did not implement the full suite of quality control features associated with the interactive digital workflow
dataQF	decayClass NotRecorded	Records collected with no decayClass; see *Special Considerations* section above for more information.

Records of land management activities, disturbances, and other incidents of ecological note that may have a potential impact are found in the Site Management and Event Reporting data product (DP1.10111.001)

6 REFERENCES

Affleck, D. L. R. 2008. A line intersect distance sampling strategy for downed wood inventory. Canadian Journal of Forest Research 38:2262-2273.

Affleck, D. L. R. 2009. Field Results for Line Intersect Distance Sampling of Coarse Woody Debris. USDA Forest Service Proceedings - RMRS-P-56. pages 1-12

Affleck, D. L. R. 2010. On the efficiency of line intersect distance sampling. Canadian Journal of Forest Research 40:1086-1094.

Barnett, D. T., P. A. Duffy, D. S. Schimel, R. E. Krauss, K. M. Irvine, F. W. Davis, J. E. Gross, E. I. Azuaje, A. S. Thorpe, D. Gudex-Cross, M. Patterson, J. M. McKay, J. T. McCorkel, C. L. Meier. 2019. The terrestrial organism and biogeochemistry spatial sampling design for the National Ecological Observatory Network. Ecosphere https://doi.org/10.1002/ecs2.2540

Fry, J., G. Xian, S. Jin, J. Dewitz, C. Homer, L. Yang, C. Barnes, N. Herold, and J. Wickham. 2011. Completion of the 2006 National Land Cover Database for the Conterminous United States. Photogrammetric Engineering and Remote Sensing 77:858-864.

Harmon, M. E., and J. Sexton. 1996. Guidelines for measurements of woody detritus in forest ecosystems. Publication No. 20 Long-term Ecological Research Network Office, University of Washington:1-42

Jordan, G. J., M. J. Ducey, and J. H. Gove. 2004. Comparing line-intersect, fixed-area, and point relascope sampling for dead and downed coarse woody material in a managed northern hardwood forest. Canadian Journal of Forest Research 34:1766-1775.

Keller, M., M. Palace, G. P. Asner, R. Pereira, and J. N. M. Silva. 2004. Coarse woody debris in undisturbed and logged forests in the eastern Brazilian Amazon. Global Change Biology 10:784-795.

Meier, C. L., K. M. Thibault, and D. T. Barnett. 2023. Spatial and temporal sampling strategy connecting NEON Terrestrial Observation System protocols. Ecosphere 14(3):e4455. https://doi.org/10.1002/ecs2.4455

Theobald, D. M., D. L. Stevens, Jr., D. White, N. S. Urquhart, A. R. Olsen, and J. B. Norman. 2007. Using GIS to generate spatially balanced random survey designs for natural resource applications. Environmental Management 40:134-146.

USDA, NRCS. 2014. The PLANTS Database (http://plants.usda.gov, 25 August 2014). National Plant Data Team, Greensboro, NC 27401-4901 USA.

7 APPENDIX

Sample R code to calculate plot-level CDW volume and mass:

```
library(neonUtilities)
library(dplyr)

#   Download cdw_fieldtally data from the 'Coarse downed wood log survey' data product.
cdw_log <- neonUtilities::loadByProduct(
  dpID = "DP1.10010.001",
  check.size = FALSE,
  package = "basic",
  #   Remove the following line if you don't have a token stored in the .Renviron file.
  token = Sys.getenv("NEON_TOKEN"))

#   Turn all tables in the list to dataframe (DF) in the global environment,
#   where name of table = name of DF.
list2env(cdw_log, envir = .GlobalEnv)

#   Download cdw_densitylog and cdw_densitydisk data from the
#   'Coarse downed wood bulk density' data product.
cdw_bulk <- neonUtilities::loadByProduct(
  dpID = "DP1.10014.001",
  check.size = FALSE,
  package = "basic",
  #   Remove the following line if you don't have a token stored in the .Renviron file.
  token = Sys.getenv("NEON_TOKEN"))

#   Turn all tables in the list to dataframe (DF) in the global environment,
#   where name of table = name of DF.
list2env(cdw_bulk, envir=.GlobalEnv)

#   If applicable, remove duplicates and non-qualifying logs as in Meier et al. 2023
#   (Meier_etal_Appendiz_S2.Rmd lines 295-407).

#   Proceed with the following R code:

#   If 2 disks were collected per log then average
cdw_densitydisk <- cdw_densitydisk %>%
  group_by(sampleID) %>%
  summarise(bulkDensDisk = mean(bulkDensDisk, na.rm = T),
            .groups = "drop")

#   Join density values with other log attributes
density <- dplyr::full_join(cdw_densitylog,
                            cdw_densitydisk,
                            by = "sampleID")
```

```
#   Create Decay x Size x Taxon (DST) identifier for the bulk density data
density$DST <- paste(density$decayClass,
                     density$scientificName,
                     density$sizeCategory,
                     sep = "_")

#   Calculate mean bulk density for each DST
densityDST <-  density %>%
  group_by(DST) %>%
  summarise(bulkDensDisk = mean(bulkDensDisk, na.rm = T),
            .groups = "drop")

#   Create Decay x Size x Taxon (DST) identifier for the survey data
cdw_fieldtally$DST <- paste(cdw_fieldtally$decayClass,
                            cdw_fieldtally$scientificName,
                            cdw_fieldtally$sizeCategory,
                            sep = "_")

#   Summarize the number of logs per DST and multiply by the volumeFactor
#   to generate CDW volume (m3/hectare)
cdwVol <- cdw_fieldtally %>%
  dplyr::filter(is.na(samplingImpractical) | samplingImpractical == "OK") %>%
  dplyr::filter(targetTaxaPresent == "Y") %>%
  dplyr::group_by(domainID,
                  siteID,
                  volumeFactor,
                  eventID,
                  plotID,
                  DST) %>%
  dplyr::summarise(logCount = n(),
                   .groups = "drop") %>%
  dplyr::mutate(cdw_m3ha  = as.numeric(volumeFactor) * logCount)

#   Multiply volume by density to obtain mass
cdwMass <- dplyr::full_join(cdwVol,
                            densityDST,
                            by = "DST")

cdwMass$cdw_Mgha <- cdwMass$cdw_m3ha * cdwMass$bulkDensDisk


#   Sum logCount, volume and mass of individual DSTs to the plot level
cdwPlot <- cdwMass %>%
```

```
dplyr::group_by(domainID,
                siteID,
                volumeFactor,
                eventID,
                plotID) %>%
summarise(logCount = sum(logCount, na.rm = T),
          cdw_m3ha = sum(cdw_m3ha, na.rm = T),
          cdw_Mgha = sum(cdw_Mgha, na.rm = T),
          .groups = "drop" )
```