NEON USER GUIDE TO WOODY PLANT VEGETATION STRUCTURE (NEON.DP1.10098) AND NON-HERBACEOUS PERENNIAL VEGETATION STRUCTURE (NEON.DP1.10045)

PREPARED BY	ORGANIZATION	DATE
Tanya Chesney	DPS	10/12/2017
Katie Jones	SCI	10/12/2017
Courtney Meier	SCI	10/12/2017

CHANGE RECORD

REVISION	DATE	DESCRIPTION OF CHANGE
A	08/30/2017	Initial Release

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 stem diamater of an individual tree 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.

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 digital workflow utilized from 2017 onward. For all data, 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. Please see the *Special Considerations* section of this document for a list of known errors that may be present in the data, and an explanation of the data quality flag codes specific to this product.

1.2 Scope

This document describes the steps needed to generate the L1 data product Woody plant vegetation structure (DP1.10098) - Structure measurements, including height, canopy diameter, and stem diameter, as well as mapped position of individual woody plants (excluding lianas) and Non-herbaceous perennial vegetation structure (DP1.10045) - Field measurements of individual non-herbaceous perennial plants (e.g. cacti, ferns) - 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 files NEON Data Variables for Woody Plant Vegetation Structure (NEON.DP1.10098) (AD[05]) and NEON Data Variables for Non-herbaceous Perennial Vegetation Structure (NEON.DP1.10045) (AD[06]), 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: Measurement of Vegetation Structure (AD[08]). The raw data that are processed in this document are detailed in the file, NEON Raw Data Validation for Vegetation Structure, Level 0 (NEON.DP0.10098) (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., '10033') 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 Level 1, Level 2 and Level 3 Data Products Catalog
AD[04]	NEON.DP0.10098.001_dataValidation.csv	NEON Raw Data Validation for Vegetation Structure, Level 0 (NEON.DP0.10098)
AD[05]	NEON.DP1.10098.001_variables.csv	NEON Data Variables for Woody Plant Vegetation Structure (NEON.DP1.10098)
AD[06]	NEON.DP1.10045.001_variables.csv	NEON Data Variables for Non-herbaceous Perennial Vegetation Structure (NEON.DP1.10045)
AD[07]	NEON.DOC.000914	TOS Science Design for Plant Biomass, Productivity, and Leaf Area Index
AD[08]	NEON.DOC.000987	TOS Protocol and Procedure: Measurement of Vegetation Structure
AD[09]	NEON.DOC.000008	NEON Acronym List
AD[10]	NEON.DOC.000243	NEON Glossary of Terms
AD[11]	OS_Generic _Transitions.pdf	NEON Algorithm Theoretical Basis Document: OS Generic Transitions
AD[12]		NEON's Ingest Conversion Language (NICL) specifications

2.2 Acronyms

Acronym	Definition
ANPP	*Annual Net Primary Productivity*
AOP	*Airborne Observation Platform*
DBH	*Diameter at Breast Height (130 cm)*
ddh	*Diameter at Decimeter Height (10 cm)*
TIS	*Terrestrial Instrument System*

3 DATA PRODUCT DESCRIPTION

Vegetation structure data includes information related to structure, spatial design, and biomass of the woody vegetation and perennial non-herbaceous plant community, including tree, sapling, shrub, liana, palm, fern, and other growth forms. Specific data collected include taxonomic identifications, diameter at breast height (DBH), diameter at decimeter height (ddh), total stem height, crown diameter, plant status (i.e. healthy, dead, damaged, etc.), and the location of measured stems ≥ 10 cm DBH (stems < 10 cm DBH may also be mapped at some sites). Parameters such as DBH, ddh, crown diameter and total stem height can then be used with allometric equations to estimate aboveground biomass and carbon (C) density values, on both a per stem and a per unit area basis.

The measurement of vegetation structure and the mapping of free-standing woody stems is an important complement to data streams generated by the NEON Airborne Observation Platform (AOP) and Terrestrial Instrument System (TIS). These ground-collected data can be used to validate LiDAR data used to map the structural complexity of vegetationn as well as enable mapping of plant biomass at the site scale. In conjunction with carbon flux data, vegetation structure data will facilitate understanding how biomass in different plant growth forms contributes to ecosystem level carbon flux.

3.1 Spatial Sampling Design

At sites with qualifying woody and/or non-herbaceous vegetation, stem mapping activities and the collection of vegetation structure data take place in up to n=20 randomly selected, spatially-balanced Distributed Plots, all Tower Plots (n=20 or n=30), and may also occur in Gradient Plots (Figure 1, Figure 2). Taken together, 40 or 50 plots are typically sampled at a given site with qualifying vegetation. At sites with short-stature woody vegetation (e.g., shrub scrub or re-generating forest), n=30 Tower Plots are sampled, and plot dimensions are 20 m x 20 m. At most forested sites, n=20 Tower Plots are sampled, and plot dimensions are 40 m x 40 m, with two randomly selected 20 m x 20 m subplots that are sampled. Gradient and Distributed Plots always have plot dimensions of 20 m x 20 m.

An offset mapping technique is used to determine the within-plot location of mapped individuals relative to permanent plot markers (pointIDs) for which high-resolution GPS data are collected (Figure 3). For additional details on the sampling design and associated protocol, see TOS Science Design for Plant Biomass, Productivity, and Leaf Area Index (AD[07]) and TOS Protocol and Procedure: Measurement of Vegetation Structure (AD[08]).

Ferns and 'other' growth forms are not measured in Distributed Plots with forested NLCD classes (Evergreen Forest, Deciduous Forest, and Mixed Forest), and are not measured in non-forested Distributed Plots when they constitute < 50% cover over the total plot area (these are the same guidelines that trigger implementation of the Herbaceous Biomass clip harvest protocol in Distributed Plots).

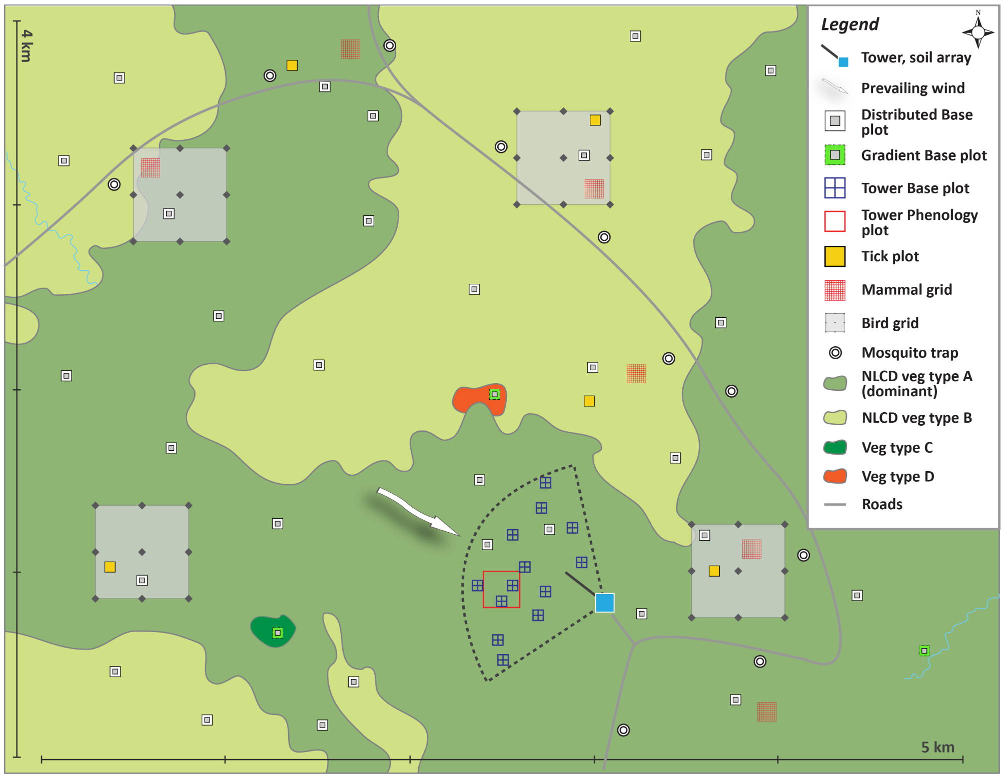

Figure 1: Generalized TOS sampling schematic, showing the placement of Distributed, Tower, and Gradient Plots.

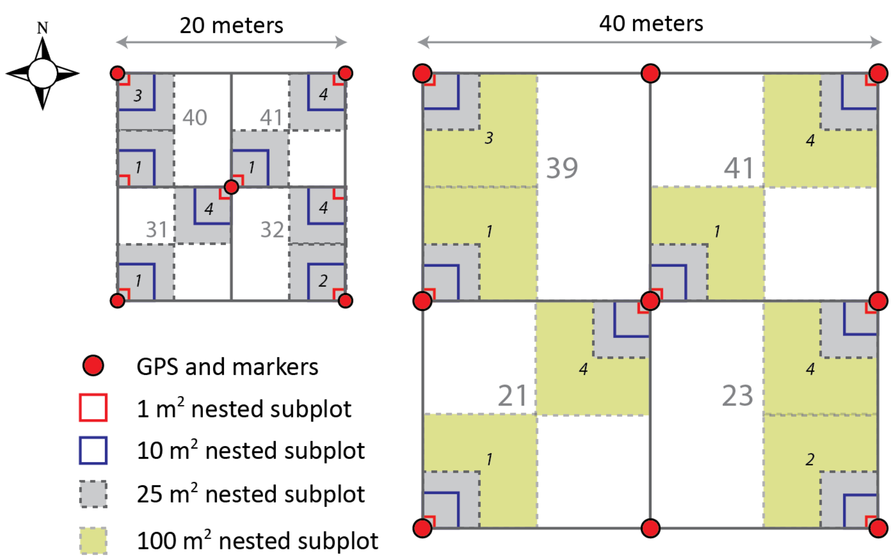

Figure 2: Illustration of a 20 m x 20 m Distributed/Gradient/Tower base plot (left), a 40 m x 40 m Tower base plot (right), and associated nested subplots used for measuring woody stem vegetation. Locations of subplots are denoted with plain text numbers, and locations of nested subplots are denoted with italic numbers.

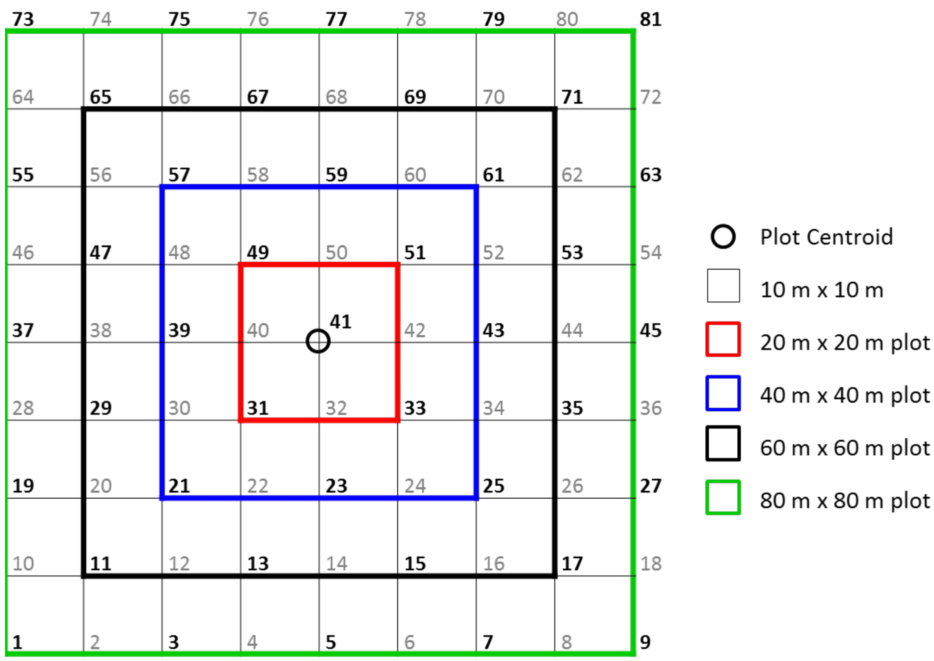

Figure 3: Plot pointIDs are numbered according to the largest theoretical plot size, 80 m x 80 m, such that a plot of any size will use a consistent numbering scheme. A pointID has been for defined for every point on a grid with 10 m spacing. The 10 m grid is used during initial plot establishment, and permanent primary and secondary markers are placed at 20 meter intervals. During stem mapping, technicians only use the points that have permanent markers in place, those identified as bold in this diagram. The subplotID within a plot is determined by the pointID of the SW corner of the subplot area. The plot centroid is always pointID = 41.

3.2 Temporal Sampling Design

At sites with a defined growing season (e.g., North temperate, Mediterranean, etc.), Vegetation Structure measurements begin only after the majority of annual growth within a given season has completed. At sites with no distinct growing season, sampling onset is scheduled at the same time every year ± 2 weeks. At all sites, sampling is completed within 4 months of onset, and before the next growing season commences.

The interannual sampling frequency for a given site depends on plot type, and also ecosystem type:

- 'Typical' relatively mesic sites:
  - Distributed Plots: Every 3 y
  - Tower Plots: Annual
- Continental cold/dry sites:
  - Distributed Plots: Every 3 y
  - Tower Plots: Every 3 y
- Boreal sites:
  - Distributed Plots: Every 6 y

- Tower Plots: Every 6 y

Cold/dry sites are sampled less frequently than relatively mesic sites because annual growth increments are often too small to be detected using DBH tapes, and/or ecosystems of this type may be fragile and susceptible to damage with annual visitation. Moreover, at continental cold/dry and boreal sites, Distributed and Tower Plots may not be sampled in the same year at a given site. See TOS Science Design for Plant Biomass, Productivity, and Leaf Area Index (AD[07]), and the TOS Protocol and Procedure: Measurement of Vegetation Structure (AD[08]) for more details.

3.3 Variables Reported

All variables reported from the field (L0 data) are listed in the file, NEON Raw Data Validation for Vegetation Structure, Level 0 (NEON.DP0.10098) (AD[04]). All variables reported in the published data (L1 data) are also provided separately in the files NEON Data Variables for Woody Plant Vegetation Structure (NEON.DP1.10098) (AD[05]) and NEON Data Variables for Non-herbaceous Perennial Vegetation Structure (NEON.DP1.10045) (AD[06]).

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 Earth Gravitational Model 96 (EGM96) for its reference gravitational ellipsoid. 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

Spatial resolution varies by growth form. Trees >10 cm diameter are mapped are mapped as points (offsets from a reference pointID; see table vst_mapping); small diameter trees and shrubs are not mapped but are identified according to the subplot or nested subplot in which they are located (see tables vst_apparentindividual and vst_shrubgroup) as are non-woody plants (see table nst_perinidividual). Lianas are not mapped but a user could infer their approximate position from the identity of the their host tree (**supportingStemID**). For growth forms that are not mapped as points, the finest spatial resolution is the level of the nested subplot, which may range in size from 1-100 m^2, depending on the density of vegetation.

Spatial hierarchy from finest to coarsest resolution=

point (optional) → **nestedSubplotID** (optional) → **subplotID** → **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 in the Document Library: http://data.neonscience.org/documents.

To derive a more precise estimate of the location of each individual tree in vst_mappingandtagging, there are two options:

1. Use the def.calc.geo.os function from the geoNEON package, available here: https://github.com/NEONScience/NEON-geolocation, or
2. Follow these steps to perform the same calculation:
    a. The namedLocation field in the vst_mappingandtagging data is the named location of plot; more precise geographic data requires the named location pointID from which relative positions of mapped stems were measured. Construct the named location of the pointID associated with each record in vst_mappingandtagging by concatenating the fields for namedLocation and pointID as: namedLocation + '.' + pointID, e.g. pointID '21' of namedLocation 'BART_001.basePlot.vst' has a complete named location of 'BART_001.basePlot.vst.21'.
    b. Use the API (http://data.neonscience.org/api; e.g. http://data.neonscience.org/api/v0/locations/BART_001.basePlot.vst.21) to query for elevation("locationElevation"), easting("locationUtmEasting"), northing("locationUtmNorthing"), coordinateUncertainty ("Value for Coordinate uncertainty"), elevationUncertainty ("Value for Elevation uncertainty"), and utmZone ("locationUtmZone") for each pointID as inputs to the next step.
    c. Calculate absolute position in UTMs of each mapped stem (**tagID**) based on **stemAzimuth** and **stemDistance** and the easting and northing values derived in step 1, using equations (1) and (2):

$$Easting = easting.pointID + d * \sin\theta \tag{1}$$

and

$$Northing = northing.pointID + d * \cos\theta \tag{2}$$

where,

$$\theta = \frac{stemAzimuth * \pi}{180} \tag{3}$$

$$easting.pointID = \tag{4}$$

the easting value of the reference pointID,

$$northing.pointID = \tag{5}$$

the northing value of the reference pointID,

$$d = \tag{6}$$

stemDistance

d. Increase **coordinateUncertainty** associated with the reference pointID by an appropriate amount (suggested 0.6 m) to account for error introduced by navigating around the plot and increase **elevationUncertainty** associated with the reference pointID by an appropriate amount to account for topographical heterogeneity (suggested 1 m).

Example: namedLocation is BART_001.basePlot.vst, pointID is 21. Easting of the pointID is 315188.32; northing of the pointID is 4879671.04; coordinateUncertainty of the pointID is 0.16; elevation of the pointID is 484.55; elevationUncertainty of the pointID is 0.2. If the stemAzimuth is 45 degrees and the stemDistance is 2m, θ *= 0.785, easting of the mapped stem = 315188.32 + 2* $\times \sin$ *0.785 = 315189.7 and northing of the mapped stem = 4879671.04 + 2* $\times \cos$ *0.785 = 4879672.0. The coordinateUncertainty is assumed to be* $\approx$ *0.8m; the elevation is assumed to be 484.55 with an elevationUncertainty of* $\approx$ *1.2m*

To derive a more precise estimate of the location of each subplot or nestedSubplot in nst_perindividual, vst_shrubgroup and/or vst_apparentindividual, there are two options:

1. Use the def.calc.geo.os function from the geoNEON package, available here: https://github.com/NEONScience/NEON-geolocation, or
2. Follow these steps to perform the same calculation and generate more precise subplot and/or nestedSubplot spatial locations within a plot:
   a. To calculate the spatial location of **subplot** centroids within a plot:
      i. Construct the named location of the pointID that lies at the Southwest corner of the desired subplot by concatenating the fields for namedLocation and pointID as: `namedLocation` + `'.'` + `pointID`.
         1. *Example*: For plotID = `BART_001` and pointID = 31, the complete pointID named location = `BART_001.basePlot.vst.31`
      ii. Use the NEON Data API to query location information for the pointID at the Southwest corner of the desired subplot (http://data.neonscience.org/api). For 20m x 20m plots, the list of southwest corner pointIDs = (31,32,40,41); for 40m x 40m Tower Plots, the list of southwest corner pointIDs = (21,23,39,41).
         1. *Example*: To query for spatial information for plotID = `BART_001` and pointID = 31 at the NEON Bartlett Experimental Forest site, use: http://data.neonscience.org/api/v0/locations/BART_001.basePlot.vst.31
      iii. For each desired pointID, query for:
         1. elevation (**locationElevation**),
         2. easting (**locationUtmEasting**),
         3. northing (**locationUtmNorthing**),
         4. coordinateUncertainty (“Value for Coordinate Uncertainty”),
         5. elevationUncertainty (“Value for Elevation Uncertainty”), and
         6. utmZone (**locationUtmZone**)
      iv. Calculate the easting and northing for the centroid of each desired subplotID. For 20m x 20m plots, add 5 m to the **locationUtmEasting** and **locationUtmNorthing** from the pointID at the SW corner of the subplot. For 40m x 40m plots, add 10 m to the **locationUtmEasting** and **locationUtmNorthing** from the pointID at the SW corner of the subplot.
      v. Increase **coordinateUncertainty** associated with the monumented pointID by an appropriate amount, to account for error introduced by navigating around the plot (suggested 0.6 m).

vi. Increase **elevationUncertainty** associated with the monumented pointID by an appropriate amount to account for topographical heterogeneity (suggested 1 m).

b. To calculate the spatial location of **nestedSubplot** centroids within a plot:
   i. Construct the named location of the nearest monumented pointID for which high-resolution GPS data are available by concatenating the fields for namedLocation and pointID as: `namedLocation` + `'.'` + `pointID`.
      1. Use plotSize and plotType information to determine the monumented pointID that is closest to each desired nestedSubplot (Figure 2).
      2. Note that in 20m x 20m plots, there are two nestedSubplots that do not have an immediately adjacent monumented pointID: `32.4` and `40.1` (using `subplotID.nestedSubplotID` format). For these two nestedSubplots, choose one of the two closest pointIDs.
      3. *Example*: For plotID = `BART_001`, subplotID = 32, and nestedSubplotID = 2, the nearest monumented pointID = 33 (Figure 2, *left*), and the complete pointID named location = `BART_001.basePlot.vst.33`
   ii. Use the NEON Data API to query location information for the closest monumented pointID (http://data.neonscience.org/api), as above.
   iii. On a per growth form basis, calculate relative easting and northing offsets for the centroid of each nestedSubplot in UTMs based on the **nestedSubplotID** and the growth-form-specific **nestedSubplotArea** from the vst_perplotperyear table for the desired **eventID**. Then, add the calculated offsets to the easting and northing values from the nearest monumented pointID.
      1. *Example*: For plotID = `BART_001`, we wish to calculate the centroid location for nestedSubplotID = `41.4` that was used to measure individuals with **growthForm** = `single shrub` (Figure 2, *left*).
         - Closest monumented pointID = 51
         - In the vst_perplotperyear table, we find that **nestedSubplotAreaShrubSapling** = 25 for the `'single shrub'` growthForm.
         - Relative to pointID = `BART_001.basePlot.vst.51`, offsetEasting and offsetNorthing = $-\frac{1}{2} * sqrt(nestedSubplotAreaShrubSapling)$ = $-\frac{1}{2} * 5$ = `-2.5 m`
         - Add offsetEasting and offsetNorthing to easting and northing values queried for `BART_001.basePlot.vst.51`.
   iv. Increase **coordinateUncertainty** associated with the monumented pointID by an appropriate amount (suggested 0.6 m) to account for error introduced by navigating around the plot.
   v. Increase **elevationUncertainty** associated with the monumented pointID by an appropriate amount to account for topographical heterogeneity (suggested 1 m).

3.5 Temporal Resolution and Extent

The finest temporal resolution at which vegetation structure records are tracked is the level of the day of year (local time) that a particular measurement occurred. All sampling at a site that occurs within a given bout is identified by a unique eventID, which is the temporal resolution that may be used to calculate site level estimates of biomass (Distributed and Tower plots) and ANPP (Tower plots only). Sampling bouts may take up to 9 months to complete.

3.6 Associated Data Streams

The **individualID** (NEON.PLA.D##.######) is a linking variable that ties vegetation structure measurements and associated metadata to the following associated data products:

- Plant phenology observations (NEON.DP1.10055),
- Plant foliar physical and chemical properties (NEON.DP1.10026), and
- Plant canopy leaf mass per area (NEON.DP1.10048)

For tagged woody individuals ≥ 10 cm diameter that fall down and become part of the Coarse Downed Wood log survey dataset (NEON.DP1.10010), the **vstTagID** variable in cdw_fieldtally corresponds to the last 6 digits of the Vegetation Structure **individualID**, and **individualID** can therefore be used to track trees/logs across the two datasets.

Additional information from collocated measurements of Litterfall and fine woody debris (NEON.DP1.10033), herbaceous clip harvests (NEON.DP1.10023), and coarse downed wood log surveys (NEON.DP1.10010) and coarse downed wood bulk density (NEON.DP1.10014) may be used together with vegetation structure measurements to estimate total Aboveground Net Primary Productivity. Remotely sensed Total biomass map - spectrometer - flight-line (DP2.30016.001) and Total biomass map - spectrometer - mosaic (NEON.DP3.30016) and Ecosystem structure (NEON.DP3.30015) data products can also be used to upscale plot-based measurements of vegetation structure.

3.7 Product Instances

At each NEON site where woody vegetation exists, sampling occurs annually at 20-30 Tower Plots with an additional 20 Distributed Plots measured once every 3 years. Each sampled plot should have one entry for vst_perplotperyear. Each plot may contain 0-150+ uniquely tagged individuals in vst_mappingandtagging, some of which may be reentered over time as corrections to taxonomy or location are noted. The number of records in vst_apparentindividual, vst_shrubgroup and nst_perindividual varies with the density of the vegetation.

3.8 Data Relationships

The Woody Plant Vegetation Structure data product is comprised of four related tables:

vst_perplotperyear → One record is created for each plotID visited at a site during a given eventID. Recorded metadata include presence/absence of various woody growth forms, and nestedSubplotArea used to measure each growth form (if applicable).

vst_mappingandtagging → One or more records expected for each tagged woody individual (**individualID**) for all time. A subset of tagged individuals are also mapped. Corrections to taxonID or mapped location through time will result in duplicate **individualIDs**. In the case of duplicates, users are advised to retain the record with the latest **date**. Non-woody individuals that are also mapped and/or tagged (e.g., cacti at some sites, small palms, etc.) are also recorded in vst_mappingandtagging.

vst_apparentindividual → For each unique woody **individualID** recorded in vst_mappingandtagging, technicians create at least one record in vst_apparentindividual for each **eventID**. Multiple records per **individualID** are created for multi-bole individuals of the following growthForms: small tree, sapling, single shrub, small shrub. To scale up measurements from individuals to the plot-scale, per growthForm measurement areas are linked by **plotID** and **eventID** in vst_perplotperyear.

vst_shrubgroup → Shrub groups are not mapped or tagged, and are assigned a temporary **groupID** that is unique within a **plotID** within a given **eventID**. For each shrub group (**groupID**), technicians create at least one record, and multiple records may exist per **groupID** in the event that the shrub group is comprised of multiple **taxonIDs**. To scale up measurements from shrub groups to the plot-scale, measurement areas are linked by **plotID** and **eventID** in vst_perplotperyear.

nst_perindividual → For each unique non-herbaceous perennial **individualID** recorded in vst_mappingandtagging, technicians create at least one record in nst_perindividual for each **eventID**. Additional records are created for individuals that are measured but not mapped or tagged (e.g., ferns, cacti at some sites, etc.). To scale up measurements from individuals to the plot-scale, measurement areas are linked by **plotID** and **eventID** in vst_perplotperyear.

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). An additional set of constraints are implemented during the process of ingest into the NEON database. In aggregate, these controls include but are not limited to:

- Constraint of data types and formats to reduce post-processing steps necessary to prepare the raw data for publication.
- Constraint of data values for a given field through the provision of dropdown options, range limitations, and filtering based on values entered into other fields.
- Checks for existing individualIDs or sampleIDs within the NEON database.
- Checks for values within required fields.

The complete list of 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 Vegetation Structure, Level 0 (NEON.DP0.10098), 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[12]).

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 of 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[11]).

4.3 Data Revision

All data are provisional until a numbered version is released; the first release of a static version of NEON data, annotated with a globally unique identifier, is planned to take place in 2020. 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 Change Log section of the data product readme downloaded from the data portal contains a history of major known errors and revisions.

4.4 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.

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

5 SPECIAL CONSIDERATIONS

5.1 Scaling from the Individual to the Plot

The vst_perplotperyear table contains sampling metadata necessary to make plot-scale inference about stem density, biomass, productivity, and other derived variables from measured individuals and/or shrub groups. There are four general groups of variables within vst_perplotperyear that work together to document the sampling effort, and enable scaling up from individuals to the total plot area:

1. For all growthForms within each plotID, technicians document whether individuals of that growth form are present anywhere within the plot.
    a. Variables in this group include:
        - **treesPresent**
        - **shrubsPresent**
        - **lianasPresent**
        - **cactiPresent**
        - **fernsPresent**
        - **yuccasPresent**
        - **palmsPresent**
        - **ocotillosPresent**, and
        - **xerophyllumPresent**
    b. Values in these fields clearly indicate that each plotID is assessed for each growthForm, and allow end-users to know when individuals from a given growthForm are absent from an entire plot (as opposed to simply forgotten or not sampled). No further documentation is provided for growthForms absent at the plot scale.

2. For woody individuals other than single- and multi-bole trees, technicians may employ nestedSubplots of various sizes to standardize the sampling effort. The **nestedSubplotArea** group of variables documents the size of nestedSubplot used per growthForm within a given plotID.
    a. Variables in this group include:
        - **nestedSubplotAreaShrubSapling**
        - **nestedSubplotAreaLiana**, and
        - **nestedSubplotAreaOther**.
    b. For the **nestedSubplotAreaOther**, one nestedSubplot size is used to assess all 'other' growthForms within the plot. That is, all cacti, ferns, yuccas, palms, ocotillo and xerophyllum individuals in a plot are combined when determining whether a nestedSubplot can be used for individuals of these growthForms.
    c. For each **plotID** and **eventID** technicians are required to record nestedSubplotArea on a per growth-Form basis when nestedSubplots are used, and to record `noneSelected` for each growthForm for which nestedSubplots were NOT used.
3. The group of **AbsentList** fields is required to account for those nestedSubplots that were assessed for each growthForm, but did not contain individuals of that growthForm. Nested subplots in these lists should be counted as having zero individuals for the specified growthForm when scaling up to the plot.
    a. Variables in this group include:
        - **treesAbsentList**
        - **shrubsAbsentList**
        - **lianasAbsentList**
        - **cactiAbsentList**
        - **fernsAbsentList**
        - **yuccasAbsentList**
        - **palmsAbsentList**
        - **ocotillosAbsentList**, and
        - **xerophyllumAbsentList**
    b. For nestedSubplots NOT in an **AbsentList** for a given growthForm, there should be records in any or all of the vst_apparentindividual, vst_shrubgroup, or nst_perindividual tables associated with the given **plotID**, **eventID** and **nestedSubplotID**.
4. The group of **totalSampledArea** fields multiplies the **nestedSubplotArea** utilized for a given growthForm by the number of expected nestedSubplots within the plot, and provides the total sampled area required to scale-up to the plot level for each growthForm.
    a. Variables in this group inclue:
        - **totalSampledAreaTrees**
        - **totalSampledAreaShrubSapling**
        - **totalSampledAreaLiana**, and
        - **totalSampledAreaOther**

5.2 Duplicates

Where location and/or taxonomic identifications of individual stems are updated through time, duplicate observations on the same individualID may be found in vst_mappingandtagging. Where duplicate records are present in the vst_mappingandtagging table, users are advised to retain the record with the latest date.

Other duplicates and/or missing data may exist where protocol and/or data entry abberations have occurred; users should check data carefully for anomalies before joining tables.

5.3 Glossary of 'dataQF' Flags

1. Flag = `legacyData`: A non-exhaustive list of known issues associated with this flag includes:
    - Diameter measurements may be missing for smaller woody growth forms (e.g., shrubs).
    - Measurements required by the protocol for selected combinations of **growthForm** and shrub **shape** may not be consistent with provided measurements (e.g., **baseCrownDiameter**(s) may be provided for shrub shapes other than inverted cone).
    - Out of range values may exist, due to incorrect units or other errors.
    - Values for **growthForm** may be missing for measured individuals.
    - There are records that should have mapping data (i.e., **pointID**, **stemDistance** and **stemAzimuth** values), but do not, due to incorrectly entered data that were removed from the dataset following initial QC checks.
    - There will eventually be duplicate **individualIDs** for these individuals as new records are created that contain corrected data (see *Duplicates* section above for guidance).
    - Some single- and multi-bole tree records do not have a value reported for **canopyPosition**, and should.

6 DATA PROCESSING STEPS

Following data entry into a mobile application of 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[11]).