



# NEON USER GUIDE TO FIELD SPECTRAL DATA (DP1.30012.001)

PREPARED BY	ORGANIZATION
Samantha Weintraub-Leff	TOS
Bridget Haas	AOP
Alok Shrestha	AOP
Kate Murphy	AOP



## CHANGE RECORD

REVISION	DATE	DESCRIPTION OF CHANGE
A	12/18/2019	Initial Release
B	05/26/2020	Included general statement about usage of neonUtilities R package and statement about possible location changes. Updated taxonomy information.
C	04/08/2022	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
D	04/17/2025	Included a statement addressing pre-standardized collection bouts and how to address samples of the same species and sample location. Added information about the new neonUtilities Python package.



## TABLE OF CONTENTS

<b>1</b>	<b>DESCRIPTION</b>	<b>1</b>
1.1	Purpose . . . . .	1
1.2	Scope . . . . .	1
<b>2</b>	<b>RELATED DOCUMENTS AND ACRONYMS</b>	<b>2</b>
2.1	Associated Documents . . . . .	2
2.2	Acronyms . . . . .	2
<b>3</b>	<b>DATA PRODUCT DESCRIPTION</b>	<b>3</b>
3.1	Spatial Sampling Design . . . . .	3
3.2	Temporal Sampling Design . . . . .	5
3.3	Theory of Measurements . . . . .	5
3.4	Variables Reported . . . . .	6
3.5	Spatial Resolution and Extent . . . . .	6
3.6	Temporal Resolution and Extent . . . . .	7
3.7	Associated Data Streams . . . . .	7
3.8	Product Instances . . . . .	7
3.9	Data Relationships . . . . .	8
3.10	Special Considerations . . . . .	9
<b>4</b>	<b>TAXONOMY</b>	<b>9</b>
<b>5</b>	<b>DATA QUALITY</b>	<b>9</b>
5.1	Data Entry Constraint and Validation . . . . .	9
5.2	Automated Data Processing Steps . . . . .	10
5.3	Data Revision . . . . .	10
5.4	Quality Flagging . . . . .	10
<b>6</b>	<b>REFERENCES</b>	<b>11</b>

## LIST OF TABLES AND FIGURES

Table 1	Descriptions of the dataQF codes for quality flagging . . . . .	10
Figure 1	Representation of NEON base plots used for field spectral sampling . . . . .	4
Figure 2	Example reflectance curve of sunlit Douglas Fir needles from the Wind River Experimental Forest . . . . .	5



# 1 DESCRIPTION

## 1.1 Purpose

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

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

## 1.2 Scope

This document describes the steps needed to generate the L1 data product Field spectral data - handheld imaging spectroradiometer measurements from plant foliage samples collected at NEON sites - and associated metadata. This document also provides details relevant to the publication of the data products via the NEON data portal, with additional detail available in the file, NEON Data Variables for Field spectral data (DP1.30012.001) (AD[04]), 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 AOP Field Spectra Protocol (AD[06]). The raw data that are processed in this document are detailed in the file, NEON Raw Data Validation for Field spectral data (DP0.30012.001) (AD[03]), provided in the download package for this data product. Please note that raw data products (denoted by 'DP0') may not always have the same numbers (e.g., '10033') as the corresponding L1 data product.



## 2 RELATED DOCUMENTS AND ACRONYMS

### 2.1 Associated Documents

AD[01]	NEON.DOC.000001	NEON Observatory Design (NOD) Requirements
AD[02]	NEON.DOC.002652	NEON Products Catalog
AD[03]	Available with data download	Validation csv
AD[04]	Available with data download	Variables csv
AD[05]	NEON.DOC.000913	TOS Science Design for Spatial Sampling
AD[06]	NEON.DOC.005214	AOP Field Spectra Protocol
AD[07]	NEON.DOC.001024	TOS Protocol and Procedure: Canopy Foliage Sampling
AD[08]	NEON.DOC.000008	NEON Acronym List
AD[09]	NEON.DOC.000243	NEON Glossary of Terms
AD[10]	NEON.DOC.004825	NEON Algorithm Theoretical Basis Document: OS Generic Transitions
AD[11]	Available on NEON data portal	NEON Ingest Conversion Language Function Library
AD[12]	Available on NEON data portal	NEON Ingest Conversion Language
AD[13]	Available with data download	Categorical Codes csv

### 2.2 Acronyms

Acronym	Definition
AOP	Airborne Observation Platform
ASD	Analytical Spectral Devices, Inc.
GPS	Global Positioning System
TOS	Terrestrial Observation System
SWIR	Short-Wave InfraRed
VNIR	Visible and Near-InfraRed



### 3 DATA PRODUCT DESCRIPTION

The Field spectra data product provides measurements of foliar spectral reflectance and transmittance taken with a field spectroradiometer. Field spectral campaigns are coordinated with data collection for TOS Protocol and Procedure: Canopy Foliage Sampling (AD[07]). Most often, the foliage chosen for reflectance measurements is a subsample of the material collected as part of sunlit canopy foliage sampling. Such samples have a host of physical and chemical measurements associated with them, including leaf mass per area (LMA), moisture content, and concentrations of carbon, nitrogen, lignin, chlorophyll, and major/minor elements. These data are delivered as part of NEON Data Product Plant foliar physical and chemical properties (DP1.10026.001). However, sometimes spectral measurements are taken on independent samples, not included in the general foliage sampling protocol. These ‘stand-alone’ samples will not be associated with any other measurements besides those delivered in the field spectra data product.

Field spectral data are collected in conjunction with overflights of the NEON Airborne Observation Platform (AOP), which conducts remote sensing of ecosystem chemical and physical characteristics using hyperspectral and LiDAR measurements. Ground-based spectral measurements are important because, when used in conjunction with AOP spectral data, they can improve our understanding of the contribution of ecosystem structure to remotely sensed reflectance. Field reflectance and transmittance curves can also serve as spectral end-members, which can be used to carry out supervised hyperspectral classification using the AOP Spectrometer L1 and L3 orthorectified surface directional reflectance data products (NEON.DOM.SITE.DP1.30006.001, NEON.DOM.SITE.DP3.30006.001) and similar applications using data products derived from the reflectance data.

Spectral samples that are associated with canopy foliage samples will be exclusively from sunlit environments, whereas stand-alone spectral samples may come from both shaded and sunlit positions. The exposure of the sample is recorded in the sample metadata table, along with other useful contextual information such as maturity and position of leaf or needle material and arrangement of the material for spectrometer readings. The EcoSIS Spectral Library (<https://ecosis.org/>) was used to inform selection of metadata variables.

#### 3.1 Spatial Sampling Design

Field spectral data collection occurs at a subset of sites performing canopy foliage sampling in a given year. Forested and savannah sites are currently the focus, though grassland sites may be added later.

Within a site, samples collected for both canopy foliage and field spectral measurements come from a subset of TOS base plots located within the NEON tower airshed (Tower plots) as well as those distributed across the landscape (Distributed plots, Figure 1). While the TOS Science Design for Spatial Sampling (AD[05]) emphasises a randomized, stratified approach for most protocols, canopy foliage and field spectra efforts do not follow this design. Since the goal is to take sunlit samples from as many species and habitat types as possible and cover the range of variation visible in the AOP imagery, technicians have the freedom to sample from all NEON plots designated for vegetation structure measurements. Sampling can occur anywhere within a NEON plot except for the nested subplots set aside for plant diversity monitoring. Stand-alone field spectral measurements may also be taken from outside of NEON plots, as long as they are within NEON site boundaries and collections adhere to site host restrictions.



In data collected prior to 2019, the AOP field teams were more opportunistic in sample selection. In early years, there were often repeat collections from the same individuals, and samples were not always collected from individuals used for canopy foliage or vegetation structure measurements. Without identification of plants from vegetation structure mapping, it is inconclusive if samples of the same species and location are due to repeat samples of the same individual or if the GPS collection device stayed in the same location while the samples were collected from a variety of individuals. For samples collected in 2019 and later, sampling is coordinated with canopy foliar sampling, and individual plants are uniquely identified by the individualID field.

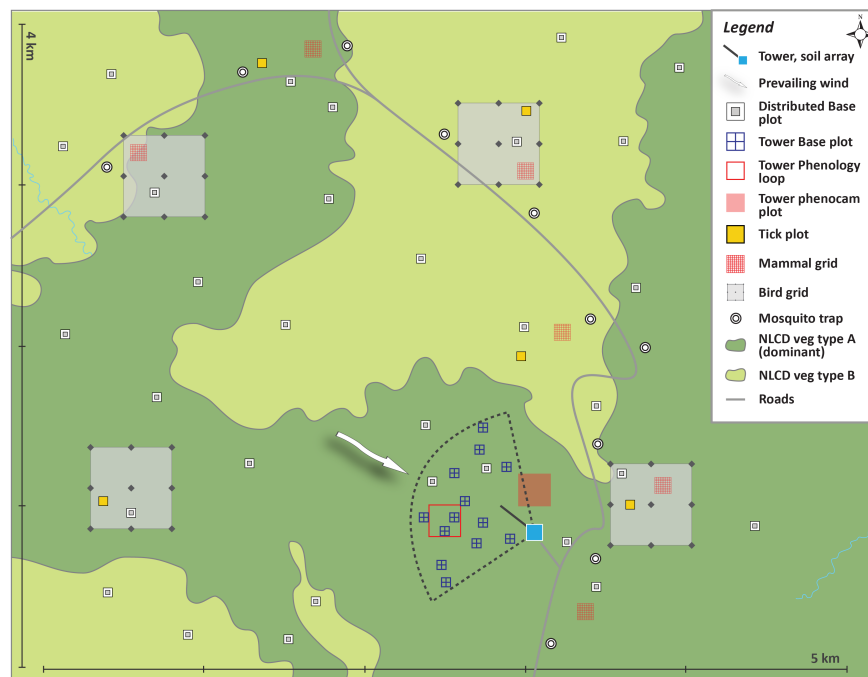


Figure 1: Representation of NEON base plots used for field spectral sampling

All field spectral measurements are georeferenced. For samples collected for canopy foliage, this is accomplished in two ways - an x,y location for the stem of each sample is provided, and starting in 2020 a shapefile of the sampled crown is also created whenever possible. Both of these can be accessed by downloading the Plant foliar physical and chemical properties data product (DP1.10026). For stand-alone field spectral samples, georeferencing is either provided as latitude/longitude of the plot centroid (for samples taken inside a NEON plot), or latitude/longitude  $\pm$  uncertainty provided from a field tablet (for samples taken outside of NEON plots).

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



### 3.2 Temporal Sampling Design

Field spectral data are collected opportunistically as a level-of-effort data product. This often means that data are collected at one of the forested or savannah sites where TOS Protocol and Procedure: Canopy Foliage Sampling is being conducted in a given year. However, some years there are no field spectral measurements taken, and other years more than one site may be sampled. All sampling campaigns for a given site-year combination occur over a 1-2 week period, coincident with the AOP overflight and the site's historical period of peak greenness. In general, foliage is measured several minutes after it is harvested, though the collection time is recorded in the data in case of deviation.

### 3.3 Theory of Measurements

The reflectance of a foliar tissue, which varies with wavelength, is the ratio of energy reflected from a sample to the energy incident on the sample. Similarly, transmittance, also varying with wavelength, is the ratio of radiant energy transmitted through a sample to the radiant energy incident on a sample. These spectral properties are determined by various biophysical properties of the foliage, including pigments (for example, chlorophyll), cell structure and composition, and water content (Figure 2). To date, NEON has used an ASD to FieldSpec 3 to collect

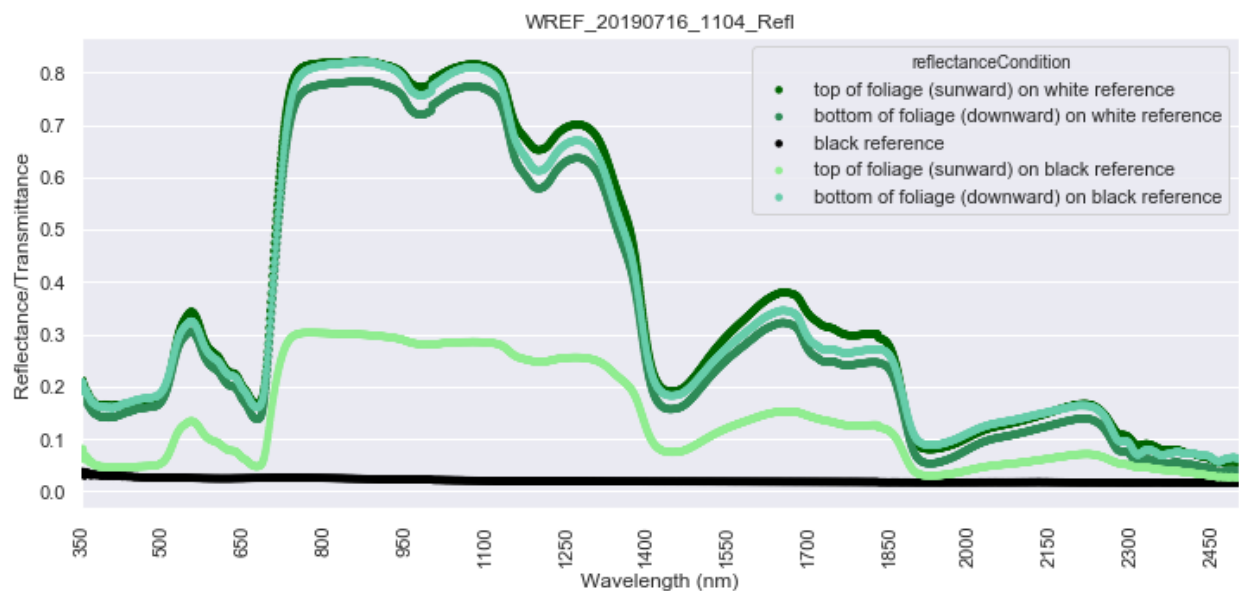


Figure 2: Example reflectance curve of sunlit Douglas Fir needles from the Wind River Experimental Forest

field spectral measurements of foliar reflectance and transmittance. The spectroradiometer measures the optical energy or electromagnetic radiation that is reflected by, absorbed into, or transmitted through a foliar sample. The radiation measured includes wavelengths ranging from visible to SWIR (360-2500 nm). Radiation is measured using a fiber optic cable, which transmits optical energy into the instrument. Wavelength components are separated and independently measured by a set of detectors for the





VNIR range (360-1000nm) and by two different detectors for the SWIR regions (1000-1830nm and 1830-2500nm).

The raw measurement made by the spectroradiometer is a combination of both the sample and the light source. An independent measure of the light source illumination on a reference of known reflectance is required to calculate the sample reflectance (or transmittance). For all measurements, a white reference standard with near 100% reflectance across the entire spectrum is used for measuring and removing the contribution of the light source. In order to provide a complete suite of spectral measurements for each sample, spectra of the front and back of the foliage sample over white and black reference standards are collected. Between each collection, a white reference panel is measured, which is used to calculate the final measurements.

From raw reflectance measurements taken in the field, a set of equations are used to create five sets of calibrated spectral reflectance measurements for each sample, for each wavelength. The reflectanceCondition field in the spectral data files differentiate these measurements from each other:

1. Spectral reflectance of the front of sample: top of foliage (sunward) on white reference
2. Spectral reflectance - back of sample: bottom of foliage (downward) on white reference
3. Spectral reflectance - black reference
4. Spectral transmittance - front of sample: top of foliage (sunward) on black reference
5. Spectral transmittance - back of sample: bottom of foliage (downward) on black reference

For further technical details related to the instrumentation and equations, see the Algorithm Theoretical Basis Document (ATBD): Field Spectral Data.

### 3.4 Variables Reported

All variables reported from the field or laboratory (L0 data) are listed in the file, NEON Raw Data Validation for Field spectral data (DP0.30012.001) (AD[04]). All variables reported in the published data (L1 data) are also provided separately in the file, NEON Data Variables for Field spectral data (DP1.30012.001) (AD[05]).

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

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

### 3.5 Spatial Resolution and Extent

The finest resolution at which spatial data are reported is a sample, taken from a plant located within or outside of a plot within a NEON site.



**sampleID** (unique ID given to the sample) → **plotID** (ID of plot within site) → **siteID** (ID of NEON site) → **domainID** (ID of a NEON domain)

For samples taken within plots, the basic spatial data included in the data download include the latitude, longitude, and elevation of the *centroid* of the plot where sampling occurred + associated uncertainty due to GPS error and plot width. Shapefiles of all NEON Terrestrial Observation System sampling locations can be found on the NEON science webpage at <https://www.neonscience.org/data-samples/data/spatial-data-maps>. For samples taken outside of plots, the basic spatial data included in the data download include the 'alt' latitude and 'alt' longitude of the location where sampling occurred + associated uncertainty due to GPS error (often much larger for outside of plot sampling).

For field spectral samples that are also measured for canopy foliage, it is possible to derive a more precise estimate of sample geolocation. There are two options:

- For samples collected in 2020 and onward, download the crown shapefiles (.shp) associated with canopy foliar collections. These can be accessed using the 'expanded' download package option. Extract geolocation info using a GIS or programming environment such as R.
- Calculate the latitude and longitude of the *stems* of sampled individuals using offset information provided in the canopy foliage data product. In the `vst_mappingandtagging` table, filter to only the **individualIDs** that appear in the field spectral metadata table, then use the `getLocTOS` function from the `geoNEON` package, available here: <https://github.com/NEONScience/NEON-geolocation>. Or, perform the same offset calculations by hand using the step by step instructions provided in the Canopy Foliage Data Product User Guide, available in the NEON data portal document library.

### 3.6 Temporal Resolution and Extent

The finest resolution at which temporal data are reported is the spectral **measurementDate**, which includes time in UTC.

The NEON Data Portal provides data in monthly files for query and download efficiency. Queries including any part of a month will return data from the entire month. Code to stack files across months is available on GitHub (<https://github.com/NEONScience/NEON-utilities>), CRAN (<https://cran.r-project.org/web/packages/neonUtilities/index.html>) or Python Package Index (PyPi; <https://pypi.org/project/neonutilities/>).

### 3.7 Associated Data Streams

**sampleID** is the linking variable that ties specific samples to the Plant foliar physical and chemical properties data product (DP1.10026.001). When provided, the **individualID** can be used to find additional information about sampled woody vegetation (e.g., tree dbh, height, etc) by downloading the NEON Woody Plant Vegetation Structure data product (DP1.10098.001) and joining to the `vst_apparentindividuals` table.

### 3.8 Product Instances

There are likely to be 20-50 field spectral measurements taken per site, and anywhere from zero to several sites sampled per year.



### 3.9 Data Relationships

The protocol dictates that before starting to take field spectral measurements for a given site-year campaign, a bout-level metadata record should be populated to report instrument parameters. This will yield a unique **eventID** in the fsp\_boutMetadata table, and this same identifier will appear in the sample-level metadata table to link samples with those instrument parameters. All samples collected for the purpose of taking a field spectral measurement are assigned a **sampleID** in the fsp\_sampleMetadata table. If they are also collected for canopy foliar physical and chemical measurements, this sampleID can be found in the tables contained in the Plant foliar physical and chemical properties data product.

Each sampleID maybe have one or more **spectralSampleIDs** associated with it, depending on how many unique leaves or needles are scanned. Each spectralSampleID is expected to appear once in the fsp\_spectralData table, including a link to download the calibrated reflectance data from NEON's cloud storage system. In the spectral files, each spectralSampleID will appear thousands of times as it is a long-format table. Duplicates and/or missing data may exist where protocol and/or data entry aberrations have occurred; users should check data carefully for anomalies before joining tables.

fsp\_boutMetadata.csv - > One record expected per eventID for all time, contains the bout-level instrument parameters

fsp\_boutMetadata.csv - > One or more records expected per sampleID, generates a unique spectralSampleID per leaf or group of leaves/needles scanned. Contains useful information about the plant and its environment.

fsp\_spectralData.csv - > One record expected per spectralSampleID, provides a link to download calibrated reflectance data (one long-form table per record, see below).

Spectral files (by download URL in fsp\_spectralData) - > One record per unique combination of spectralSampleID, reflectanceCondition, and wavelength

*Note: if the expanded download package is selected, reflectance files will be downloaded automatically*

Foliar tissue used directly for field spectral measurements will be discarded following measurement. However, if the spectra are taken from canopy foliage samples, there will be an archive subsample of dried/ground tissue saved for posterity. This material can be searched for and requested via the NEON Biorepository data portal, <https://biorepo.neonscience.org/portal/index.php>.

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



### 3.10 Special Considerations

Water has strong absorption bands around ~1450nm, ~1950nm, and 2500nm; as such, we recommend excluding these regions prior to carrying out analyses using the field spectral 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 options, which reduces the number of processing steps necessary to prepare the raw data for publication. An additional set of constraints are implemented during the process of ingest into the NEON database. The product-specific data constraint and validation requirements built into data entry applications and database ingest are described in the document NEON Raw Data Validation for Field spectral data (DP0.30012.001), provided with every download of this data product. Contained within this file is a field named 'entryValidationRulesForm', which describes syntactically the validation rules for each field built into the data entry application. Data entry constraints are described in Nicl syntax in the validation file provided with every data download, and the Nicl language is described in NEON's Ingest Conversion Language (NICL) specifications ([AD[13]).

Data collected prior to 2019 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 meta-data through to publication on the NEON Data Portal are detailed in the NEON Algorithm Theoretical Basis Document: OS Generic Transitions (AD[12]). The steps used to process raw spectral measurements on black and white backgrounds into reflectance and transmission values are detailed in the NEON Algorithm Theoretical Basis Document: Field Spectral Data.

## 5.3 Data Revision

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

## 5.4 Quality Flagging

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

Table 1: Descriptions of the dataQF codes for quality flagging

fieldName	value	definition
dataQF	legacyData	Data recorded using a paper-based workflow that did not implement the full suite of quality control features associated with the interactive digital workflow

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



(DP1.10111.001)

## 6 REFERENCES

Malvern Panalytical Ltd. ASD, FieldSpec® 3 User Manual, ASD Document 600540 Rev. J; Analytical Spectral Devices, Inc.: Boulder, CO, USA, 2010.

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