NEON USER GUIDE TO THE VEGETATION STRUCTURE DATA PRODUCT (DP1.10098.001)

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

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.

REVISION	DATE	DESCRIPTION OF CHANGE
D	02/10/2023	Significant changes by section: • Throughout document: Updated table names to reflect bundling of woody and non-woody tables into one consolidated Vegetation structure data product (DP1.10098); Non-herbaceous perennial vegetation structure product (DP1.10045) is deprecated. • Throughout document: Added new ’palm tree’, ’small palm’, ’large tree fern’, ’small tree fern’ growth forms; the ’palm’ and ’tree fern’ growth forms are now deprecated. New growth forms align nested subplot sampling strategy with that used for trees, small trees, and saplings. • Section 3.1 and 3.2: Updated Spatial Sampling Design and Temporal Sampling Design to reflect the current sampling design as documented in protocol version M. • Section 3.3: Added change entries for new nestedSubplotAreaFerns and totalSampledAreaFerns, and nonwoodyCollectDate fields, as well as new palm tree, small palm, large tree fern, and small tree fern growth forms. • Section 3.5: Added new 50 m2 nestedSubplot area for use in 40m x 40m Tower plots. • Section 6: Revised Special Considerations vst_perplotperyear subsection to reflect removal of Absence Lists, and to more clearly document which growth forms are associated with nestedSubplotAreaShrubSapling and nestedSubplotAreaOther field values.
D.1	12/10/2023	Added Identification History documentation: • Section 3.9 Data Relationships: Added description of new vst_identificationHistory table. • Section 4 Taxonomy: Added Section 4.1 Identification History to document how changes to taxonomy are captured in the data.

REVISION	DATE	DESCRIPTION OF CHANGE
E	03/04/2024	Significant changes by section: • Section 3: Added updated subplotID naming convention information throughout section, including new Figure 2. • Section 3.5: Updated instructions for using geoNEON functions to obtain spatial location data for mapped and unmapped individuals. • Section 3.9: Clarified that total sampled area fields in vst_perplotperyear are required to accurately scale up data from different growth forms when nested subplots are used. Added that individualID in vst_mappingandtagging is a linking variable with vst_non-woody for tagged tree palms and tree ferns. • Section 6.3: Updated to describe how vst_apparentindividual subplotID field can be used to locate centroid of measurement area for unmapped woody individuals. • Section 6.4: Updated to describe how vst_non-woody subplotID field can be used to locate centroid of measurement area for unmapped non-woody individuals.
E.1	11/25/2024	Added multiPlotDuplicate and measurementError QFs: • Section 5.4.2 Data QF: Added description of new vst data QFs.
F	03/27/2025	Added information about the new neonutilities Python package.
F.1	01/06/2026	Updated to reflect removal of eventID, subplotID, and nestedSubplotID fields from the vst_mappingandtagging table. Numerous minor edits throughout the document to improve clarity.
G	03/23/2026	Added documentation of 2026 sampling modifications that impact the temporal sampling design. Updated information about sampling location histories and added information in the automated data processing steps section.

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 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:

- 'Vegetation structure' (DP1.10098) - Field measurements from woody and non-woody perennial plants that include: height, crown diameter, stem diameters (breast height and/or basal), other growth form specific measurements, taxonID, as well as mapped position of qualifying woody and non-woody individuals. Non-woody perennial plants include: cacti, ferns, ocotillo, palms, tree ferns, xerophyllum, and yucca.
- 'Non-herbaceous perennial vegetation structure' (DP1.10045) - From NEON Data Release 2022 and onward, this product is deprecated and the 'non-woody' table formerly associated with this product is now published with the 'Vegetation Structure' product (DP1.10098). Because NEON Data Releases are stable, this data product persists in Releases prior to 2022.

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., '10098') as the corresponding Level 1 data product.

2 RELATED DOCUMENTS AND ACRONYMS

2.1 Associated Documents

AD[01]	NEON.DOC.000001	NEON Observatory Design
AD[02]	NEON.DOC.000913	TOS Science Design for Spatial Sampling
AD[03]	NEON.DOC.002652	NEON 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]	Available on NEON data portal	NEON's Ingest Conversion Language (NICL) specifications
AD[13]	NEON.DOC.005424	NEON Algorithm Theoretical Basis Document: OS Data Quality Control

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

The NEON Terrestrial Observation System (TOS) 'Vegetation Structure' data product includes information related to structure, taxonomy, spatial location, and biomass of woody and non-herbaceous perennial vegetation. These vegetation types include tree, sapling, shrub, liana, cactus, palm, fern, tree 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 taxonID, 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], Barnett *et al.* 2019, Theobald *et al.* 2007). According to this design, plots are allocated across the landscape in proportion to the area of dominant National Land Cover Database (NLCD) cover types (AD[02], Fry *et al.* 2011). Within the tower airshed, a spatially-balanced randomized sampling design is employed (Barnett *et al.* 2019).

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 all 10 m x 10 m subplots are sampled. 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 made up of four 10 m x 10 m subplots. Regardless of size, all NEON TOS plots are established from 10 m x 10 m base units (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 make up ≥ 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 sampled within nested subplots 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 palm tree or large tree fern is present. If palm trees or large tree ferns 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 palm tree or large tree fern 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, palm trees and large tree ferns are mapped and measured throughout the plot sampling area. If stem density thresholds are met, smaller qualifying individuals may be sampled within nested subplots to standardize the sampling effort across plots (Figure 2).

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

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.

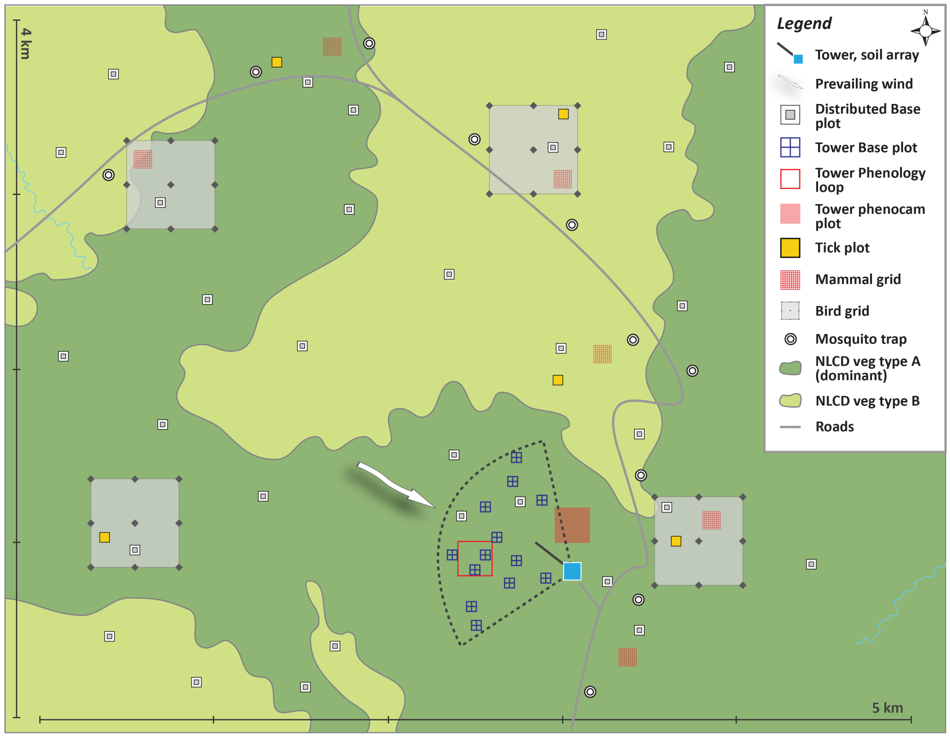

Figure 1: Generalized TOS sampling schematic, showing the placement of Distributed and Tower plots.

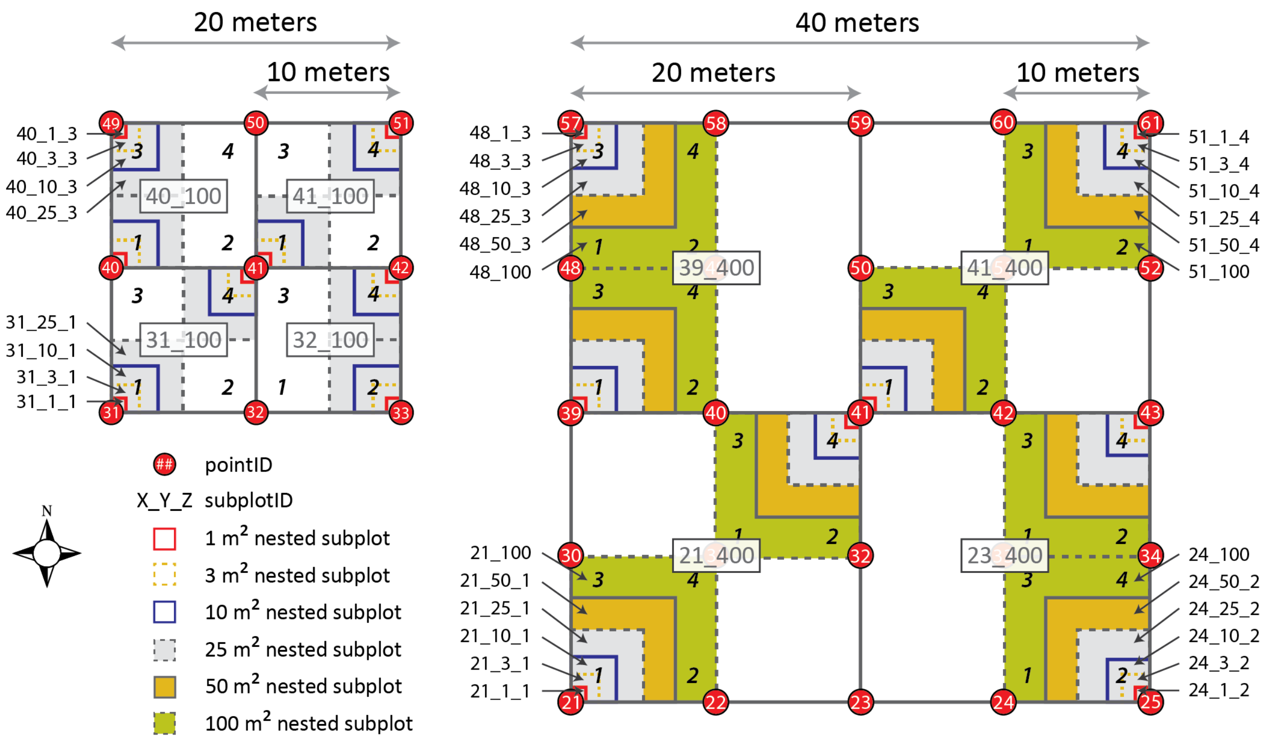

Figure 2: Illustration of a 20 m x 20 m Distributed or Tower base plot (left), a 40 m x 40 m Tower base plot (right), and associated subplots and nested subplots used for measuring woody and non-herbaceous perennial vegetation. Subplots that are 100 m^2 and larger are labeled like ’XX_YYY’, where ’XX’ denotes the pointID that anchors the southwest corner and ’YYY’ indicates the area in meters squared. Nested subplots smaller than 100 m^2 are identified like ’XX_YY_Z’, where the additional ’Z’ component designates the corner (black italics) in which the nested subplot is established within the 10 m x 10 m base unit.

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 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 sampling and the full complement of Tower plots is sampled 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 sampled 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 so that cross-site data are comparable. To determine phenologically consistent sampling dates across widely different ecosystems, the TOS utilizes MODIS-EVI phenology data for the majority of sites (Didan 2021), 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 relatively invariant throughout the year, and precipitation data are therefore used to guide the vegetation structure sampling dates. In these two domains, vegetation structure sampling is scheduled to begin once the rainy season concludes. For all sites, sampling must be completed before growth begins the following season or within 4 months of sampling onset, whichever is sooner.

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.

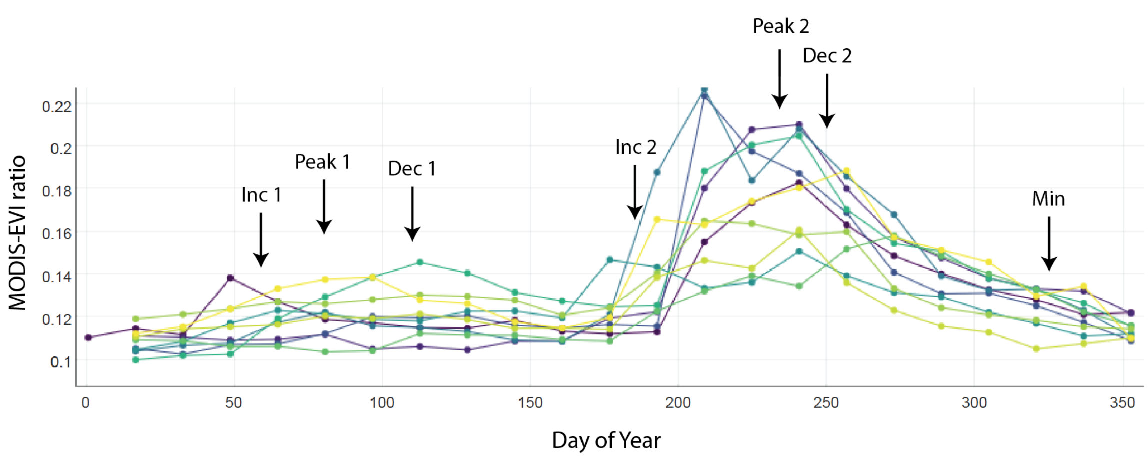

Figure 3: MODIS-EVI time series data for the Domain 14 Santa Rita Experimental Range NEON site (D14 SRER), 2005-2014. There is an initial minor green-up early in the year centered around day 80, and another major green-up associated with the summer monsoon centered around day 240. Colored traces represent data from different years; ‘Inc’ = average date at which EVI greenness increases; ‘Peak’ = average date of EVI greenness peak; ‘Dec’ = average date at which EVI decreases; the ‘1’ and ‘2’ denote the two distinct green-up cycles at the site.

3.3 Sampling Design Changes

The NEON TOS has implemented several sampling design changes affecting the 'Vegetation structure' data product over the course of data collection (Table 2). Such changes arise due to continual evaluation of the sampling design for best practices, both internally and in collaboration with external technical working groups. Changes to protocols and data products also occur when NEON improves data product usability and structure, and when optimization of sampling designs is necessary to ensure that allocation of sampling effort is poised to maximize returns to the scientific community.

Table 2: Summary of significant sampling design and data product changes for the NEON TOS 'Vegetation structure' data product.

Change Date	Affected Sites	Change Summary Description
2017-03-09	MOAB, JORN, SRER, ONAQ	Tower plots are sampled on a multi-year interval instead of annually at affected sites after 2017-03-09. Following this change, productivity must be calculated as a multi-year average rather than annually. The sampling interval was increased from annual to every 5 years due to slow growth rates and to mitigate damage to sensitive soils and vegetation.
2017-03-09	RMNP, YELL, NIWO, WREF, SJER, SOAP, TEAK, BONA, DEJU, HEAL	The Tower plot sampling interval was changed from annual to every 5 years at affected sites due to low-quality stem diameter increment data resulting from small annual growth increments. After 2019-08-30, NEON installed dendrometer bands at affected sites on a subset of trees in Tower plots and banded trees are measured annually. Data users may use annual dendrometer band data to model annual productivity across all Tower plots, or productivity may be calculated as a multi-year average using data from all Tower plots sampled every 5 years.
2018-07-26	All sites for which woody and non-woody vegetation structure data have been published.	At all sites: 1) Sampling of Distributed plots is every 5 years (was every 3 years); 2) Sampling of the full complement of Tower plots is every 5 years (was every 3 years); 3) Sampling of Distributed and all 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) A spatially-balanced subset of Tower plots (n=5) is sampled annually except for at MOAB, JORN, SRER, and ONAQ (see above). This change means data users can no longer expect to obtain Distributed plot and Tower plot data from a site for the same year, and the sampling interval is longer for all plots.

Change Date	Affected Sites	Change Summary Description
2020-03-05	ABBY, BART, DELA, GRSM, HARV, LENO, MLBS, SCBI, SERC, STEI, TALL, TREE, UNDE, WREF	NEON discontinued measurement of ferns and fern allies in the annually sampled Tower plot subset at forested sites where these growth forms are present. 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. This change means fern productivity must be calculated as a multi-year average rathern than annually.
2020-03-05	All sites for which woody vegetation structure data have been published.	To uniquely identify multiple qualifying boles for small tree, sapling, single shrub, and small shrub growth forms, the 'tempStemID' field was added to the vst_apparentindividual table. 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. This change reduces the number of apparent duplicates: Prior to adding 'tempStemID', multiple qualifying boles with the same 'individualID' within affected growth forms appeared as duplicates based on the 'individualID' alone.
2020-03-05	DSNY, JERC, OSBS, KONZ	Individuals of affected species (listed below by site) are only tagged and assigned a permanent 'individualID' after they graduate to DBH ≥ 1 cm. Individuals of affected species with DBH < 1 cm are assigned recordType = 'temporary' in the vst_mappingandtagging table and given a temporary-type 'individualID'. This change was made because routine fire management at affected sites resulted in a lack of longitudinal data for species prone to re-sprouting after fire. Following the change, data users cannot calculate annual or multi-year average productivity for individuals with DBH < 1 cm on a per 'individualID' basis for affected species. 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

Change Date	Affected Sites	Change Summary Description
2020-03-05	All sites for which vegetation structure data have been published for ferns.	To create unique records within the vst_non-woody 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 above). For fern species for which stemDiameter is measured (e.g., Pteridium aquilinum), prior to the change it was possible for multiple individuals with the same plotID, taxonID, and stemDiameter to incorrectly appear as duplicates. See the Special Considerations section of this document and the protocol for more detail (AD[08]).
2022-03-01	All sites for which non-woody vegetation structure data have been published.	After the specified date, woody and non-woody perennial vegetation structure data are published as part of the same unified 'Vegetation structure' data product (DP1.10098), and the 'Non-herbaceous perennial vegetation structure' data product (DP1.10045) is deprecated (though it will continue to exist in past NEON data releases). Splitting the data into two products made the non-herbaceous perennial vegetation structure product harder to use and less discoverable. Following the change, downloads of the 'Vegetation structure' data product will include the vst_non-woody table by default.
2023-03-01	All sites for which vegetation structure data have been published for ferns.	New 'nestedSubplotAreaFerns' and 'totalSampledAreaFerns' variables were created in the vst_perplotperyear table. The 'totalSampledAreaFerns' value is only populated when ferns are scheduled for sampling, and this variable remains NULL when ferns are not sampled for a given Tower plot bout. This change enables more accurate assessment of plot-level biomass for ferns. Prior to the change, ferns and additional 'other' growth forms were considered together when reporting the 'nestedSubplotAreaOther' and 'totalSampledAreaOther' values in the vst_perplotperyear table, potentially creating confusion for data users because a sampling area was reported for ferns even when ferns were not scheduled for sampling. In addition to the new fields introduced for ferns, the vst_perplotperyear 'fernsPresent' column now has 'present - sampled', 'present - sampling criteria not met', and 'not present' options, to clarify when downstream records for ferns are expected for a given plotID in the vst_non-woody table.

Change Date	Affected Sites	Change Summary Description
2023-03-01	DSNY, GUAN, JERC, OSBS	The 'palm tree' and 'small palm' growth forms were introduced in the vst_non-woody table, replacing the deprecated 'palm' growth form. Individuals classified as 'palm tree' are always sampled throughout the plot wherever they occur, whereas 'small palm' individuals may be sampled using nested subplots. When nested subplots are used to sample 'small palm' individuals, the 'totalSampledAreaOther' field in the vst_perplotperyear table must be used to correctly calculate plot-level biomass of small palms. The plot-level biomass of 'palm tree' individuals is calculated analogously to trees using the 'totalSampledAreaTrees' field in the vst_perplotperyear table. Prior to the change, it was not clear which palm individuals were sampled within a nested subplot and which were not when a nestedSubplotAreaOther value was reported in the vst_perplotperyear table. Data edits were made to assign all previously measured palm individuals to the correct new 'palm tree' or 'small palm' growth form.
2023-03-01	PUUM	The 'large tree fern' and 'small tree fern' growth forms were introduced in the vst_non-woody table, replacing the deprecated 'tree fern' growth form. Individuals classified as 'large tree fern' are always sampled throughout the plot wherever they occur, whereas 'small tree fern' individuals may be sampled using nested subplots. When nested subplots are used to sample 'small tree fern' individuals, the 'totalSampledAreaOther' field in the vst_perplotperyear table must be used to correctly calculate plot-level biomass of small tree ferns. The plot-level biomass of 'large tree fern' individuals is calculated analogously to trees using the 'totalSampledAreaTrees' field in the vst_perplotperyear table. Prior to the change, it was not clear which tree fern individuals were sampled within a nested subplot and which were not when a nestedSubplotAreaOther value was reported in the vst_perplotperyear table. Data edits were made to assign all previously measured tree fern individuals to the correct new 'large tree fern' or 'small tree fern' growth form.

Change Date	Affected Sites	Change Summary Description
2023-03-01	All sites for which ferns or pad-forming Opuntia spp. are sampled at an earlier date than woody vegetation and other non-woody vegetation.	To more accurately describe multiple sampling dates that are nonetheless still part of the same Vegetation Structure 'eventID', the vst_perplotperyear table now contains a 'nonwoodyCollectDate' field. When a plot is sampled earlier in the season for ferns or pad-forming Opuntia individuals, the sampling date is recorded in the new 'nonwoodyCollectDate' field; the sampling date associated with woody and other non-woody individuals is reported in the 'date' field. Prior to the change, a single sampling 'date' was reported in the vst_perplotperyear table every time a plot was sampled, regardless of whether multiple visits were made to the plot at different times of year to sample different growth forms.
2023-03-01	All sites	Multiple growth-form-specific 'AbsenceList' fields are no longer published in the vst_perplotperyear table. The 'AbsenceList' fields were intended to document which plots or subplots were surveyed but contained no individuals of the specified growth forms. However, accurate assessment of presence/absence by growth form at the time of vst_perplotperyear record creation in the field was not consistently possible. Moving forward, the presence or absence of downstream records in the vst_apparentindividual and vst_non-woody tables provides analogous information.
2024-01-29	All sites	For woody individuals, the 'subplotID' is now published in the vst_apparentindividual table rather than the vst_mappingandtagging table. In addition, the subplotID structure was changed to always contain the point that anchors the southwest corner as well as the subplot area in meters squared (see Figure 2). For subplots smaller than 100 m2, the subplotID also contains corner information that was previously reported in the 'nestedSubplotID' field in the vst_mappingandtagging table. The 'subplotID' in all past data have been edited to be consistent with the new subplotID structure. When nested subplots are used to sample smaller woody and non-woody growth forms, this change means that when the subplotID is used as a grouping variable the spatial scale of the individuals grouped together is clearly specified. Moreover, subplotID centroid spatial data are now available via the NEON API.

Change Date	Affected Sites	Change Summary Description
2026-01-06	All sites	The 'eventID', 'subplotID', and 'nestedSubplotID' fields are no longer published in the vst_mappingandtagging table. The 'eventID' was removed because it is not a relevant grouping variable for the vst_mappingandtagging table. The 'subplotID' and 'nestedSubplotID' fields were removed because the 'subplotID' variable introduced to the vst_apparentindividual table in 2024 now contains the information from these fields.
2026-01-29	All sites	Data collection for those Distributed plots and Tower plots sampled on a 5-year interval was suspended in 2026, leading to a 6-year interval between sampling for affected plots rather than the expected 5-year interval. The subset of Tower plots sampled annually (n=5), as well as annual sampling of dendrobanded trees, is unaffected by this change and will be carried out in 2026 as usual. The sites with Distributed and Tower plots sampled on a 5-year interval that were originally scheduled for 2026 will be sampled in 2027, and the schedule originally intended for 2027 will be implemented in 2028 and so forth, leading to a consistent 6-year interval between sampling events for affected plots across all sites in the schedule until 2032.

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 Geoid12A geoid model for its vertical reference surface. Latitudes and longitudes are denoted in decimal notation to six decimal places, with longitudes indicated as negative west of the Greenwich meridian.

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

3.5 Spatial Resolution and Extent

Trees with DBH ≥ 10 cm diameter, as well as palm trees and large tree ferns with DBH ≥ 10 cm and stem-Length > 130 cm, are always mapped wherever they occur. Trees are mapped by recording offset data (**stemAzimuth** and **stemDistance**) relative to a nearby **pointID** (i.e., plot marker) that is associated with high-resolution GPS data; offset mapping data are reported in the vst_mappingandtagging table and **pointID** coordinates are available via the NEON API. Consult the 'Mapped Individuals' section below for information on how to calculate the mapped location for trees. Lianas are not mapped, but it is possible to infer the approximate position of a liana from the identity of the host tree (identified via the **supportingStemID** field in the vst_apparentindividual table).

Small trees (DBH < 10 cm), saplings, shrubs, small palms, small tree ferns, cacti, yucca, and additional 'other' growth forms are not typically mapped (see details in Spatial Sampling Design section above). However, unmapped individuals are always identified according to the **subplotID** in which they are located, which is reported in either the vst_apparentindividual or vst_non-woody table, depending on the growth form. Spatial data associated with subplotID centroids are available via the NEON API. Consult the 'Unmapped Individuals' section below for information on how to retrieve subplotID centroids.

For 20 m x 20 m Distributed plots and small-stature Tower plots, available subplot sizes are (Figure 2):

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

For 40 m x 40 m large-stature Tower plots, there are additional subplot sizes:

- 50 m^2
- 400 m^2

In all TOS plots, the naming convention for subplots consists of the pointID in the southwest corner of the subplot, the scale or size of the subplot in meters squared, and for those nested subplots smaller than 100 m^2, the corner of the 100 m^2 subplot in which the smaller nested subplot is located. For example, '21_400' refers to a subplot with pointID 21 in the southwest corner and an area of 400 m^2 (20 m x 20 m). Alternatively, '31_1_1' is a subplot with pointID 31 in the southwest corner, an area of 1 m^2, and is located in corner 1 of the 100 m^2 subplot with pointID 31 in the southwest corner (Figure 2). The subplot or nested subplot size selected for sampling varies by plot depending on the density of individuals with DBH < 10 cm in a given plot.

Spatial hierarchy from finest to coarsest resolution=

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

The spatial data included in the vst_perplotperyear table correspond to the centroids of sampled **plotIDs** (i.e., pointID 41 in Figure 2), and include the latitude, longitude, and elevation of the plot centroid, plus associated uncertainty due to GPS error and plot width.

3.5.1 Mapped individuals

To derive the location of mapped individual trees reported in the vst_mappingandtagging table, download the `R geoNEON` package (https://github.com/NEONScience/NEON-geolocation) and use the `geoNEON::getLocTOS()` function with the following arguments:

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

Alternatively, use the steps below to perform the same calculation, 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='31', the pointNamedLocation = 'BART_006.basePlot.vst.31'
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"). If the location data includes more than one record, use the coordinates corresponding to the sampling date.
   a. An example URL: http://data.neonscience.org/api/v0/locations/BART_006.basePlot.vst.31?history=true
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 484.6 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 in which the individual is recorded, then determining the coordinates of the subplot centroid. The **subplotID** associated with individuals is provided in the vst_apparentindividual, vst_non-woody, and vst_shrubgroup tables. To find the subplot centroid coordinates, download the `R` `geoNEON` package (https://github.com/NEONScience/NEON-geolocation/geoNEON) and use the `geoNEON::getLocByName()` function with the following steps:

1. Construct the named location of the **subplotID** associated with each record in the input data frame (e.g., the vst_apparentindividual table) by concatenating the plot-level **namedLocation** with the **subplotID** like this: namedLocation + '.' + subplotID
   a. For example: For namedLocation = 'BART_001.basePlot.vst' and subplotID='31_100', the subplotNamedLocation = 'BART_001.basePlot.vst.31_400'
2. Use the `geoNEON::getLocByName()` function with the following arguments:
   - `data = df`, where 'df' is a data frame containing subplot named locations as described immediately above.
   - `locCol = [name of column containing subplot named location identifiers]`, e.g., **subplotNamedLocation** in the example above.
   - `token = NEON_TOKEN`, where 'NEON_TOKEN' is an optional token generated via a NEON user account. While tokens are not required, they provide enhanced download speeds.
   - *NOTE*: For data collected prior to 2024, some **subplotID** values may have an "unknown" component and `getLocByName()` will not return spatial data for these subplotIDs.

Alternatively, if not using `geoNEON::getLocByName()`, manually retrieve spatial data for **subplotID** centroids using the NEON API:

1. Construct the named location of the **subplotID** as above.

2. To obtain coordinates for the **subplotID**, build a URL for the subplot named location that is compatible with the NEON API (http://data.neonscience.org/api/v0/locations), and obtain easting ("locationUtmEasting"), northing ("locationUtmNorthing"), coordinateUncertainty ("Value for Coordinate uncertainty"), elevation ("locationElevation"), elevationUncertainty ("Value for Elevation uncertainty"), utmZone ("locationUtmZone"), and additional spatial data values. If the location data includes more than one record, use the coordinates corresponding to the sampling date.

   - An example URL: http://data.neonscience.org/api/v0/locations/BART_001.basePlot.vst.31_100?history=true
   - *NOTE*: As above, **subplotID** named location URLs with an "unknown" component will not return spatial data via the API.

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)

A small subset of the **individualIDs** reported in the 'Vegetation structure' data product also have data in the products listed above.

For tagged woody individuals with DBH ≥ 10 cm that fall down and become part of the 'Coarse Downed Wood log survey' dataset (DP1.10010), the **vstTagID** variable in the cdw_fieldtally table 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), herbaceous clip harvests (DP1.10023), and coarse downed wood log surveys (DP1.10010) and coarse downed wood bulk density (DP1.10014) may be used together with vegetation structure data to estimate total Aboveground Net Primary Productivity. Remotely sensed Total biomass map - spectrometer - flight-line (DP2.30016) and Total biomass map - spectrometer - mosaic (DP3.30016) and Ecosystem structure (DP3.30015) 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 (typically n=5), and all Tower plots are sampled every 5 years; at sensitive desert sites there is no annual Tower plot sampling and all Tower plots are sampled every 5 years. At all NEON sites with qualifying vegetation, Distributed plots are sampled 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 record 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_non-woody**: One record is expected for each measured individual each time a scheduled plot with qualifying non-herbaceous perennial vegetation is sampled. 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 vst_non-woody records are created per site per year.

- **vst_shrubgroup**: A minimum of one record is expected per shrub group per sampling event. In the event a shrub group comprises 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.

3.9 Data Relationships

The ‘Vegetation structure’ data product comprises six related tables:

- **vst_perplotperyear** → Data reported in this table define sampling **eventIDs** and provide **totalSampledArea** metadata by growth form for each sampled **plotID** that are critical to scaling up vegetation structure data to the plot- and site-level when nested subplots are used. 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 except vst_mappingandtagging. For the latter, the **plotID** is the only linking variable.
- **vst_mappingandtagging** → Each 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 offset mapping data are reported here, and the **date** and **individualID** fields are primary keys. Plot-level data associated with individuals is linked in the vst_perplotperyear table via the **plotID** field, and the **individualID** is a linking variable to longitudinal data in both the vst_apparentindividual table (woody individuals) and the vst_non-woody table (e.g., tagged cacti, tree ferns, tree palms, yucca, etc). 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, height, etc). 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 offset mapping data in the vst_mappingandtagging table.
- **vst_non-woody** → 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., large tree ferns, small tree ferns, palm trees, small palms, large-stature cacti, some yucca), the **individualID** is a linking variable to invariant offset mapping 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’ (e.g., ferns are not tracked through time).

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

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

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. 2023. 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 areas including the Lower 48 U.S. States, Alaska, Hawaii, and Puerto Rico.

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.

4.1 Identification History

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

5 DATA QUALITY

5.1 Data Entry Constraint and Validation

Many quality control measures are implemented at the point of data entry within a mobile data entry application or web user interface (UI). For example, data formats are constrained and data values controlled through the provision of dropdown 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 or web user interface, the steps used to process the data through to publication on the NEON Data Portal are detailed in the NEON Algorithm Theoretical Basis Document: OS Generic Transitions (AD[11]).

Published data are reviewed for completeness, timeliness, and validity using an internal set of tests and metrics, as detailed in the NEON Algorithm Theoretical Basis Document: OS Data Quality Control (AD[13]). These quality tests are used to guide process improvements, audits of analytical facilities, and data updates, but do not generate quality flags in published data.

5.3 Data Revision

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

5.4 Quality Flagging

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).

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 3).

Table 3: 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.

Table 3: Field-collected vegetation structure data quality variables and descriptions.

Table Name	Field Name	Description
vst_apparentindividual	changedMeasurement Location	Indicates whether one or more stem diameter measurement locations on the individual changed from a previous measurement location. 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_mappingandtagging vst_non-woody vst_shrubgroup	identification Qualifier	A standard term to express the determiner's doubts about the taxonomic identification.
vst_apparentindividual	heightQualifier	Indicates common situations that introduce uncertainty into reported height data.

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 'Vegetation structure' data product are provided in Table 4.

Table 4: 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.
dataQF	multiPlotDuplicate	IndividualID is present in different plots, resolutions to correct location or tagID are pending further investigation by field teams.
dataQF	measurementError	Height and stemDiameter values unavailable

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**
      - **treeFernsPresent**
      - **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 not sampled). No further documentation is provided for growthForms absent at the plot scale. If a growthForm is recorded as 'present - sampled', downstream records are expected in the vst_apparentindividuals or vst_non-woody tables.

c. The **fernsPresent** field is unique in that 'present - sampling criteria not met' is an option. This option is selected when ferns may be present in a plot but ferns do not meet %cover requirements for sampling, or ferns are present but are not scheduled for sampling based on the eventType (e.g., ferns are not sampled when eventType = 'towerSubset').
d. The **palmsPresent** field indicates 'palm tree' and/or 'small palm' individuals are present in the plot.
e. The **treeFernsPresent** field indicates 'large tree fern' and/or 'small tree fern' individuals are present in the plot.

2. For 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**; nested subplot area used for the sum of 'single shrub' + 'small shrub' + 'small tree' + 'sapling' growth forms.
      - **nestedSubplotAreaLiana**
      - **nestedSubplotAreaFerns**
      - **nestedSubplotAreaOther**; nested subplot area used for the sum of 'cacti' + 'yuccas' + 'small palms' + 'small tree ferns' + 'ocotillo' + 'xerophyllum' growth forms
   b. 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 **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 include:
      - **totalSampledAreaTrees**
      - **totalSampledAreaShrubSapling**
      - **totalSampledAreaLiana**
      - **totalSampledAreaFerns**, and
      - **totalSampledAreaOther**
4. 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 sampled in a given **eventID** 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 scheduled 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 **even-**

tID; 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 to appropriately use data in the vst_mappingandtagging table.

1. Duplicate records based on **individualID** may exist. The combination of **date** and **individualID** is unique in this table, and when **individualID** duplicates are found, data users should always select 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` to enable publication of **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 dimensions, 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 unique **individualID** that is assigned per bole.
b. Multi-bole small trees, saplings, shrubs, and small shrubs have duplicate **individualIDs** within a sampling event, and crown dimensions, 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. Within an individual, the **tempStemID** does not correspond to the same bole from year-to-year.

2. For woody individuals that are not mapped, the **subplotID** can be used to obtain coordinates for the centroid of the subplot in which the individual was measured for a given **eventID** (see Section 3.5.2 for step-by-step details). For example, when un-mapped saplings within a plot are sampled using 10 m^2 nested subplots, an individual with subplotID = '40_10_3' can be located to the area bounded by the 10 m^2 box in the upper-left corner of the plot (Figure 2).
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., additional measurements are needed for broken individuals, or measurements are not required for individuals that were previously sampled but cannot be relocated in the current sampling event.

Table 5: 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)

Growth Form	**Diameter**	**Height**	**Crown Diameter**	**Other**
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)
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 (i.e., baseCrownHeight)

Growth Form	Diameter	Height	Crown Diameter	Other
Small shrub	• basalStemDiameter (ddh, per bole) • basalMeasurement-Height (per bole) • One record per qualify-ing 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 baseCrown-Diameter reported • If shape = ellipse: height to crown base reported (i.e., baseCrownHeight)

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

6.4 The vst_non-woody table

The vst_non-woody 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 cannot be used to track an individual's growth 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 tree, small palm, large tree fern, or small tree fern.
3. For the majority of non-woody individuals that are not mapped, the **subplotID** can be used to obtain coordinates for the centroid of the subplot in which the individual was measured for a given **eventID** (see Section 3.5.2 for step-by-step details). For example, when un-mapped ferns within

a plot are sampled using 3 m^2 nested subplots, an individual with subplotID = '24_3_2' can be located to the area bounded by the 3 m^2 box in the lower-right corner of the plot (Figure 2).

4. The **growthForm** assigned to an individual in the field primarily determines which measurements are reported (Table 6). For cactus, the **shape** field additionally informs required measurements.

Table 6: Summary of non-herbaceous perennial vegetation structure data by growth form in the vst_non-woody table.

Growth Form	**Diameter**	**Height**	**Crown Diameter**	**Other**
Fern	*Pteridium aquilinum* only	Not measured	Not measured	For species other than *P. aquilinum*: • leafNumber • meanLeafLength
Large tree fern, small tree fern	• stemDiameter or basalStemDiameter, depending on stemLength • measurementHeight	Per individual	Not measured	• stemLength • canopyPosition
Palm tree, small palm	• stemDiameter or basalStemDiameter, depending on stemLength (*Arecaceae spp.* only) • measurementHeight (*Arecaceae spp.* only)	Per individual	Per individual; shrub palms only (i.e., *Leucothrinax morrissii, Serenoa repens, Sabal etonia*)	Tree palms (*Arecaceae spp.*): • stemLength • canopyPosition Shrub palms: • leafNumber • meanPetioleLength • meanBladeLength • canopyPosition
Ocotillo	Not measured	Per individual	Basal crown diameter, per individual	• basal stem count • canopyPosition
Bear grass	• basalStemDiameter (i.e., average tussock diameter at the base) • 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

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

Didan, K. MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V061. 2021. NASA EOSDIS Land Processes DAAC. https://doi.org/10.5067/MODIS/MOD13Q1.061

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. 2023. The PLANTS Database (http://plants.usda.gov, 10 February 2023). National Plant Data Team, Greensboro, NC 27401-4901 USA.