NEON USER GUIDE TO COARSE DOWNED WOOD BULK DENSITY SAMPLING (DP1.10014.001)

PREPARED BY	ORGANIZATION
Samuel Simkin	SCI
Courtney Meier	SCI

CHANGE RECORD

REVISION	DATE	DESCRIPTION OF CHANGE
A	12/29/2017	Initial Release
B	10/09/2019	Section on 'Quality Assurance Samples' added.
C	05/26/2020	Included general statement about usage of neonUtilities R package and statement about possible location changes. Updated taxonomy information. Updated 'Spatial Resolution and Extent' section. Made other clarifications within document.
D	04/11/2022	Added language in section 4 Taxonomy addressing RTE species obfuscation in the data. Updated section 5.4 Data Revision with latest information regarding data release. Updated information regarding the geoNEON package
D.1	12/08/2023	Added identification history section 4.2 to Taxonomy and updated Data Relationships section (3.8) and Analytical Facility Data Quality section (5.5) to reference identification history.
E	04/09/2024	Minor formatting updates
F	03/20/2025	New dataQF value of bagNotTared added. Added information about the new neonUtilities Python package.

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 volumes and dry weights of disks or disk subsamples collected from 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 Bulk Density Sampling and associated metadata from input data. The processed data include parent log characteristics, disk or disk subsample dry weights, disk fresh weights, disk dimensions, and calculated wood bulk density values. 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 Bulk Density Sampling (DP1.10014.001) (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 Bulk Density Sampling (DP0.10014.001) (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., '10014') 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.DP0.10014.001 _dataValidation.csv	NEON Raw Data Validation for Coarse Downed Wood Bulk Density Sampling (DP0.10014.001)
AD[05]	NEON.DP1.10014.001 _variables.csv	NEON Data Variables for Coarse Downed Wood Bulk Density Sampling (DP1.10014.001)
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.000912	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.004825	NEON Algorithm Theoretical Basis Document: OS Generic Transitions
AD[14]	Available on NEON data portal	NEON's Ingest Conversion Language (NICL) specifications

2.2 Acronyms

Acronym	Definition
CDW	Coarse Downed Wood
DBH	Diameter at Breast Height
DST	Decay class x size category x taxonID combination
LIDS	Line Intercept Distance Sampling

3 DATA PRODUCT DESCRIPTION

The Coarse Downed Wood Bulk Density Sampling data product provides bulk density data from downed logs across all terrestrial NEON sites where such logs exist according to CDW tally data (DP1.10010.001)(see Affleck 2008, 2010 for a discussion of the LIDS tally method). Log bulk density is calculated by collecting disks from downed logs, measuring disk dimensions to enable calculation of disk volume, and weighing disks after oven-drying to constant mass (105 °C for a minimum of 48 h). For disks larger than 500 g fresh weight, up to two subsamples representative of the entire disk cross-section are generated, and the fresh mass of the intact disk and the subsample(s) are collected. When disks are collected from a log, the log's decay class is defined according to Forest Service guidelines (Maser et al. 1979), the taxonID is recorded, and the log is parsed into one of three different diameter size categories, as defined by Keller et al. (2004): 2-5 cm, 5-10 cm, and greater than or equal to 10 cm. For the smaller two size categories, disks are only collected from logs greater than or equal to 1 m length; for logs greater than or equal to 10 cm diameter, disks are collected from logs greater than or equal to 2 m length (similar to tally rules defined by Harmon and Sexton 1996).

To target disk collection to the most abundant logs, CDW tally data (DP1.10010.001) are used to create a ranked list of the most abundant logs in each unique decay class x size category x taxonID (DST) combination. A sample size of n=10 is desired when the size category is greater than or equal to 10 cm diameter, and a sample size of n=5 is the goal for size categories < 10 cm diameter; the desired sample size may not always be attained for a given DST. Bulk density sampling is considered complete when the desired sample size is met for those DSTs that cumulatively comprise the most abundant 80% of all DSTs tallied at a given site.

When combined with the CDW tally data product, which provides an estimate of CDW volume per DST, CDW bulk density data enable calculation of CDW mass per DST, as well as total CDW mass at the site scale.

3.1 Spatial Sampling Design

Coarse downed wood surveys are completed at all terrestrial NEON sites that support growth of woody vegetation with diameter at breast height (DBH) greater than or equal to 2 cm, and at those sites where tallies are non-zero, logs are also sampled for bulk density. In general, because CDW bulk density does not scale to a per unit area, sampling is dispersed over available plots, but is not strictly guided by spatial sampling principles. Instead, sampling is driven by the number of DSTs at each site, and the desired sample size per DST. All Distributed Plots (up to 32) and all Tower Plots (either 20 or 30) at a given site may be searched for target logs during bulk density disk sampling, and when permission is granted by the site host, logs may also be sampled outside of established NEON plot boundaries, and up to 50 m from the nearest plot marker. Typical total disk sample sizes range from 100-300 disks per site, and bulk density samples are not collected from logs that intersect LIDS transects. Disk sampling is avoided for these logs because sampling would alter the length of the logs intersecting the transect, and affect the long-term tally dataset.

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.

As much as possible, sampling occurs in the same locations over the lifetime of the Observatory. However, over time some sampling locations may become impossible to sample, due to disturbance or other local changes. When this occurs, the location and its location ID are retired. A location may also shift to slightly different coordinates. Refer to the locations endpoint of the NEON API for details about locations that have been moved or retired: https://data.neonscience.org/data-api/endpoints/locations/

3.2 Temporal Sampling Design

Coarse downed wood bulk density sampling is scheduled twice per site. The first bout occurs within the first 3 years of site operations, and the second bout occurs 5-6 years after the first bout. For each bout, disks are ideally collected from logs within a period that spans no more than 12 months from the time that sampling is initiated. However, samples may be collected over a period of 24 months if target sample sizes for ranked DSTs cannot be obtained during a 12 month period. The two discrete sampling periods in this data product are intended to provide a decay time series in which quantitative density estimates are linked to more uncertain qualitative decay class assessments.

3.3 Variables Reported

All variables reported from the field or laboratory technician (L0 data) are listed in the file, NEON Raw Data Validation for Coarse Downed Wood Bulk Density Sampling (DP0.10014.001) (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 Bulk Density Sampling (DP1.10014.001) (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.

3.4 Spatial Resolution and Extent

The **mappingMethod**, **logDistance**, **logAzimuth**, **sampleEasting**, and **sampleNorthing** fields are primarily used internally by NEON staff for purposes of relocating the subset of logs that need to be sampled more than once. With that caveat, the specific location of some logs can be determined as described below.

The finest resolution at which spatial data are reported is the point location of the log from which disks were collected.

The spatial hierarchy from finest to coarsest resolution is:

point (log location relative to plot) → **plotID** (ID of plot within site) → **siteID** (ID of NEON site) → **domainID** (ID of NEON domain)

The basic spatial data included in the data downloaded include the latitude, longitude, and elevation of the centroid of the plot closest to 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.

This data product includes three mappingMethod options: A) Not mapped, B) GPS, or C) Relative

A) Not mapped: For this mappingMethod option, no finer resolution spatial coordinates beyond the plot centroid are available.

B) GPS: For this mappingMethod option, more precise estimates of log location are provided directly via the **sampleEasting** and **sampleNorthing** fields; the spatial data (UTM, using the zone appropriate for the site) in these fields is collected at the time of log sampling with a recreational grade GPS unit.

C) Relative: For this mappingMethod option, the log location can be calculated using **logDistance** and **logAzimuth** offsets from a known **pointID** (i.e., a NEON plot marker for which high-resolution GPS data have been collected during plot establishment) via two possible options:

- Use the getLocTOS function from the geoNEON package, available here: https://github.com/NEONScience/NEON-geolocation, or
- Perform the calculation manually.

To perform the calculation manually:

1. The namedLocation field in the data is the named location of the plot.
2. Use the API (http://data.neonscience.org/api; e.g. http://data.neonscience.org/api/v0/locations/HARV_052.basePlot.div) to query for elevation("locationElevation"), easting("locationUtmEasting"), northing("locationUtmNorthing"), coordinateUncertainty ("Value for Coordinate uncertainty"), elevationUncertainty ("Value for Elevation uncertainty"), and utmZone ("locationUtmZone") as inputs to the next step. Note that in this data product, unlike in "Coarse Downed Wood Log Survey", logs may be sampled from plots of type cdw or from the larger set of plots of type div.

3. Calculate the geolocation of each log according to the equations:

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

and

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

where,

$$\theta = \frac{(lidsAzimuth + 180deg) * \pi}{180} \tag{3}$$

$$locationUtmEasting = \tag{4}$$

the easting value associated with the pointID

$$locationUtmNorthing = \tag{5}$$

the northing value associated with the pointID

$$d = \tag{6}$$

logDistance

4. Increase coordinateUncertainty by an appropriate amount to account for error introduced during log mapping. Additional error may be introduced due to tape stretching and drift from a compass heading.

3.5 Temporal Resolution and Extent

The finest resolution at which temporal data are reported is the **collectDate**, which is the date a disk sample is collected from a log in the field. Disks collected as part of the same sampling bout at a given site are grouped according to the same **eventID**. The temporal hierarchy within the data is therefore:

collectDate (date a disk was sampled) → **eventID** (bout-level identifier)

3.6 Associated Data Streams

The Coarse Downed Wood Bulk Density data product (DP1.10014.001) is tightly integrated with the Coarse Downed Wood Log Survey data product (DP1.10010.001). Because the Log Survey product provides estimates of CDW volume, taken together, these two data products allow calculation of CDW dry mass per sizeCategory per decayClass per taxonID, as well as total CDW dry mass at the site scale. Although these two data products are related, logs that are tagged during Log Survey are specifically avoided for Bulk Density sampling.

The Woody and Non-Herbaceous Perennial Vegetation Structure data products (DP1.10098.001 and DP1.10045.001, respectively) are also linked to the CDW Bulk Density product due to the fact that individuals tagged and measured while living may be sampled for bulk density once they have fallen to the ground. In the event that a bulk density sample is 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.

3.7 Product Instances

At sites with woody vegetation greater than or equal to 2 cm diameter and average diversity and abundance of CDW particles, between 100-300 disks are collected per bout, and there are 2 bouts per site. Thus, each site with qualifying downed woody particles should produce between 200-600 product instances over the lifetime of the Observatory.

3.8 Data Relationships

Each unique log sampled is assigned a **logID**. Logs that are greater than or equal to 10 cm diameter and that may be repeat sampled across both sampling bouts are assigned a numeric **logID**, and are tagged in the field. Logs < 10 cm diameter that will only be sampled once are assigned an alpha-numeric **logID** that begins with 'L'. Because logs greater than or equal to 10 cm diameter may be sampled during both bulk density bouts at a given site, the **logID** is used in combination with 'year' data to construct a unique **sampleID** within each sampling **eventID**. Because 1 or 2 disks may be collected per log during an **eventID**, a unique **subsampleID** is generated for each disk. 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 field tally table with other data products. To summarize:

cdw_densitylog.csv → One record with unique **sampleID** expected each time a disk is collected from a log during an **eventID**

cdw_densitydisk.csv → 1 or 2 records with unique **subsampleID(s)** (disks) expected per **sampleID**

cdw_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_densitylog tables. The cdw_identificationHistory table is populated with all prior names used for specimen(s) in the data product. When data are populated in the cdw_identificationHistory table, **identificationHistoryID** is used as a linking variable between the cdw_identificationHistory table and the cdw_densitylog table where updates were made.

Recently downed logs in the CDW Bulk Density dataset that were previously measured while standing are identified by the **vstTagID** field. To match **vstTagID** with the linked **individualID** field used 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 data products:
    - Woody Plant Vegetation Structure (DP1.10095.001)
    - Plant Canopy Leaf Mass per Area (DP1.10048.001)
    - Plant Foliar Stable Isotopes (DP1.10053.001)
    - 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 neonUtilities package in R and the neonutilities package in Python contain functions to merge these files across sites and months into a single file for each table. The neonUtilities R 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 neonutilities package in Python 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. # 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.

3.9 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 alpha-numeric 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.

3.10 Identification History

Beginning in 2023, the 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, 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. Upon approval, the existing record in cdw_densitylog 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_identificationHistory table where the previous taxonomic information is archived along with the date the change was made.

4 DATA QUALITY

4.1 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 Bulk Density Sampling (DP0.10014.001), 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[14]). 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.

4.2 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[13]).

4.3 Quality Assurance Samples

The "qaSample" field specifies whether records are replicates for quality assurance purposes. For records with qaSample = "Y" on or after July 1, 2019 the intent was for both field and lab measurements of a sample to receive replicate measurements for quality assurance purposes. Prior to July 1, 2019, the qaSample field indicates whether legacy quality assurance measurements included solely field masses ("legacyFreshMassQA"), lab masses ("legacy DryMassQA"), or both ("legacyFreshAndDryMassQA").

4.4 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.

4.5 Quality Flagging

The **dataQF** field in each data record is a quality flag for known errors applying to the record. Please see below for an explanation of **dataQF** codes specific to this product.

Table 1: 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	bagNotTared	Bag included in field measurement of diskFreshMass. The average cloth bag mass of 43.6 g has been subtracted from all samples with this dataQF. The maximum expected error after correction ranges from 0.7 to 17.4 percent

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)

4.6 Analytical Facility Data Quality

If errors in identification are detected through QAQC processes after data publication, then corrected taxonomy is provided in the cdw_densitylog table and previous taxonomic information is preserved in the cdw_identificationHistory (see Section 4.2 Identification History and Section 3.8 Data Relationships above for more details).

5 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. 2010. On the efficiency of line intersect distance sampling. Canadian Journal of Forest Research 40:1086-1094.

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

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.

Maser, C., R. G. Anderson, K. Cromack Jr., J. T. Williams, and R. E. Martin. 1979. Dead and Downed Woody Material, in Wildlife Habitats in Managed Forests, J.W. Thomas, editor, Agricultural Handbook No. 553., U.S. Department of Agriculture.

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