NEON USER GUIDE TO SOIL PHYSICAL PROPERTIES (DISTRIBUTED PERIODIC) (NEON.DP1.10086)

PREPARED BY	ORGANIZATION	DATE
Lee Stanish	FSU	05/31/2017
Samantha Weintraub	FSU	05/31/2017

CHANGE RECORD

REVISION	DATE	DESCRIPTION OF CHANGE
A	04/20/2017	Initial Release

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

1 DESCRIPTION

1.1 Purpose

This document provides an overview of the data included in this NEON Level 1 data product, the quality controlled product generated from raw Level 0 data, and associated metadata. In the NEON data products framework, the raw data collected in the field - for example, soil temperature 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 Soil physical properties (Distributed periodic), which encompasses field and laboratory measurements of soil temperature, moisture content, and pH as well as generating subsamples used for chemical and microbial analyses. Data from the subsamples can be found in the related data products listed below. 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 Soil physical properties (Distributed periodic) (NEON.DP1.10086) (AD[05]), 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: Soil Biogeochemical and Microbial Sampling (AD[08]), or TOS Standard Operating Procedure: Wetland Soil Sampling (AD[09]) if the site is a wetland. The raw data that are processed in this document are detailed in the file, NEON Raw Data Validation for Soil physical properties (Distributed periodic) (NEON.DP0.10086) (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.DP0.10086.001_dataValidation.csv	NEON Raw Data Validation for Soil physical properties (Distributed periodic) (NEON.DP0.10086)
AD[05]	NEON.DP1.10086.001_variables.csv	NEON Data Variables for Soil physical properties (Distributed periodic) (NEON.DP1.10086)
AD[06]	NEON.DOC.000906	TOS Science Design for Terrestrial Biogeochemistry
AD[07]	NEON.DOC.000908	TOS Science Design for Terrestrial Microbial Diversity
AD[08]	NEON.DOC.014048	TOS Protocol and Procedure: Soil Biogeochemical and Microbial Sampling
AD[09]	NEON.DOC.004130	TOS Standard Operating Procedure: Wetland Soil Sampling
AD[10]	NEON.DOC.000008	NEON Acronym List
AD[11]	NEON.DOC.000243	NEON Glossary of Terms
AD[12]	NEON.DOC.004130	TOS Standard Operating Procedure: Wetland Soil Sampling
AD[13]	OS_Generic _Transitions.pdf	NEON Algorithm Theoretical Basis Document: OS Generic Transitions
AD[14]		NEON's Ingest Conversion Language (NICL) specifications

3 DATA PRODUCT DESCRIPTION

The Soil physical properties (Distributed periodic) data product is derived from the TOS Protocol and Procedure: Soil Biogeochemical and Microbial Sampling (AD[08]), or TOS Standard Operating Procedure: Wetland Soil Sampling (AD[09]) if the site is a wetland. The sampling plan implements the guidelines and requirements described in the Science Designs for TOS Terrestrial Biogeochemistry (AD[06]) and Microbial Diversity (AD[07]). All accompanying field and laboratory data are reported at the spatial resolution of a single soil sampling location, e.g., an x,y coordinate (+/- 0.5 meters) within a NEON plot, exept for microbial metagenomic samples, which represent a pooled, plot-level composite and are thus reported at the scale of a NEON plot. For all samples, the temporal resolution is that of a single collection date.

Soils are sampled by horizon type (organic or mineral, Figure 1) to a maximum depth of 30 cm. Where possible, users can validate the technician designation of horizon type using measured carbon content from the Soil chemical properties (Distributed periodic) (NEON.DP1.10078) data product. The type of device used to collect soils varies based on local soil types and seasonal conditions and is recorded for each sample. At each x,y sampling location, in-situ soil temperature is measured and soil samples are removed and homogenized. Subsamples for microbial community analyses are immediately frozen on dry ice. Following field sampling, gravimetric soil moisture and pH measurments are conducted at NEON domain support facilites, following the methods outlined in Robertson et al. (1999).

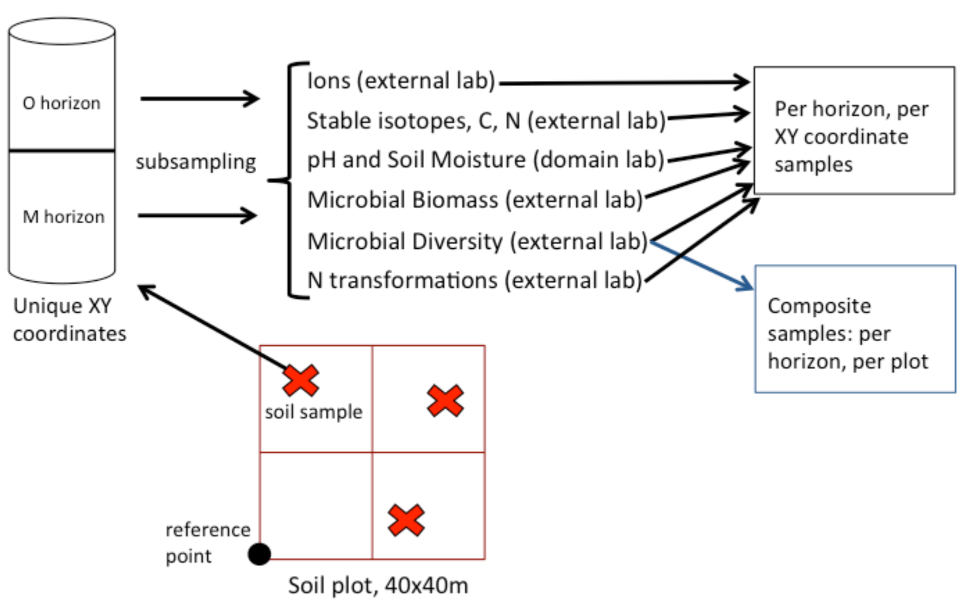

Figure 1: Overview of soil field sampling and analysis workflow

Measurements of soil physical properties help to elucidate the constraints on microbial acitvity, nutrient cycling, and carbon storage in soils at the plot, site, and continental scales. They also provide essential data for understanding change in soil microbial and biogeochemical dynamics over time.

3.1 Spatial Sampling Design

Soil biogeochemical and microbial sampling is executed at all terrestrial NEON sites. Soils are sampled from three pre-determined, randomly assigned x,y locations per 40 x 40 meter plot (Figure 1). If large rocks, tree roots, animal burrows, or other significant impediments are encountered at a pre-determined sampling location, technicians proceed down a list of random, alternate locations until an acceptable one is found. Ten plots per site are sampled, four within the tower airshed (Figure 2) and six others distributed across the landscape, located in dominant vegetation types. Tower plots are selected using a random spatial design, while distributed plots follow a random stratified design based on National Land Cover Database (NLCD) vegetation class. The number of distributed plots within each NLCD class are proportional to the percent coverage of that class. See AD[02] for further details on the NEON spatial design.

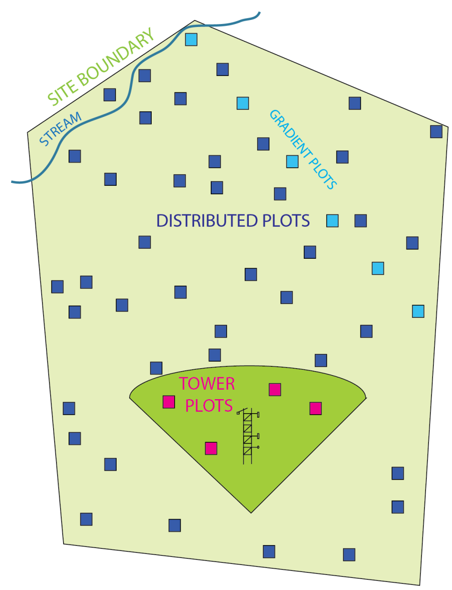

Figure 2: Representation of a NEON site with Tower and Distributed plots shown

3.2 Temporal Sampling Design

Soil sampling for a suite of microbial analyses (including a plot-level, pooled sample for metagenomics), physical properties (temperature, moisture), and pH occurs annually at each site during the period of peak greenness. At nearly all sites, soil sampling occurs two additional times per year, usually bracketing the peak greeness window and aimed at capturing seasonal transitions in microbial activity. Sampling frequency is reduced in arctic and boreal sites, which are only sampled during peak greenness. During transitional season sampling, a pooled metagenomics sample is not collected. On an interannual basis, soils are utilized for additional downstream analyses. For instance, incubation and processing of soil cores to determine inorganic nitrogen pools and net transformation rates occurs every five years, following the same seasonal sampling schedule outlined above (Figure 3). Measurement of soil chemical and isotopic composition plus creation of an air-dried subsample for archive occurs once every 5-10 years, utilizing peak greenness samples only.

Thus, of the five data tables contained in this data product, field collection, pH, and moisture tables will contain data for 1-3 sampling events per site per year, the metagenomics pooling table will contain data for 1 sampling event per site per year, and the biogeochemistry subsampling table will contain data for one sampling event per site, every 5-10 years. Additionally, once every 5 years per site, inorganic nitrogen pools in initial and incubated cores will be available as part of the Soil inorganic nitrogen pools and transformations (NEON.DP1.10080) data product.

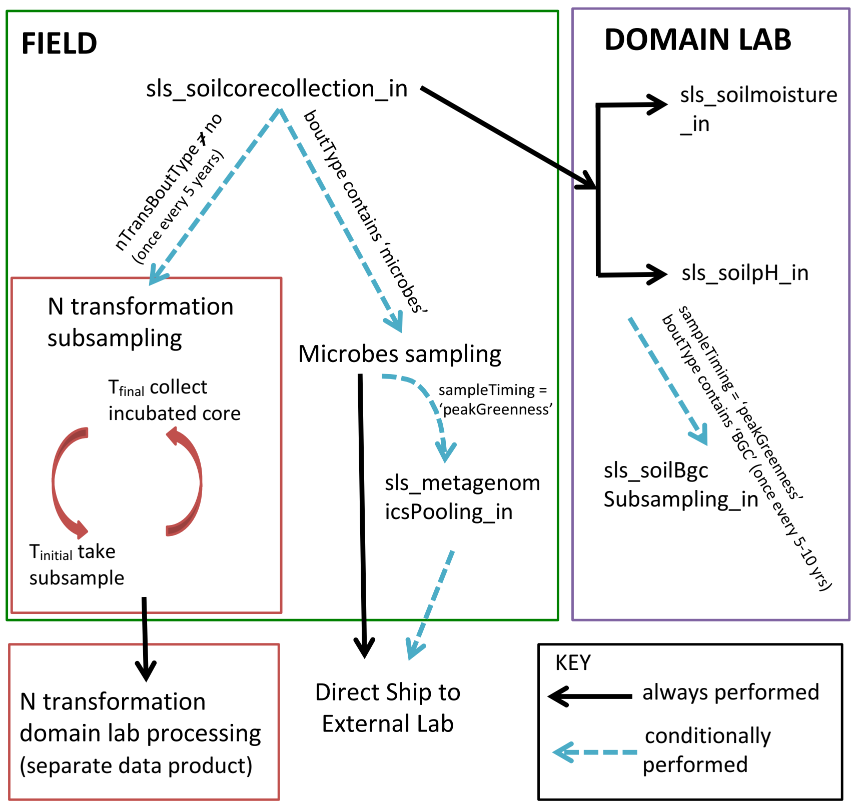

Figure 3: Soil field and laboratory workflows. Arrows indicate which data tables are generated by each step in the soil sampling process and under what conditions. Data ingest table names follow the convention: moduleID_tableName_in, for example sls_soilcorecollection_in

3.3 Variables Reported

All variables reported from the field or laboratory technician (L0 data) are listed in the file, NEON Raw Data Validation for Soil physical properties (Distributed periodic) (NEON.DP0.10086) (AD[04]). All variables reported in the published data (L1 data) are also provided separately in the file, NEON Data Variables for Soil physical properties

(Distributed periodic) (NEON.DP1.10086) (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), and the VegCore data dictionary (https://projects.nceas.ucsb.edu/nceas/projects/bien/wiki/VegCore; accessed 16 February 2014), where applicable. NEON TOS spatial data employs the World Geodetic System 1984 (WGS84) for its fundamental reference datum and Earth Gravitational Model 96 (EGM96) for its reference gravitational ellipsoid. Latitudes and longitudes are denoted in decimal notation to six decimal places, with longitudes indicated as negative west of the Greenwich meridian.

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

3.4 Spatial Resolution and Extent

The finest resolution at which spatial data are reported is a single x,y sampling location.

sampleID (unique ID given to the individual soil sampling location and horizon, based on x,y coordinates relative to the southwest corner of the plot) → **plotID** (ID of plot within site) → **siteID** (ID of NEON site) → **domainID** (ID of a NEON domain.

The basic spatial data included in the data downloaded include spatial location (northing and easting) and elevation of the *centroid* of the plot where sampling occurred, plus associated uncertainty due to GPS error and plot width. Shapefiles of all NEON Terrestrial Observation System sampling locations can be found in the Document Library: http://data.neonscience.org/documents.

To derive a more precise estimate of the location of each x,y soil sampling location, there are two options:

- Use the def.calc.geo.os function from the geoNEON package, available here: https://github.com/NEONScience/NEON-geolocation
- Or follow these steps to perform the same calculation:

1. Calculate the actual northing of the x,y location. This can be accomplished by subtracting 20 meters from plot centroid northing, then adding the y-coordinate value. Use the following formula:

$$yCoordNorthing = northing - 20m + coreCoordinateY$$

2. Calculate the actual easting of the x,y location. This can be accomplished by subtracting 20 meters from plot centroid easting, then adding the x-coordinate value. Use the following formula:

$$xCoordEasting = easting - 20m + coreCoordinateX$$

3. Increase coordinateUncertainty by an appropriate amount to account for variance in sampling location area (+/- 0.5 m), plus the error introduced by technicians navigating within plots, stretching meter tapes to navigate to x,y locations, etc. This uncertainty, on average, will be +/- 1 m, but will vary by site based on terrain heterogeneity and density of vegetation.

3.5 Temporal Resolution and Extent

The finest resolution at which temporal data are reported is collectDate. All samples associated with a sampling event have collectDates within a ~14-day window and are categorized by the target season or seasonal transition (e.g. wet-dry transition, winter-spring transition, peak greenness). The total number of sampling events per year will vary among sites, based on the length of the growing season. It is expected that 3 sampling events will occur annually for microbial sampling in most sites, except those in arctic and boreal regions.

3.6 Associated Data Streams

Field and laboratory data from the Soil physical properties (Distributed periodic) data product will be necessary to interpret and utilize several other related data products. For Soil microbe community composition (NEON.DP1.10081), Soil microbe group abundances (NEON.DP1.10109), and Soil microbe marker gene sequences (NEON.DP1.10108) **geneticSampleID** is the variable name that links samples and their associated metadata from field to external laboratory data tables. For Soil microbe biomass (NEON.DP1.10104), this variable is **biomassID**, while for Soil microbe metagenome sequences (NEON.DP1.10107), the linking varaible name is **compositeSampleID**. For the Soil inorganic nitrogen pools and transformations (NEON.DP1.10080) data product, **sampleID** is the variable name needed to join external laboratory and field data tables. For the Soil chemical properties (Distributed periodic) (NEON.DP1.10078) and Soil stable isotopes (Distributed periodic) (NEON.DP1.10100) data products, **cnSampleID** should be used to join the data products together.

3.7 Product Instances

Soil samples will be collected at all terrestrial NEON sites. A maximum of 10 plots will be sampled at every site at a frequency of 3 times per year (except for arctic and boreal sites, which are sampled once). For each soil horizon (maximum of 2, organic or mineral), 3 samples per plot are collected. Thus at most sites, it is expected that soil sampling will result in 90-180 unique soil samples per site per year, analyzed for moisture and pH. In any given year, each unique sample may result in 0-8 subsamples used for additional analyses and archive.

3.8 Data Relationships

The protocol dictates that each x,y location sampled yields a unique sampleID per horizon per collectDate (day of year, local time) in sls_soilCoreCollection. A record from sls_soilCoreCollection may have zero or one child records in sls_soilpH and sls_soilMoisture; a given sls_soilCoreCollection.sampleID is expected to be sampled only once. Depending on the type of bout and time of year (Figure 3), a record from sls_soilCoreCollection may have zero or one child records in sls_metagenomicsPooling and in sls_bgcSubsampling. 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*.

sls_soilCoreCollection.csv - > One record expected per sampleID. Depending upon boutType, each sampleID generates up to a single geneticSampleID, up to 3 geneticArchiveSamplesIDs (labeled geneticArchiveSampleXIDs where X is a number one to three), and up to a single biomassID. Generates samples used in Soil microbe community composition (NEON.DP1.10081), Soil microbe group abundances (NEON.DP1.10109), Soil microbe marker gene sequences (NEON.DP1.10108), and Soil microbe biomass (NEON.DP1.10104). Additionally, subsamples generated from soil sampleIDs are used in Soil inorganic nitrogen pools and transformations (NEON.DP1.10080), though inorganic N data does not appear in this data table.

sls_soilpH.csv - > One record expected per sampleID, generates a single pHSampleID.

sls_soilMoisture.csv - > One record expected per sampleID, generates a single moistureSampleID.

sls_metagenomicsPooling.csv - > One record expected per plotID per horizon per collectDate (day of year, local time). Record represents a mixture of the samples collected in a plot (listed in toCompositeSampleIDList). Each record generates a single compositeSampleID, used in Soil microbe metagenome sequences (NEON.DP1.10107).

sls_bgcSubsampling.csv - > One record expected per sampleID, generates a single cnSampleID and bgcArchiveID. Generates samples used in Soil chemical properties (Distributed periodic) (NEON.DP1.10078) and Soil stable isotopes (Distributed periodic) (NEON.DP1.10100).

sampleIDs will be generated for each unique physical soil sample created during a collection event, and **sampleBarcodes** may also be generated. Each sampleID will yield between one to three frozen archive subsamples (**geneticArchiveSampleXID**), to be stored at -80C. When a collection event includes soil chemistry and isotope analyses, an air-dried archive subsample (**bgcArchiveID**) will also be created. If not, following pH and moisture measurments, any remaining soil material will be discarded at the end of the calendar year.

3.9 Special Considerations

Every five years, incubation and processing of soil cores occurs in order to measure net nitrogen transformation rates. Incubated cores, designated with **nTransBoutType = tfinal**, will appear intermingled with non-incubated samples throughout the sls_soilCoreCollection and sls_soilMoisture data tables. However, they will not be measured for soil pH and will not be subsampled for microbial community or soil chemistry and isotopic analyses.

4 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 Soil physical properties (Distributed periodic) (NEON.DP0.10086), 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[14]).

5 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[13]). Additionally, the methods used to create calculated fields from the raw data (soil moisture, water-to-solution ratios for pH measuremnts) are detailed in NEON Raw Data Validation for Soil physical properties (Distributed periodic) (NEON.DP0.10086).

6 REFERENCES

Robertson, G. P. (1999). Standard soil methods for long-term ecological research (Vol. 2). Oxford University Press on Demand.