# NEON USER GUIDE TO SOIL PHYSICAL PROPERTIES (DISTRIBUTED INITIAL CHARACTERIZATION) (NEON.DP1.10047) AND SOIL CHEMICAL PROPERTIES (DISTRIBUTED INITIAL CHARACTERIZATION) (NEON.DP1.10008)

| PREPARED BY | ORGANIZATION | DATE |
|---|---|---|
| Samantha Weintraub | FSU | 08/30/2019 |
| Lee Stanish | FSU | 08/30/2019 |

# CHANGE RECORD

| REVISION | DATE | DESCRIPTION OF CHANGE |
|---|---|---|
| A | 02/19/2018 | Initial Release |
| B | 08/30/2019 | Additional theory and method details provided, especially for bulk density |

# TABLE OF CONTENTS



# LIST OF TABLES AND FIGURES

# 1 DESCRIPTION

## 1.1 Purpose

This document provides an overview of the data included in these NEON Level 1 data products, the quality controlled products 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 depth of a soil horizon - 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 products Soil physical properties (distributed initial characterization) (NEON.DP1.10047) and Soil chemical properties (distributed initial characterization) (NEON.DP1.10008). Soil Physical Properties includes soil profile descriptions as well as bulk density and particle size (texture) measurements; Soil Chemical Properties includes a wide array of geochemical constituents measured in those soil profiles as well as coarse fragment content. This document 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 Soil physical properties (distributed initial characterization) (NEON.DP1.10047) (AD[05]) and NEON Data Variables for Soil chemical properties (distributed initial characterization) (NEON.DP1.10008) (AD[06]), provided in the download packages for these data products.

This document describes the process for ingesting and performing automated quality assurance and control procedures on soil physical and chemical property data produced during initial soil characterization efforts. How the Level 0 data are processed is detailed in the file, NEON Raw Data Validation for Soil chemical and physical properties (distributed initial characterization), Level 0 (AD[04]), provided in the download package for this data product. Please note that raw data products (denoted by ‘DP0’) may not always have the same numbers (e.g., ‘10033’) as the corresponding L1 data product.

# 2 RELATED DOCUMENTS AND ACRONYMS

## 2.1 Associated Documents

| | | |
| --- | --- | --- |
| AD[01] | NEON.DOC.000001 | NEON Observatory Design (NOD) Requirements |
| AD[02] | NEON.DOC.000913 | TOS Science Design for Spatial Sampling |
| AD[03] | NEON.DOC.002652 | NEON Level 1, Level 2 and Level 3 Data Products Catalog |
| AD[04] | NEON.DP0.10008.001 _dataValidation.csv | NEON Raw Data Validation for Soil chemical and physical properties (distributed initial characterization), Level 0 |
| AD[05] | NEON.DP1.10047.001 _variables.csv | NEON Data Variables for Soil physical properties (distributed initial characterization) (NEON.DP1.10047) |
| AD[06] | NEON.DP1.10100.001 _variables.csv | NEON Data Variables for Soil chemical properties (distributed initial characterization) (NEON.DP1.10008) |
| AD[07] | NEON.DOC.000906 | TOS Science Design for Terrestrial Biogeochemistry |
| AD[09] | NEON.DOC.000008 | NEON Acronym List |
| AD[10] | NEON.DOC.000243 | NEON Glossary of Terms |
| AD[11] | OS_Generic _Transitions.pdf | NEON Algorithm Theoretical Basis Document: OS Generic Transitions |
| AD[12] | | NEON's Ingest Conversion Language (NICL) specifications |

# 3 DATA PRODUCT DESCRIPTION

The Soil physical properties (distributed initial characterization) (NEON.DP1.10047) and Soil chemical properties (distributed initial characterization) (NEON.DP1.10008) data products provide information about soils measured during the course of an initial soil characterizaion effort at each NEON site. This effort was executed by the Soil Science Division of the Natural Resoures Conservation Service (NRCS), in partnership with the USDA Agriculture Research Service (ARS). The goals of the intial soil characterization effort were to describe the taxonomic, physical and geochemical properties of soils at each NEON site in locations where Terrestrial Observation System (TOS) sampling occurs. Moreover, the effort sought to capture the range of variability in soil characteristics, especially across Distributed base plots (Figure 1). The Guidelines for the NEON Soil Characterization Effort document contains more thorough descriptions of NRCS sampling procedures and can be downloaded from the Neon Data Portal.

Initial characterizaton measurements of soil physical and chemical properties help to implement the guidelines and requirements described in the TOS Science Design for Terrestrial Biogeochemistry (AD[07]). Field and laboratory data are reported at the spatial resolution of a soil horizon within a NEON plot. The temporal resolution is that of a single collection date.

Measurements of soil physical and chemical properties from the initial characterizaion effort will help to reveal drivers of variation in belowground element storage and cycling, weathering dynamics, and soil developement at plot, site, and continental scales. They may also prove useful in interpreting patterns of soil biogeochemical, microbial, and vegetation dynamics at NEON sites. The bulk density measurements included in this data product will be essential for data users interested in converting distributed periodic measurements of soil carbon and nitrogen from concentrations to stocks.

## 3.1 Spatial Sampling Design

Soils were sampled at all terrestrial NEON sites. At each site, 10-34 base plots were sampled from up to 4 Tower and 30 Distributed plots (Figure 1). The number of plots varied for each NEON site and was determined by NRCS, drawing on their extensive expertise with soil characterization along with knowledge of site variability and number of soil map units present. Relief, landforms, slope position, parent materials, and aspect are important parameters used to ensure site heterogeneity was captured by initial soil characterization efforts. See AD[02] for further details on the overall NEON spatial design and the Guidelines for the NEON Soil Characterization Effort document for more information on initial soil characterization spatial sampling.

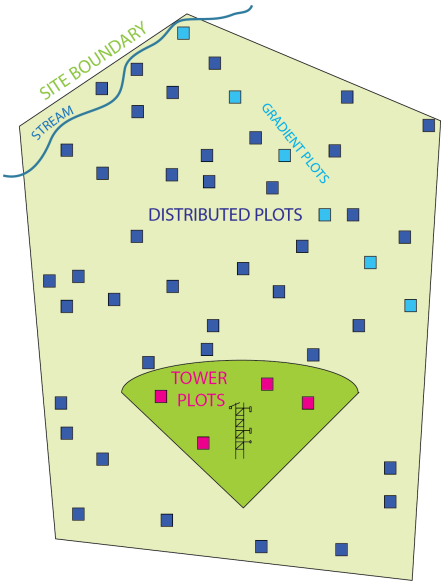


Figure 1: Representation of a NEON site with select Tower and Distributed plots shown

In most Distributed base plots, NRCS excavated a 1 m x 1 m x 1 m soil pit. In Tower plots and sites where pit sampling was not permitted, NRCS used a bucket auger to collect several 10 cm diameter cores down to 1 m. Whenever possible, all cores were taken from within a 1 m x 1 m square. Soil pits or cores were only taken from the destructive 'BGC/microbes' outer perimeter sampling zone of TOS plots, with the inner core reserved for non-destructive vegetation monitoring (Figure 2). The precise sampling locations within a plot were recorded and removed from consideration for future TOS soil sampling to ensure that future biological and chemical measurements are not influenced by this disturbance.

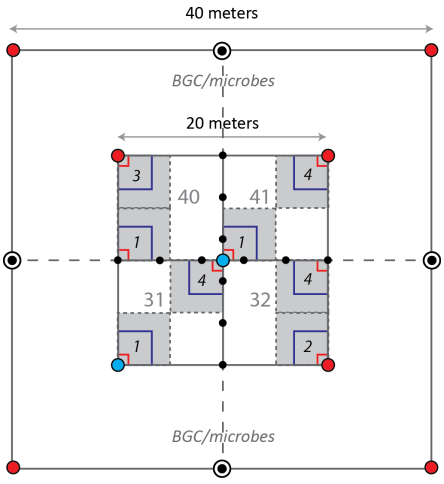


Figure 2: Representation of a NEON base plot used for soil initial characterization sampling

Upon excavating a pit or collecting cores, NRCS described the profile and all major horizons, visually estimated

very coarse (> 20 mm fragment) volumes, collectd bulk density samples (most often by the clod method), and then collected enough material to conduct all laboratory analyses. Field sampling and profile descriptions followed the methods outlined in the NRCS Field Book for Describing and Sampling Soils, version 3.0, available at https://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/nrcs142p2_052523.pdf or the Neon Data Portal.

## 3.2 Temporal Sampling Design

Soil physical and chemical properties from the initial characterizaion effort were measured once during the lifetime of each NEON site.

## 3.3 Theory of Measurement

All soils were sent to the Kellogg Soil Survey Laboratory (KSSL) in Lincoln, Nebraska for physical and geochemical analyses. Analysis methods follow standard operating procedures outlined in the Soil Survey Laboratory Methods Manual, Report No. 42, Version 5, 2014. This manual can be downloaded at https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/ref/?cid=nrcs142p2_054247 and is also available at the Neon Data Portal. For each set of analyses, the specific method used by KSSL is recorded in the relevant data table, with enough information to identify that method in the KSSL manual. For example, bulkDensMethod = Bulk Density, Clods (3B1).

### 3.3.1 Bulk Density

Bulk density (BD), or mass of soil per unit volume, is a critical measurement as it is used to convert element concentrations to stocks in the soil profile. The standard means by which NRCS measures BD is the clod method. In the spc_bulkdensity table, two different BD values are reported for clods: 1/3 bar (bulkDensThirdBar), which is essentially at field capacity, and oven dry (bulkDensOvenDry). The sample has a maximum density when it is oven-dry (since the volume shrinks), conversely the sample has a minimum density at 1/3 bar as the volume expands. For clayey soils, there can be significant difference between the two BD values, but sandy soils have little shrink-swell thus the two measurements tend to be similar. The NRCS views 1/3 bar as an excellent standard moisture content, and for most applications users will wish to use the 1/3 bar values for their calculations.

In some situations, obtaining an undisturbed clod was not possible. This includes soils with very high coarse fragment content, sandy materials, saturated conditions, pervasive woody roots, and high concentrations of organic material. Where clods could not be obtained, NRCS attempted to use other BD methods, such as compliant cavity or field cores. These method variations are recorded in the data. For non-clod methods, bulk density is reported on field-moist soil (bulkDensFieldMoist).

Regardless of the BD method used, it is important to account for *horizon* coarse fragment content when estimating element stocks, since the BD values are reported on a < 2 mm basis (e.g., 'soil'). In order to adjust for coarse fragments 2-20 mm in size, the variables coarseFrag2To5 and coarseFrag5To20, measured in grams per kilogram and reported in the spc_particlesize table, should be summed, then converted to % by multiplying by 0.1. Given the size of the NRCS pits, coarse fragment measurements are accurate for 2-20 mm fragments. If fragments > 20 mm occur in the soil, the volume estimates documented in site-specific pedon descriptions can be converted to

weight values and horizon fragment content adjusted accordingly. This is a more involved adjustment, but may be necessary on colluvial mountain settings and cobbly glacial soils.

Lastly, in areas where NRCS could not dig pits but had to take 10 cm diameter cores instead, no intact soil volume could be obtained, thus no bulk density measurement was possible. For those requiring a BD estimate, the chemical and physical data provided for those horizons can be used to estimate BD values using pedo-transfer functions.

### 3.4 Variables Reported

All variables reported from the laboratory (L0 data) are listed in the file, NEON Raw Data Validation for Soil chemical and physical properties (distributed initial characterization), Level 0 (AD[04]). All variables reported in the published data (L1 data) are provided separately in the files NEON Data Variables for Soil physical properties (distributed initial characterization) (NEON.DP1.10047) (AD[05]) and NEON Data Variables for Soil chemical properties (distributed initial characterization) (NEON.DP1.10008) (AD[06]).

Field names have been standardized with Darwin Core terms (http://rs.tdwg.org/dwc/; accessed 16 February 2014), the Global Biodiversity Information Facility vocabularies (http://rs.gbif.org/vocabulary/gbif/; accessed 16 February 2014), and the VegCore data dictionary (https://projects.nceas.ucsb.edu/nceas/projects/bien/wiki/VegCore; accessed 16 February 2014), where applicable. NEON TOS spatial data employs the World Geodetic System 1984 (WGS84) for its fundamental reference datum and GEOID09 for its reference gravitational ellipsoid. Latitudes and longitudes are denoted in decimal notation to six decimal places, with longitudes indicated as negative west of the Greenwich meridian.

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

### 3.5 Spatial Resolution and Extent

The finest resolution at which spatial data are reported is a unique soil pit location and horizon within a NEON plot.

**horizonID** (unique ID given to the individual soil horizon) → **pitID** (unique ID given to the individual soil pit) →**plotID** (ID of plot within site) → **siteID** (ID of NEON site) → **domainID** (ID of a NEON domain.

The basic spatial data included in the soil physical properties data download includes latitude, longitude, and elevation of the *centroid* of the plot where sampling occurred, plus associated uncertainty due to GPS error. The basic spatial data included in the soil chemistry data product is the location of the NEON plot. Shapefiles of all NEON Terrestrial Observation System plot locations can be found in the Document Library: http://data.neonscience.org/documents.

In the soil physical properties data product, if **referenceCorner**, **sampleDistance** and **sampleBearing** are provided, users will be able to calculate more precise geolocations of the soil pit using the following procedure:

1. Obtain easting and northing for the plot centroid. This can be accomplished either by using the def.extr.geo.os function in the geoNEON R package (https://github.com/NEONScience/NEON-geolocation),

or using the NEON API (http://data.neonscience.org/data-api; e.g. http://data.neonscience.org/api/v0/locations/BART_001.basePlot.all) to query for easting ("locationUtmEasting") and northing ("locationUtmNorthing") of plot named locations.

2. Calculate northing of the reference corner by either subtracting or adding 10 or 20 meters from plot centroid northing, depending on the entry for **referenceCorner** (Figure 3).

For example, if **referenceCorner** = SW20,

$$referencecornernorthing = plotcentroidnorthing - 10m$$

If **referenceCorner** = NE40,

$$referncecornernorthing = plotcentroidnorthing + 20m$$

3. Calculate easting of the reference corner by either subtracting or adding 10 or 20 meters from plot centroid easting, depending the the choice in **referenceCorner**.

For example, if **referenceCorner** = SW20,

$$referencecornereasting = plotcentroideasting - 10m$$

If **referenceCorner** = NE40,

$$referencecornereasting = plotcentroideasting + 20m$$

4. Calculate northing and easting of the pit based on **sampleDistance** and **sampleBearing** using the following equations:

$$pitnorthing = referencecornernorthing + d * \sin\theta \quad (1)$$

and

$$piteasting = referencecorninereasting + d * \cos\theta \quad (2)$$

where, if sampleBearing < 90

$$\theta = 90 - sampleBearing \quad (3)$$

else

$$\theta = 450 - sampleBearing \quad (4)$$

and

$$d = sampleDistance \tag{5}$$

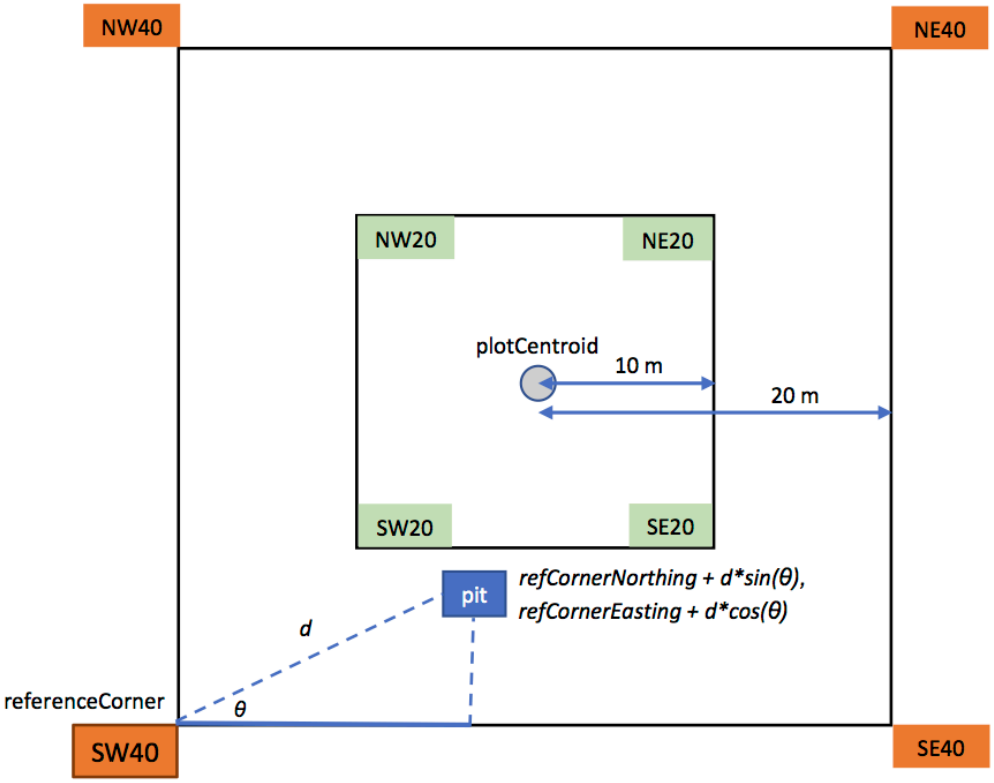


Figure 3: Diagram showing the locations of reference corners and an example calculation of sample location using distance and bearing measurements.

5. Increase **coordinateUncertainty** associated with the pit location by an appropriate amount (suggested 1 m) to account for error introduced by measurement and navigation within the plot.

### 3.6 Temporal Resolution and Extent

The finest resolution at which temporal data are reported is a **collectDate**. The total number of sampling events will be one per site for the lifetime of NEON.

### 3.7 Associated Data Streams

These data products are self-contained, meaning that together, they contain all parent and child samples. Along with physical and chemical data, NRCS also provides the following:

- Site Level Plot Summaries. These are narrative summaries that place the sampled soils in the broader context of soils and geomorphology for the entire NEON site.
- Pedon descriptions. These are pit-level observations and field measurements and are reported using the standard NRCS format. They contain volume estimates for coarse fragments > 20 mm where applicable

Both forms of documentation are available on the NEON Data Portal, under Resources > Document Library > Soil Characterization Summaries > Distributed Plots.

In addition, certain horizons had left-over soil material, which has been archived in collaboration with the International Soil Carbon Network (ISCN) and may be available upon request. For details, refer to the ISCN webpage, https://iscn.fluxdata.org/.

### 3.8 Product Instances

At each NEON site, 10-34 base plots will be sampled for initial soil taxonomic, physical and chemical properties. The most common sampling method is excavation of a single soil pit per plot, although collection of 1-6 cores (which are composited by horizon for characterization) is also possible. The composite of all core locations is treated as a 'pit' for this data product. The type of sampling method employed is recorded in the variable **soilSamplingMethod** in the table spc_perplot. Each pit is sampled by soil horizon and characterized for taxonomic, physical and chemical properties. Bulk density is measured on all horizons when possible. Assuming pits have 4-8 horizons, this will result in 120-816 unique records of physical and chemical properties per site. It is anticipated that all terrestrial NEON sites will be sampled, yielding a range of 5,460-38,358 data records across these two data products.

### 3.9 Data Relationships

Guidelines for the NEON Soil Characterization Effort dictates that each soil pit sampled yields a unique pitID in the spc_perplot table of the Soil physical properties data product. A record from spc_perplot then has several child records, one for each horizon in the pit, in spc_perhorizon, provided in that same data product. Each horizon record from spc_perhorizon will then have zero or one child records in spc_bulkdensity and spc_particlesize (in the Soil physical properties data product) and spc_biogeochem (in the Soil chemical properties data product). It is expected that child sample identifiers will only appear once per table, but duplicates and/or missing data may exist where protocol and/or data entry abberations have occurred. *Users should check data carefully for anomalies before joining tables*.

*Soil physical properties:*

spc_perplot.csv - > One record expected per **pitID**

spc_perhorizon.csv - > Multiple records expected per **pitID**, generates unique **horizonID**s for each soil horizon in the pit

spc_bulkdensity.csv - > One record expected per **horizonID**, generates a single **bulkDensIDnrcs** used to measure bulk density

spc_particlesize.csv - > One record expected per **horizonID**, generates a single **biogeoIDnrcs** used to measure particle size and coarse fragment (2-20 mm) content

*Soil chemical properties:*

spc_biogeochem.csv - > One record expected per **horizonID**, generates a single **biogeoIDnrcs** used to measure biogeochemical properties

spc_externalSummary.csv - > One record expected per **laboratoryName** x **analyte** x **sampleType** x **labSpecificStartDate** combination. Can use corresponding variables in biogeochem table to associate sample data with relevant uncertainty values. Not all analytes will be reported.

# 4 DATA QUALITY

## 4.1 Data Entry Constraint and Validation

Constraints and data validation are implemented during the process of ingest into the NEON database. The product-specific data constraint and validation requirements built into data entry and database ingest are described in the document NEON Raw Data Validation for Soil chemical and physical properties (distributed initial characterization), Level 0, provided with every download of this data product. Contained within this file is a field named 'entryValidationRulesParser', which describes syntactically the validation rules for each field built into the data ingest process. 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]).

## 4.2 Automated Data Processing Steps

Following data entry into a mobile application or web user interface, the steps used to process the data through to publication on the NEON Data Portal are detailed in the NEON Algorithm Theoretical Basis Document: OS Generic Transitions (AD[14]).

## 4.3 Data Revision

All data are provisional until a numbered version is released; the first release of a static version of NEON data, annotated with a globally unique identifier, is planned to take place in 2020. During the provisional period, QA/QC is an active process, as opposed to a discrete activity performed once, and records are updated on a rolling basis as a result of scheduled tests or feedback from data users. The Change Log section of the data product readme, provided with every data download, contains a history of major known errors and revisions.

### 4.4 Quality Flagging

The **dataQF** field in each data record is a quality flag for known errors applying to the record. There are currently no **dataQF** codes in use in this data product.

### 4.5 Analytical Facility Data Quality

Data in these products are subject to the standard analytical quality control procedures used by the NRCS, as documented in the Soil Survey Laboratory Methods Manual, Report No. 42, Version 5, 2014. Analyses conducted at the Lincoln Laboratory include standards run as unknowns alongside samples in order to gauge run acceptability. Long-term analytical precision and accuracy of these standard analyses are reported to allow users to interpret and model soil geochemical data in the context of its uncertainty range. The data table spc_externalLabSummary (available in the expanded package) contains the long-term precision and accuracy of lab analyses.

Analytical results below the detection limit of the method or instrument are reported as zero, equivalent to choosing 'Replace trace-and-dash notation with zero' when downloading data directly from the National Cooperative Soil Survey.