NEON USER GUIDE TO WOODY PLANT VEGETATION STRUCTURE (DP1.10098.001) AND NON-HERBACEOUS PERENNIAL VEGETATION STRUCTURE (DP1.10045.001)

PREPARED BY	**ORGANIZATION**
Courtney Meier	SCI
Tanya Chesney	DPS
Katie Jones	SCI

CHANGE RECORD

REVISION	DATE	DESCRIPTION OF CHANGE
A	01/15/2018	Initial Release
B	10/15/2020	Significant changes by section: • Section 3.1 and 3.2: Updated Spatial Sampling Design and Temporal Samplign Design to reflect the current sampling design as documented in protocol versions H and J. • Section 3.3: New Sampling Design Changes section to document specific protocol changes that affect the data. • Section 3.8: Updated Product Instances based on changes made to the spatial and temporal sampling design. • Section 3.9: Revised Data Relationships to focus more clearly on primary keys within tables, linking variables across tables, and to include neonUtilities R package information. • Section 5.4: Developed sub-sections to communicate QC Data recorded in the field, and post-hoc quality flagging. • Section 6: Revised Special Considerations to contain sub-sections per data table, and added information to help data users understand and utilize data structure and new fields introduced in protocol versions H and J.
C	04/25/2022	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.

TABLE OF CONTENTS

1 DESCRIPTION **1**

1.1 Purpose . . . 1

1.2 Scope . . . 1

2 RELATED DOCUMENTS AND ACRONYMS **2**

2.1 Associated Documents . . . 2

2.2 Acronyms . . . 3

3 DATA PRODUCT DESCRIPTION **4**

3.1 Spatial Sampling Design . . . 4

3.2 Temporal Sampling Design . . . 7

3.3 Sampling Design Changes . . . 8

3.4 Variables Reported . . . 12

3.5 Spatial Resolution and Extent . . . 12

3.5.1 Mapped individuals . . . 13

3.5.2 Unmapped individuals . . . 15

3.6 Temporal Resolution and Extent . . . 16

3.7 Associated Data Streams . . . 16

3.8 Product Instances . . . 17

3.9 Data Relationships . . . 18

4 TAXONOMY **19**

5 DATA QUALITY **20**

5.1 Data Entry Constraint and Validation . . . 20

5.2 Automated Data Processing Steps . . . 20

5.3 Data Revision . . . 20

5.4 Quality Flagging . . . 20

5.4.1 QC Data Recorded in the Field . . . 21

5.4.2 Post-Hoc Quality Flagging . . . 21

6 SPECIAL CONSIDERATIONS **22**

6.1 The vst_perplotperyear table 22

6.2 The vst_mappingandtagging table 24

6.3 The vst_apparentindividual table 25

6.4 The nst_perindividual table 27

7 DATA PROCESSING STEPS **29**

8 REFERENCES **29**

LIST OF TABLES AND FIGURES

Table 1 Summary of significant sampling design changes for the NEON TOS 'Woody plant vegetation structure' and 'Non-herbaceous perennial vegetation structure' data products. 9

Table 1 Summary of significant sampling design changes for the NEON TOS 'Woody plant vegetation structure' and 'Non-herbaceous perennial vegetation structure' data products. 10

Table 1 Summary of significant sampling design changes for the NEON TOS 'Woody plant vegetation structure' and 'Non-herbaceous perennial vegetation structure' data products. 11

Table 2 Field-collected vegetation structure data quality variables and descriptions. 21

Table 3 Vegetation structure post-hoc data quality flagging codes and descriptions. 22

Table 4 Summary of woody vegetation structure data by growth form in the vst_apparentindividual table. Note that individuals may change growth form from one year to the next, potentially necessitating different measurements from year-to-year (e.g., a change from 'sapling' to 'small tree'). 26

Table 4 Summary of woody vegetation structure data by growth form in the vst_apparentindividual table. Note that individuals may change growth form from one year to the next, potentially necessitating different measurements from year-to-year (e.g., a change from 'sapling' to 'small tree'). 27

Table 5 Summary of non-herbaceous perennial vegetation structure data by growth form in the nst_perindividual table. 28

Table 5 Summary of non-herbaceous perennial vegetation structure data by growth form in the nst_perindividual table. 29

Figure 1 Generalized TOS sampling schematic, showing the placement of Distributed and Tower plots. 6

Figure 2 Illustration of a 20 m x 20 m Distributed or Tower base plot (left), a 40 m x 40 m Tower base plot (right), and associated nested subplots used for measuring woody and non-herbaceous perennial vegetation. Locations of subplots are denoted with plain text numbers, and locations of nested subplots are denoted with italic numbers. 6

Figure 3 MODIS-EVI timecourse data for the Domain 14 Santa Rita Experimental Range site (D14 SRER). Data are from the years 2005-2014. There is an initial minor green-up early in the year centered around day 80, and another major green-up associated with the summer monsoon centered around day 240. Arrows indicate key greenness transitions that inform the onset of TOS sampling. 8

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 are considered the lowest level (Level 0); an example of Level 0 data is the stem diameters of individual trees. 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. This document provides a discussion of measurement theory and implementation, data product provenance, quality assurance and control methods, and approximations and/or assumptions made during generation of published Level 1 data.

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 *Data Quality* 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 Level 1 data products:

- 'Woody plant vegetation structure' (DP1.10098) - Field measurements from woody plants that include: height, crown diameter, stem diameters (breast height and/or basal), as well as mapped position of qualifying individual woody individuals.
- 'Non-herbaceous perennial vegetation structure' (DP1.10045) - Field measurements from non-herbaceous perennial plants that include: height, crown diameter, basal crown diameter, shape, leaf counts and dimensions, and other growth-form-specific variables. Non-herbaceous perennial vegetation includes the following growth forms: cacti, ferns, ocotillo, palms, tree ferns, xerophyllum, and yucca.

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 (DP1.10098.001) (AD[05]) and NEON Data Variables for Non-herbaceous Perennial Vegetation Structure (DP1.10045.001) (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 (DP0.10098.001) (AD[04]), provided in the download package for this data product. Please note that raw Level 0 data products (denoted by 'DP0') may not always have the same numbers (e.g., '10033') as the corresponding Level 1 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 (DP0.10098.001)
AD[05]	NEON.DP1.10098.001 _variables.csv	NEON Data Variables for Woody Plant Vegetation Structure (DP1.10098.001)
AD[06]	NEON.DP1.10045.001 _variables.csv	NEON Data Variables for Non-herbaceous Perennial Vegetation Structure (DP1.10045.001)
AD[07]	NEON.DOC.000914	TOS Science Design for Plant Biomass and Productivity
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	*Above-ground 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*
TOS	*Terrestrial Observation System*

3 DATA PRODUCT DESCRIPTION

NEON Terrestrial Observation System (TOS) vegetation structure data includes information related to structure, spatial location, and biomass of woody and non-herbaceous perennial vegetation. These vegetation types include tree, sapling, shrub, liana, cactus, 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 individuals ≥ 10 cm DBH (individuals < 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 plot-level per unit area basis. Longitudinal data collected from the same individuals or plots through time enable estimation of an important component of above-ground annual net primary productivity (ANPP) and associated uncertainty.

The measurement of vegetation structure and the mapping of free-standing woody individuals 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 characteristics of vegetation as well as enable mapping of plant biomass at the site scale. In conjunction with carbon flux data, vegetation structure data facilitate understanding how biomass in different plant growth forms contributes to ecosystem level carbon flux.

3.1 Spatial Sampling Design

Within each NEON terrestrial site, NEON TOS vegetation structure 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 (AD[02], 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 (AD[02], Fry *et al.* 2011). Within the tower airshed, a spatially-balanced randomized sampling design is employed.

At sites with qualifying woody and/or non-herbaceous perennial vegetation, stem mapping activities and the collection of vegetation structure data take place in up to n=20 Distributed plots. Vegetation structure data are collected from all Tower plots. Taken together, 40 or 50 plots are typically sampled at a given site with qualifying vegetation. At sites with short-stature vegetation (e.g., shrub scrub or re-generating forest), n=30 Tower plots are sampled, plot dimensions are 20 m x 20 m, and plots are comprised of four 10 m x 10 m subplots. At most forested sites, n=20 Tower plots are established, plot dimensions are 40 m x 40 m, and two randomly selected 20 m x 20 m subplots within each plot are sampled. Distributed plots always have plot dimensions of 20 m x 20 m and are comprised of four 10 m x 10 m subplots (Figure 2).

Each Distributed plot is sampled in a given bout if at least one tree with Diameter at Breast Height (DBH) ≥ 10 cm is present; if trees with DBH ≥ 10 cm are absent, Distributed plots are sampled if smaller woody individuals comprise ≥ 10% aerial cover of the plot. Tower plots are sampled if at least one tree with DBH

≥ 10 cm is present in ≥ 10% of tower plots, or if smaller woody individuals comprise ≥ 10% of aerial cover averaged across all Tower plots. Within both Distributed and Tower plots, all individuals with DBH ≥ 10 cm are mapped and measured throughout the plot sampling area. Individuals with DBH < 10 cm are mapped if a) individuals with DBH ≥ 10 cm are absent from the plot, and b) they are visible to airborne remote-sensing instruments. If standardized stem density thresholds are met, individuals with DBH < 10 cm may be measured within nested subplots in order to standardize the sampling effort across plots (Figure 2). 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 2).

Non-herbaceous perennial vegetation structure data are also collected from Distributed and Tower plots. Each Distributed plot is sampled in a given bout if at least one tree palm is present. If tree palms are absent, smaller qualifying non-herbaceous perennial vegetation is considered together with smaller qualifying woody vegetation; both non-herbaceous perennial and woody vegetation are sampled if all qualifying individuals of both types comprise ≥ 10% aerial cover of the plot. For example, a Distributed plot with 10% aerial cover comprised of a mix of small shrubs and yucca would be sampled for yucca even if total yucca cover is < 10% aerial cover. Ferns, Xerophyllum, and Yucca have additional sampling criteria within Distributed plots: These three growth forms are not measured when the NLCD vegetation type of the plot is Deciduous Forest, Evergreen Forest, or Mixed Forest. In addition, when Distributed plot NLCD vegetation type is not forest, ferns are only sampled when aerial cover is ≥ 50% within the plot. Tower plots are sampled if at least one tree palm is present in ≥ 10% of tower plots, or if smaller qualifying non-herbaceous perennial individuals plus smaller qualifying woody vegetation comprise ≥ 10% of aerial cover averaged across all Tower plots. Within Distributed and Tower plots, tree palms are mapped and measured throughout the plot sampling area. If stem density thresholds are met, smaller qualifying individuals may be measured within nested subplots to standardize the sampling effort across plots (Figure 2).

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/

See TOS Science Design for Plant Biomass and Productivity (AD[07]), and the TOS Protocol and Procedure: Measurement of Vegetation Structure (AD[08]) for more details.

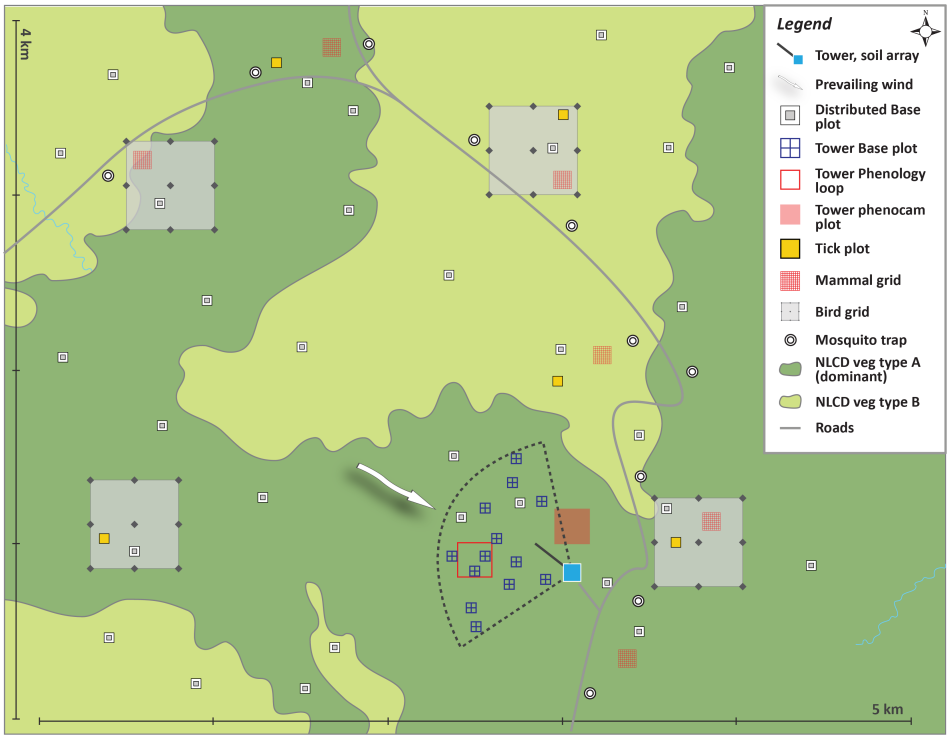

Figure 1: Generalized TOS sampling schematic, showing the placement of Distributed and Tower plots.

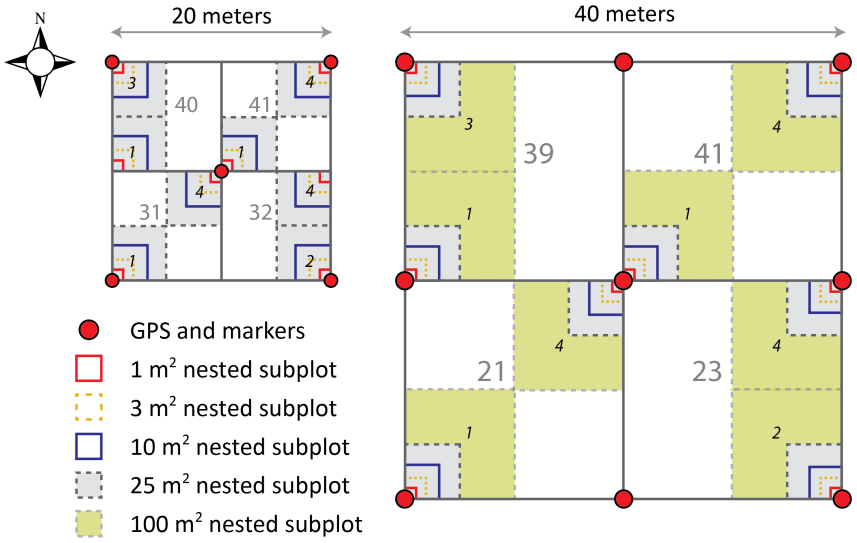

Figure 2: Illustration of a 20 m x 20 m Distributed or Tower base plot (left), a 40 m x 40 m Tower base plot (right), and associated nested subplots used for measuring woody and non-herbaceous perennial vegetation. Locations of subplots are denoted with plain text numbers, and locations of nested subplots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

3.2 Temporal Sampling Design

At all sites at which qualifying woody and non-herbaceous perennial vegetation is present, Distributed plots are sampled every 5 years. At most sites with qualifying vegetation, a spatially-balanced subset of n=5 Tower plots are sampled annually, and the full complement of Tower plots is sampled every 5 years. This temporal design allows annual growth increment in all Tower plots to be modeled based on data collected annually in the Tower plot subset. At slow-growth-increment sites (RMNP, YELL, NIWO, WREF, SJER, SOAP, TEAK, BONA, DEJU, HEAL), dendrometer bands are measured annually on a subset of individuals within tower plots. Trees are selected for banding in order to represent the biomass distribution of the population of trees within the tower airshed, with a bias toward selecting the largest trees. At desert sites with sensitive vegetation and soils (i.e., MOAB, JORN, SRER, ONAQ), there is no annual tower plot measurement and the full complement of tower plots is measured every 5 years only.

At all sites, Distributed plot sampling and sampling of the full complement of Tower plots is scheduled in a staggered manner such that sites generate data either from all Distributed plots or all Tower plots every 2-3 years. Distributed and Tower plots that do not have qualifying woody vegetation are surveyed every 5 years to determine whether in-growth of qualifying woody vegetation has occurred. In the event that surveys report newly qualifying woody or non-woody vegetation, vegetation structure sampling is scheduled for the year following the survey.

Woody and non-herbaceous perennial vegetation is sampled according to the same temporal design for most growth forms. However, ferns are only sampled in Tower plots when all Tower plots are sampled every 5 years. That is, ferns are not measured annually in the n=5 Tower plot subset.

Within a sampling season, it is very important that vegetation structure data are collected at phenologically consistent times across sites in order for cross-site data to be comparable. To determine phenologically consistent sampling dates across widely different ecosystems, the TOS utilizes MODIS-EVI phenology data for the majority of sites (Didan 2015), and more specifically, per site averages from the most recent 10 years that are updated every 5 years (e.g., Figure 3). Vegetation structure sampling is scheduled to begin no earlier at a given site than the date at which peak greenness begins to decrease. At sites with seasonal senescence, the actual onset of vegetation structure sampling in a given year is triggered by senescence of ≥ 50% of either deciduous canopy individuals or understory individuals depending on site vegetation. For the Domain 04 and Domain 20 tropical sites, MODIS-EVI is uninformative and vegetation structure sampling is scheduled to begin once the rainy season concludes. For most temperate sites, sampling must be completed before growth begins the following season or within 4 months of sampling onset, whichever is sooner. At sites with no distinct season, sampling begins within ± 2 weeks of the same date, and must be completed within 60 days of sampling onset.

See TOS Science Design for Plant Biomass and Productivity (AD[07]), and the TOS Protocol and Procedure: Measurement of Vegetation Structure (AD[08]) for more details.

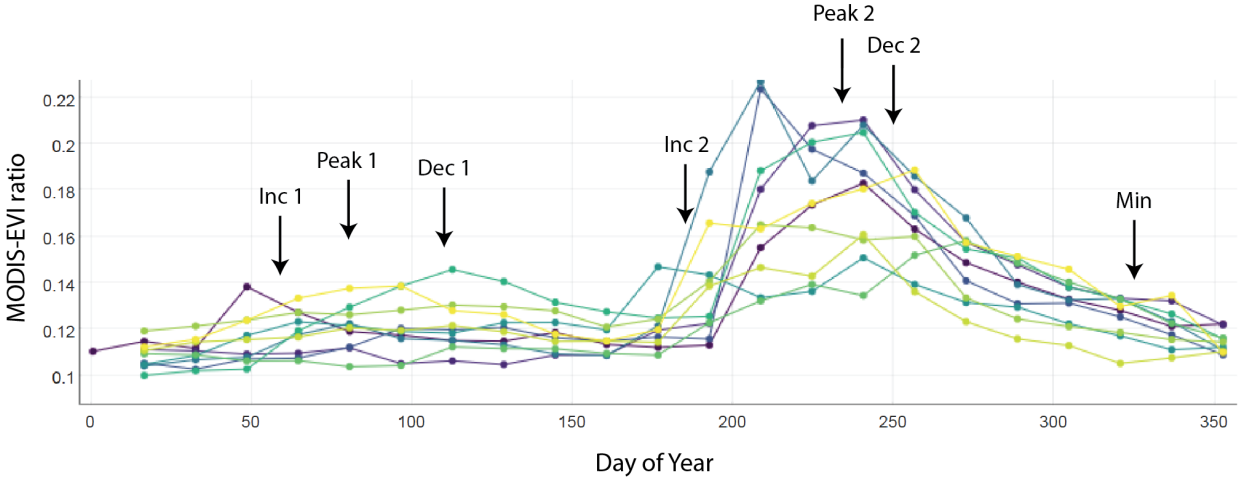

Figure 3: MODIS-EVI timecourse data for the Domain 14 Santa Rita Experimental Range site (D14 SRER). Data are from the years 2005-2014. There is an initial minor green-up early in the year centered around day 80, and another major green-up associated with the summer monsoon centered around day 240. Arrows indicate key greenness transitions that inform the onset of TOS sampling.

3.3 Sampling Design Changes

The NEON TOS has implemented several sampling design changes affecting the 'Woody plant vegetation structure' and 'Non-herbaceous perennial vegetation structure' data products over the course of data collection (Table 1). Such changes arise due to continual evaluation of the sampling design for best practices, both internally in collaboration with external technical working groups. Design changes also occur when optimization 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 changes for the NEON TOS 'Woody plant vegetation structure' and 'Non-herbaceous perennial vegetation structure' data products.

Change Date	Affected Sites	Change Summary Description
2017-03-09	MOAB, JORN, SRER, ONAQ	Prior to the change date, annual sampling of tower plots at desert sites caused unacceptable damage to sensitive soils and/or vegetation – e.g., soil biocrust damage, or incidents of broken vegetation when measuring brittle, multi-stem desert shrubs. To address this problem, annual sampling of Tower plots was discontinued at affected sites, and all Tower plots at these sites are now sampled on a multi-year time interval. The sampling interval was every 3-years between 2017-03-09 and 2018-07-26, and was changed to every 5-years after 2018-07-26.
2017-03-09	RMNP, YELL, NIWO, MOAB, SRER, JORN, ONAQ, WREF, SJER, SOAP, TEAK, BONA, DEJU, HEAL	Prior to the change date, low quality annual stem diameter increment data were observed from slow-growth-increment sites. Annual sampling of Tower plots was therefore discontinued at slow-growth increment sites; the sampling interval for Tower plots at affected sites was changed to every 3-years between 2017-03-09 and 2018-07-26, and was changed to every 5-years between 2018-07-26 and 2019-08-30.
2018-07-26	All sites for which woody and non-herbaceous perennial vegetation structure data have been published.	Prior to the change date, all Tower plots were measured annually at faster-growth sites, and at slow-growth increment sites, Tower plots were measured every 5 years; Distributed plots were measured every 5 years and in the same year that Tower plots were measured. To enable slow-growth-increment sites to generate data more frequently, to maintain efficient annual sampling at faster-growth sites, and to standardize the sampling schedule across sites, the following changes were made: 1) At all sites, sampling of Distributed plots is every 5 years; 2) At all sites, sampling of the full complement of Tower plots is every 5 years; 3) Sampling of Distributed and Tower plots is staggered such that a site generates data from all Distributed plots or all Tower plots every 2-3 years (see Section 4 in AD[08] for details); 4) At faster-growth sites, annual sampling of a spatially-balanced subset of Tower plots is scheduled (n=5 plots); faster-growth sites are all those not listed as slow-growth-increment sites immediately above.

Table 1: Summary of significant sampling design changes for the NEON TOS 'Woody plant vegetation structure' and 'Non-herbaceous perennial vegetation structure' data products.

Change Date	Affected Sites	Change Summary Description
2019-08-30	RMNP, YELL, NIWO, MOAB, SRER, JORN, ONAQ, WREF, SJER, SOAP, TEAK, BONA, DEJU, HEAL	Slow-growth increment sites lacked annual Tower plot data between 2017-03-09 and 2019-08-30, making calculation of annual above-ground net primary productivity at affected sites impossible. To enable annual ANPP estimation across the Observatory at all sites, NEON installed dendrometer bands on selected trees within a subset of Tower plots at slow-growth-increment sites. Banded trees are measured annually to enable estimation of ANPP.
2020-03-05	ABBY, BART, DELA, GRSM, HARV, LENO, MLBS, SCBI, SERC, STEI, TALL, TREE, UNDE, WREF	The sampling effort required to measure ferns annually in Tower plots was very high at forested sites with abundant fern cover relative to perceived fern contributions to plot-level biomass and productivity. Based on Technical Working Group feedback, NEON discontinued annual measurement of ferns and fern allies at sites where these growth forms are present in forested Tower plots. Ferns and fern allies are still sampled every 5-years in all Tower plots where they occur, and no changes were made to fern and fern ally sampling guidelines in Distributed plots.
2020-03-05	All sites for which woody vegetation structure data have been published.	Prior to the change, the vst_apparentindividual table lacked a mechanism to uniquely identify multiple boles for multi-bole small trees, saplings, single shrubs, and small shrub growth forms. Consequently, multiple qualifying boles for individuals with these growth forms appeared as duplicates based on the 'individualID'. To uniquely identify multiple qualifying boles for these individuals, the 'tempStemID' field was added to the protocol (AD[08], revJ). The value of the 'tempStemID' field increments within a multi-bole individual, and when combined with the 'individualID' and the 'eventID', provides a unique identifier. The value of 'tempStemID' assigned to a given bole within a multi-bole individual does not stay constant from year to year.

Table 1: Summary of significant sampling design changes for the NEON TOS 'Woody plant vegetation structure' and 'Non-herbaceous perennial vegetation structure' data products.

Change Date	Affected Sites	Change Summary Description
2020-03-05	DSNY, JERC, OSBS, KONZ	Prior to the change, routine fire management at affected sites resulted in re-tagging sapling and/or small shrub growth forms of certain species after every burn, due to root collar re-sprouts after above-ground parts of small individuals were killed by fire. For small individuals of affected species, re-tagging after fire resulted in increased labor costs and lack of desired longitudinal data for species prone to re-sprouting after fire. From the specified date, individuals of affected species are only tagged and assigned a permanent 'individualID' after they graduate to DBH = 1 cm (i.e., graduate into either the small-tree or single-shrub growth forms). Individuals with DBH < 1 cm are assigned recordType = 'temporary' in the vst_mappingandtagging table and given a corresponding temporary-type 'individualID'. See the Special Considerations section of this document and the protocol for more detail (AD[08], revJ). Affected taxa by site are: 1) DSNY and OSBS: Quercus spp., Lionia spp., Ilex glabra, Vaccinium spp., Diospyros virginiana, Rhus copallinum, Sassafras albidum, and Cratagus spp.; 2) JERC: Quercus spp., Vaccinium spp., Diospyros virginiana, Rhus spp., Sassafras albidum, Cratagus spp., and Prunus spp.; 3) KONZ: Cornus drummondii
2020-03-05	All sites for which non-herbaceous perennial vegetation structure data have been published for ferns.	Prior to the change, the nst_perindividual table lacked a mechanism to uniquely identify qualifying ferns that were measured but not tagged due to annual death of above-ground parts. Qualifying ferns of affected taxa (e.g., Pteridium aquilinum) therefore appeared as duplicates within a plotID for a given sampling event if taxonID and stemDiameter were identical for = 2 individuals in the plot. To create unique records within the nst_perindividual table, a temporary 'individualID' was created that is identical in structure to that used for small woody individuals affected by fire management in the vst_mappingandtagging table. See the Special Considerations section of this document and the protocol for more detail (AD[08], revJ).

3.4 Variables Reported

All variables reported during field collection (Level 0 data) are listed in the file, NEON Raw Data Validation for Vegetation Structure, Level 0 (DP0.10098.001) (AD[04]). All variables reported in the published data (Level 1 data) are also provided separately in the files NEON Data Variables for Woody Plant Vegetation Structure (DP1.10098.001) (AD[05]) and NEON Data Variables for Non-herbaceous Perennial Vegetation Structure (DP1.10045.001) (AD[06]).

Where applicable, 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). NEON TOS spatial data employs the World Geodetic System 1984 (WGS84) for its fundamental reference datum and GEOID09 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.

3.5 Spatial Resolution and Extent

Spatial resolution varies by growth form and is reported exhaustively by **individualID** in the vst_mappingandtagging table. Trees with DBH ≥ 10 cm diameter are always mapped as points wherever they occur (offsets from a reference pointID). Trees with DBH < 10 cm and shrubs are not typically mapped (see details in Spatial Sampling Design section above), but these individuals are identified according to the **subplotID** or **nestedSubplotID** in which they are located, 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 (identified via the **supportingStemID** field in the vst_apparentindividuals table). For growth forms that are not mapped as points, the finest spatial resolution is the level of either the subplot or the nested subplot if the latter is employed. Subplots are 100 m^2 in Distributed plots and small-stature Tower plots, and are 400 m^2 in large-stature Tower plots (Figure 2). Nested subplot size varies by plot depending on the density of individuals with DBH < 10 cm in a given plot. Available nested subplot sizes are:

- 1 m^2
- 3 m^2
- 10 m^2
- 25 m^2
- 100 m^2

Spatial hierarchy from finest to coarsest resolution=

point (optional) → **nestedSubplotID** (optional) → **subplotID** → **plotID** → **siteID** → **domainID**.

The basic spatial data included in the data download package correspond to the centroids of sampled plots, and include the latitude, longitude, and elevation of the plot centroid, plus associated uncertainty due to GPS error and plot width. Shapefiles of all NEON Terrestrial Observation System sampling locations can be found in the 'Spatial Data' section of the NEON Document Library (http://data.neonscience.org/documents). It is possible to obtain finer-grained spatial data for both mapped individuals and individuals that are not mapped.

3.5.1 Mapped individuals

To derive the mapped location of each individual tree in vst_mappingandtagging, download the `R geoNEON` package (https://github.com/NEONScience/NEON-geolocation) and use the `getLocTOS()` function.

Alternatively, use the steps below to perform the same calculation, using the high-resolution GPS data associated with the **pointID** from which each tree was mapped, plus **stemAzimuth** and **stemDistance** data for each tree in the vst_mappingandtagging table:

1. Construct the named location of the **pointID** associated with each record in vst_mappingandtagging by concatenating the plot-level named location with the pointID, like this: plotNamedLocation + '.' + pointID
    a. For example: For plotNamedLocation = 'BART_001.basePlot.vst' and pointID='21', the pointNamedLocation = 'BART_001.basePlot.vst.21'
2. To obtain easting and northing UTM data for the **pointID**, build a URL for the named location that is compatible with the NEON API (http://data.neonscience.org/api/v0/locations), and obtain elevation("locationElevation"), easting("locationUtmEasting"), northing("locationUtmNorthing"), coordinateUncertainty ("Value for Coordinate uncertainty"), elevationUncertainty ("Value for Elevation uncertainty"), and utmZone ("locationUtmZone").
    a. An example URL: http://data.neonscience.org/api/v0/locations/BART_001.basePlot.vst.21
3. Calculate the easting and northing in UTMs of each tree (i.e., each unique **individualID**) using **stemAzimuth** and **stemDistance**, and the easting and northing values for the **pointID** derived above. Easting and northing are derived according to equations (1) and (2) below.
4. To estimate horizontal and elevation uncertainty for the mapped tree location:
    a. Horizontal uncertainty: Increase the **coordinateUncertainty** value for the **pointID** by an appropriate amount (suggested 0.6 m) to account for error introduced by the laser rangefinder used for offset mapping.
    b. Elevation uncertainty: Increase the **elevationUncertainty** value for the **pointID** by an appropriate amount to account for topographical heterogeneity (suggested 1 m).

$$Easting = easting.pointID + d * \sin\theta \quad (1)$$

and

$$Northing = northing.pointID + d * \cos\theta \quad (2)$$

where,

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

$$easting.pointID = \quad (4)$$

the easting value of the reference pointID,

$$northing.pointID = \quad (5)$$

the northing value of the reference pointID,

$$d = \quad (6)$$

stemDistance

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 2 meters:
θ = 0.785,
easting = 315188.32 + 2 $\times$ sin (0.785) = 315189.7,
northing = 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

3.5.2 Unmapped individuals

For individuals that are not mapped, finer-grained spatial information may be derived by identifying the subplot or nested subplot in which the individual was recorded, then determining the coordinates of the subplot or nestedSubplot centroid. The **subplotID** or **nestedSubplotID** associated with individuals is recorded in the tables: vst_apparentindividual and nst_perindividual. For nested subplots, the location of the centroid depends on the area of the nested subplot; nested subplot area information is recorded in the vst_perplotperyear table. Use the steps below to calculate the location of the subplot or nested subplot centroid in which an unmapped individual occurs.

1. Join the vst_apparentindividual or nst_perindividual table with vst_perplotperyear by **plotID**. Select the appropriate **nestedSubplotArea** column from vst_perplotperyear depending on which table was joined. This new joined dataframe is referred to as `vst_join` in steps below.

    a. For vst_apparentindividual joins: Select **nestedSubplotAreaShrubSapling** to determine the area of nested subplot used for the sum of all single shrubs, small shrubs, small trees, and saplings within the plot. A value of 'noneSelected' means no nested subplot was used and the subplot centroid is the finest resolution available for these growth forms.
    b. For nst_perindividual joins: Select **nestedSubplotAreaOther** to determine the area of nested subplot used for the sum of all ferns, tree ferns, palms, ocotillo, xerophyllum, and yucca within the plot.

2. In the new `vst_join` data frame, for each individual identify the scale for which spatial data are available within the plot by creating a new variable by concatenating the **plotID**, **subplotID**, **nestedSubplotID**, and **nestedSubplotArea**. Do not concatenate values of `noneSelected` for **nestedSubplotArea** when no nested subplot was used. For example:

    a. `spatialScale = ABBY_001.31.1.10` (here, an unmapped individual was recorded in a nested subplot with area of 10 m^2)
    b. `spatialScale = ABBY_001.31` (here, an unmapped individual was recorded at the subplot scale and no nested subplot was used)

3. Retrieve the NEON TOS Spatial Data from the 'Spatial Data' section of the NEON Document Library (https://data.neonscience.org/documents). Download the file `All_NEON_TOS_Plots_VX.zip`, where 'X' is the version number.

    a. Extract the file: `All_NEON_TOS_Plot_Points_VX.csv` and create a data table.

4. In the `All_NEON_TOS_Plot_Points_VX` table:

    a. Create the same **spatialScale** variable as in `vst_join` by concatenating the **plotID** and **pointID** columns. Note that coordinates for nested subplots with area = 3 m^2, 25 m^2, and 100 m^2 are not currently available for download; anticipated availability is Q1 2022.
    b. Filter using the **appMods** column to retain only those rows that contain the "div" string.

5. Join `vst_join` with `All_NEON_TOS_Plot_Points_VX` by **spatialScale** and select the following columns from `All_NEON_TOS_Plot_Points_VX`:

    - latitude
    - longitude
    - datum
    - utmZone
    - easting
    - northing
    - horzUncert
    - elevation
    - vertUncert

6. The **easting** and **northing** values from `All_NEON_TOS_Plot_Points_VX` are the coordinates of the centroid for either the subplotID or nestedSubplotID (of the correct area) in which the unmapped individual is located. The **elevation** is the elevation of the subplot or nested subplot centroid.

7. The **horzUncert** and **vertUncert** values from `All_NEON_TOS_Plot_Points_VX` are the horizontal and vertical uncertainty for the subplot or nested subplot centroids, respectively.

3.6 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 are scheduled to be complete within 4 months of sampling onset, but may take longer in rare circumstances.

3.7 Associated Data Streams

The **individualID** (e.g., `NEON.PLA.DOM.SITE.######`) is a linking variable that ties vegetation structure measurements and metadata to the following associated data products:

- Plant phenology observations (DP1.10055.001),
- Plant foliar traits (DP1.10026.001), and
- Field spectral data (DP1.30012.001)

For the products listed above, data exist for a small subset of **individualIDs** reported in the Vegetation Structure data products.

For tagged woody individuals with DBH ≥ 10 cm that fall down and become part of the 'Coarse Downed Wood log survey' dataset (DP1.10010.001), 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 (DP1.10033.001), herbaceous clip harvests (DP1.10023.001), and coarse downed wood log surveys (DP1.10010.001) and coarse downed wood bulk density (DP1.10014.001) may be used together with vegetation structure measurements to estimate total Aboveground Net Primary Productivity. Remotely sensed Total biomass map - spectrometer - flightline (DP2.30016.001) and Total biomass map - spectrometer - mosaic (DP3.30016.001) and Ecosystem structure (DP3.30015.001) data products can also be used to upscale plot-based measurements of vegetation structure.

3.8 Product Instances

At typical NEON sites where qualifying woody and non-herbaceous perennial vegetation exists, sampling occurs annually at a subset of Tower plots, and all Tower plots are measured every 5 years; at sensitive desert sites there are no annual Tower plot measurements and all Tower plots are measured every 5 years. At all NEON sites with qualifying vegetation, Distributed plots are measured once every 5 years. The combination of these spatial and temporal factors gives rise to the following product instance information by table:

- **vst_perplotperyear**: One record per plot is expected per sampling event, and in years when a sampling event is scheduled, no more than one sampling event is expected per year. In other words, one record is created for each **plotID** visited at a site during a given **eventID**. Based on the sampling design above, it is expected that 5-30 vst_perplotperyear records are generated per site per year.
- **vst_mappingandtagging**: A minimum of one records exists for each tagged individual measured at any point during the lifetime of NEON. Following the initial tagging and mapping effort that occurs after plot establishment, the number of records created per site per year is variable and depends on recruitment - i.e., the number of newly qualifying individuals that must be tagged within the plots scheduled for sampling for a given **eventID**.
- **vst_apparentindividual**: Field staff create at least one record in the vst_apparentindividual table for each individual each time it is measured. Greater than one record per individual is created for small multi-stem trees, saplings, shrubs, and small shrubs - i.e., one record per qualifying stem per individual is created for multi-stem individuals. The number of expected records per site per year is therefore variable, and depends on the number of plots scheduled for sampling, the vegetation density within plots, and the number of stems associated with multi-stem individuals within the scheduled plots. Typically 300-3000 vst_apparentindividual records are created per site per year.
- **vst_shrubgroup**: A minimum of one record is expected per shrub group per sampling event for for shrub groups that occur within scheduled plots. In the event a shrub group is comprised of multiple taxa, additional records per shrub group are expected, one for each taxon making up the shrub group. The number of expected records per site per year is therefore variable, and depends on the number of plots scheduled for sampling, the density of the vegetation, and the number of taxa that typically occur within shrub groups. Typically < 100 vst_shrubgroup records are created per site per year.

- **nst_perindividual**: One record is expected for each measured individual each time a scheduled plot with qualifying non-herbaceous perennial vegetation is measured. The number of records per site per year is therefore variable, and depends on the number of plots scheduled for sampling, and the density of qualifying non-herbaceous perennial vegetation within plots. Typically 100-3000 nst_perindividual records are created per site per year.

3.9 Data Relationships

The 'Woody plant vegetation structure' and 'Non-herbaceous perennial vegetation structure' data products are comprised of five related tables:

- **vst_perplotperyear** → Data reported in this table define sampling **eventIDs** and provide sampling metadata for each sampled **plotID** that are critical to scaling up vegetation structure data to the plot- and site-level. Taken together, the **eventID** and **plotID** variables in each record are primary keys and link each vst_perplotperyear record to all other vegetation structure tables.
- **vst_mappingandtagging** → Each tagged woody and non-herbaceous perennial plant that is tagged is tracked in this table via the **individualID** field. Data that are invariant through time such as taxonID and spatial location are reported here, and the **eventID** and **individualID** fields are primary keys. Plot-level data associated with individuals is linked in the vst_perplotperyear table via the **plotID** and **eventID** fields, and the **individualID** is a linking variable to longitudinal data in the vst_apparentindividual table. The **recordType** field distinguishes between individuals that are tagged only, those that are mapped and tagged, and those that have dendrometer band data reported in the vst_apparentindividual table.
- **vst_apparentindividual** → Woody plant data that are expected to change through time are reported in this table (e.g., stem diameters). The **eventID**, **individualID**, and **tempStemID** variables are primary keys. Measurements from individuals are linked to plot-level data in the vst_perplotperyear table via the **eventID** and **plotID** variables. The **individualID** is a linking variable to invariant taxonID and spatial location data in the vst_mappingandtagging table.
- **vst_shrubgroup** → Data reported in this table describe the occurrence, structure, and taxonomic composition of shrub groups within a sampling event. Shrub groups are not mapped or tagged, and are assigned a temporary **groupID** that is unique within a given **plotID** x **eventID** combination.The **eventID**, **plotID**, **groupID**, and **taxonID** variables are primary keys. Measurements from shrub groups are linked to plot-level data in the vst_perplotperyear table via the **eventID** and **plotID** variables.
- **nst_perindividual** → Non-herbaceous perennial plant data that are expected to change through time are reported in this table (e.g., stem diameters). The **eventID** and **individualID** variables are primary keys. Measurements from individuals are linked to plot-level data in the vst_perplotperyear table via the **eventID** and **plotID** variables. For individuals with an **individualID** beginning with 'NEON' (e.g., tree ferns, palms, large-stature cacti), the **individualID** is a linking variable to invariant spatial location data in the vst_mappingandtagging table. However, there is

no link to the vst_mappingandtagging table for the majority of records with an **individualID** that begins with 'TEMP'.

Data downloaded from the NEON Data Portal are provided in separate data files for each site and month requested. The neonUtilities R package contains functions to merge these files across sites and months into a single file for each table described above. The 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. For instructions on using neonUtilities 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

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 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.

The master taxonomy list includes geographic range and nativity as described by the USDA PLANTS Database. A list for each NEON domain includes those species with ranges that overlap the domain as well as nativity designations - 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.

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.

The full master taxonomy lists are available on the NEON Data Portal for browsing and download: http://data.neonscience.org/static/taxon.html.

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 lists, 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 Vegetation Structure, Level 0 (DP0.10098.001), provided with every download of this data product. Contained within this file is a field named 'entryValidationRulesForm', which describes syntactically the validation rules for most fields that are 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]).

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

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 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]).

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

NEON Observation System quality flagging consists of product-specific quality observations entered routinely during data collection, and of post-hoc quality flags based on one or more observations or tests. Additionally, records of land management activities, disturbances, and other incidents of ecological note that may have a potential impact on data quality or interpretation are found in the Site Management and Event Reporting data product (DP1.10111.001).

5.4.1 QC Data Recorded in the Field

Situations that arise routinely in the field during the course of vegetation structure data collection and which affect data quality are documented via product-specific quality fields (Table 2).

Table 2: Field-collected vegetation structure data quality variables and descriptions.

Table Name	Field Name	Description
vst_apparentindividual	tagStatus	Description of state or condition of the physical tag on an individual. Values other than 'ok' indicate uncertainty in the continuity of the tagID through time.
vst_apparentindividual	changedMeasurement Location	Indicator for whether one or more stem diameter measurement locations on the individual changed from previous measurements. Values other than 'noChange' indicate a discontinuity in the measurement location and a disruption to stem increment time-series.
vst_apparentindividual	dendrometer Condition	Condition of the dendrometer band. Values other than 'ok' indicate uncertainty in the continuity of dendrometer gap measurements due to band disturbance, damage, re-setting, or replacement.
vst_apparentindividual vst_mappingandtagging vst_shrubgroup nst_perindividual	identification Qualifier	A standard term to express the determiner's doubts about the taxonomic identification.

See AD[08] for more detail on field-collected data quality variables.

5.4.2 Post-Hoc Quality Flagging

The **dataQF** field in each data record is a post-hoc quality flag for known errors that apply to the record. Explanations of **dataQF** codes specific to the 'Woody plant vegetation structure' and 'Non-herbaceous perennial vegetation structure' data products are provided in Table 3.

Table 3: Vegetation structure post-hoc data quality flagging codes and descriptions.

Field Name	Value	Description
dataQF	legacyData	Data recorded using a paper-based workflow that did not implement the full suite of front-end quality control features associated with the mobile data entry application workflow.
dataQF	noMapTagData	Measurements of individuals in the vst_apparentindividual table with no corresponding individual in the vst_mappingandtagging table; affected individuals may not have mapping data and/or taxonID data. Only applied to records collected before 2016.

Additional 'dataQF' Flag Details:

1. dataQF = `legacyData`: A non-exhaustive list of known issues associated with this flag includes:
    a. Diameter measurements may be missing for smaller woody growth forms (e.g., shrubs).
    b. 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).
    c. Out of range values may exist, due to incorrect units or other errors.
    d. Values for **growthForm** may be missing for measured individuals.
    e. 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.
    f. Some single- and multi-bole tree records do not have a value reported for **canopyPosition** and should.

6 SPECIAL CONSIDERATIONS

6.1 The vst_perplotperyear table

The vst_perplotperyear table contains sampling metadata necessary to make plot-scale and site-scale inference about stem density, biomass, productivity, volume, mean canopy height, and other derived variables from measured individuals and/or shrub groups. There are several 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 and to the site area:

1. For all growthForms within each plotID, field staff document whether individuals of that growth form are present anywhere within the plot for a given sampling event.
    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 and eventID.

   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 non-herbaceous perennial 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** field staff record nestedSubplotArea on a per growthForm basis when nestedSubplots are used, and `noneSelected` is recorded for each growthForm for which nestedSubplots were NOT used.

3. The group of **AbsentList** fields is used 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**

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. Finally, there is a group of fields that describes the nature of the scheduled sampling effort within and across plots. These fields enable appropriate use of the data when plot- and site-level estimation of response variables is desired.

    a. **samplingImpractical**: Provides insight into the completeness of the sampling effort. Records in vst_perplotperyear with a value in this field other than 'ok' indicate **plotIDs** that will not generate data for a given **eventID** for a variety of reasons. See AD[08] for more details.
    b. **eventType**: Indicates which plots were measured in a given sampling event according to the Temporal Sampling Design described in Section 3.6. For example, if an analysis requires data from Distributed plots or all Tower plots from multiple years, this field enables filtering to all events that include the desired plots.
    c. **dataCollected**: Values of `allGrowthForms` or `dendrometerOnly` indicate whether all individuals of all growth forms within the plot were measured, or whether only trees fitted with dendrometer bands were measured for a given **eventID**. When `dendrometerOnly` is selected it is not possible to scale response variables to the plot-scale for the selected **eventID**; instead, dendrometer gap data from banded trees within the plot may be used to model growth increment for the Tower airshed based on the banded trees that were measured.

6.2 The vst_mappingandtagging table

There are several important factors to consider in order to appropriately use data in the vst_mappingandtagging table.

1. Duplicate records based on **individualID** may exist. The combination of **eventID** and **individualID** are unique in this table, and when **individualID** duplicates are found, data users should always use the most recently created record. Duplicates based on **individualID** occur due to the following common scenarios:

    a. Field staff have updated the **taxonID** when errors in taxonomic identification were discovered and staff are certain which individuals were incorrectly identified and what the correct identification should be.
    b. Field staff have updated any or all of the **pointID**, **stemAzimuth**, or **stemDistance** fields that are used for mapping the location of qualifying individuals within the plot. Re-mapping may be required to correct data entry errors or to accommodate plot marker movement undertaken to correct plot establishment errors.

c. A dendrometer band has been installed on a tree, or an existing dendrometer band has been replaced due to damage.

2. For multi-bole trees, each qualifying bole is identified in this table with a unique **individualID**. The structure of the **individualID** indicates which related records are part of the same individual.

    a. *Example:* An **individualID** like `NEON.PLA.D05.STEI.02249` is reported for the largest bole and has a standard structure. Subsequent qualifying boles for this multi-bole individual have the structure `NEON.PLA.D05.STEI.02249A`, `NEON.PLA.D05.STEI.02249B`, etc.

3. At the KONZ, DSNY, JERC, and OSBS sites, frequent fire management results in individuals with DBH < 1 cm of certain species being consumed by fire before the next sampling event. For these individuals tagging is impractical and records are created in vst_mappingandtagging with **recordType** = `temporary` in order to publish **taxonID** data for these individuals. When **recordType** = `temporary`, the **individualID** is like `TEMP.PLA.SITE.YYYY.NN###`, where:

    a. `SITE` = the standard NEON 4-letter siteID
    b. `YYYY` = the year the sampling event began
    c. `NN` = the numerical portion of the **plotID**, and
    d. `###` = an incrementing numerical identifier generated by field staff that is unique within a plot for a given sampling event.
    e. 'Temporary' individualIDs do NOT correspond to the same individual from one sampling event to the next.

6.3 The vst_apparentindividual table

The vst_apparentindividual table reports longitudinal vegetation structure data collected from qualifying individuals. Important points to consider when using these data include:

1. For multi-bole trees (individuals with DBH ≥ 10 cm for at least one bole), and multi-bole small trees, saplings, shrubs, and small shrubs, more than one record exists in this table for a given sampling event. Note that:

    a. Multi-bole trees have **individualIDs** that persist through time for each bole with DBH ≥ 10 cm, and these related **individualIDs** are discerned by a trailing alpha-character. For example: Largest bole = `NEON.PLA.D17.SJER.00047`; second bole = `NEON.PLA.D17.SJER.00047A`; third bole = `NEON.PLA.D17.SJER.00047B`, etc. Crown dimension, height, canopy position and other variables recorded at the level of the individual are recorded for the largest bole only. The **tempStemID** is auto-assigned to each bole but is not relevant for multi-bole tree individuals due to the uniquely assigned per bole **individualID**.
    b. Multi-bole small trees, saplings, shrubs, and small shrubs have duplicate **individualIDs** within a sampling event, and crown dimension, height, canopy position and other variables recorded at the level of the individual are recorded for the largest bole only. The **tempStemID** variable increments sequentially for each qualifying bole, and when combined with the **individualID**, uniquely identifies the bole within a sampling event. The **tempStemID** does not correspond to the same bole from year-to-year.

2. The **growthForm** assigned to an individual in the field primarily determines which measurements are reported (Table 4). Secondarily, the **plantStatus** affects required measurements - e.g., additional measurements are needed for broken individuals, or measurements are not required for individuals that were previously measured but cannot be relocated in the current sampling event.

Table 4: Summary of woody vegetation structure data by growth form in the vst_apparentindividual table. Note that individuals may change growth form from one year to the next, potentially necessitating different measurements from year-to-year (e.g., a change from 'sapling' to 'small tree').

Growth Form	Diameter	Height	Crown Diameter	Other
Liana	• stemDiameter (DBH) • measurementHeight	Not measured	Not measured	• plantStatus
Single-bole tree	• stemDiameter (DBH) • measurementHeight	Per individual	• maxCrownDiameter • ninetyCrownDiameter	• plantStatus • canopyPosition
Multi-bole tree	• stemDiameter (DBH, per bole) • measurementHeight (per bole)	Per individual (largest bole), not per bole	• Per individual (largest bole), not per bole • maxCrownDiameter • ninetyCrownDiameter	• plantStatus (per bole) • canopyPosition (largest bole, not per bole)
Small tree	• stemDiameter (DBH, per bole) • measurementHeight (per bole) • One record per qualifying bole	Per individual (largest bole), not per bole	• Per individual (largest bole), not per bole • maxCrownDiameter • ninetyCrownDiameter	• plantStatus (per bole) • canopyPosition (largest bole, not per bole)
Sapling	• basalStemDiameter (ddh, per bole) • basalMeasurementHeight (per bole) • One record per qualifying bole	Per individual (largest bole), not per bole	Not measured	• plantStatus (per bole) • canopyPosition (largest bole, not per bole)

Table 4: Summary of woody vegetation structure data by growth form in the vst_apparentindividual table. Note that individuals may change growth form from one year to the next, potentially necessitating different measurements from year-to-year (e.g., a change from 'sapling' to 'small tree').

Growth Form	Diameter	Height	Crown Diameter	Other
Single shrub	• basalStemDiameter and stemDiameter (ddh and DBH per bole; DBH from largest diameter fork only) • measurementHeight and basalMeasurementHeight (per bole) • One record per qualifying bole	Per individual (largest bole), not per bole	• Per individual (largest bole), not per bole • maxCrownDiameter • ninetyCrownDiameter	• plantStatus (per bole) • canopyPosition (largest bole, not per bole) • If shape = inverted cone: max and ninety baseCrownDiameter reported • If shape = ellipse: height to crown base reported
Small shrub	• basalStemDiameter (ddh, per bole) • basalMeasurementHeight (per bole) • One record per qualifying bole	Per individual (largest bole), not per bole	• Per individual (largest bole), not per bole • maxCrownDiameter • ninetyCrownDiameter	• plantStatus (per bole) • canopyPosition (largest bole, not per bole) • If shape = inverted cone: max and ninety baseCrownDiameter reported • If shape = ellipse: height to crown base reported

See AD[08] for more detail on growth form dependent measurement requirements for woody individuals.

6.4 The nst_perindividual table

The nst_perindividual table reports vegetation structure data collected from qualifying non-herbaceous perennial individuals. Important points to consider when using these data include:

1. The majority of individuals are not tagged (e.g., ferns), and are assigned 'temporary'-type **individualIDs** with structure like `TEMP.PLA.SITE.YYYY.NN###`. Components of temporary non-woody **individualIDs** are identical to those defined above for temporary woody individuals in the vst_mappingandtagging table.
    a. Temporary **individualID** values are not unique from year-to-year.
2. A subset of non-herbaceous perennial individuals are tagged and are associated with **individualIDs** with the standard structure shown above in the vst_mappingandtagging section - e.g., `NEON.PLA.D14.SRER.00047`
    a. Individuals are typically tagged because mapping and longitudinal data collection for the individual is possible.
    b. Tagged individuals usually have growthForm = cactus, palm, or tree fern.
3. The **growthForm** assigned to an individual in the field primarily determines which measurements are reported (Table 5). Secondarily, the **plantStatus** affects required measurements - e.g., measurements are not required for individuals that were previously measured but cannot be relocated in the current sampling event.

Table 5: Summary of non-herbaceous perennial vegetation structure data by growth form in the nst_perindividual table.

Growth Form	Diameter	Height	Crown Diameter	Other
Fern	*Pteridium aquilinum* only	Not measured	Not measured	For species other than *P. aquilinum*: • leafNumber • meanLeafLength
Tree fern	• stemDiameter or basalStemDiameter, depending on height • measurementHeight	Per individual	Not measured	• stemLength • canopyPosition
Palms	• stemDiameter or basalStemDiameter, depending on height (tree palms only) • measurementHeight (tree palms only)	Per individual	Per individual; shrub palms only (i.e., *Leucothrinax morrissii, Serenoa repens, Sabal etonia*)	Tree palms: • stemLength • canopyPosition Shrub palms: • leafNumber • meanPetioleLength • meanBladeLength • canopyPosition

Table 5: Summary of non-herbaceous perennial vegetation structure data by growth form in the nst_perindividual table.

Growth Form	Diameter	Height	Crown Diameter	Other
Ocotillo	Not measured	Per individual	Basal crown diameter, per individual	• basal stem count • canopyPosition
Bear grass	• basalStemDiameter (per flower stalk) • measurementHeight	Not measured	Not measured	• meanLeafLength
Yucca (includes agave)	Not measured	Per individual	Per individual	

See AD[08] for more detail on growth form dependent measurement requirements for non-herbaceous perennial individuals.

7 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]).

8 REFERENCES

Didan, K. 2015. MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V006. NASA EOSDIS Land Processes DAAC. https://doi.org/10.5067/MODIS/MOD13Q1.006.

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.

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.