NEON USER GUIDE TO SOIL PHYSICAL AND CHEMICAL PROPERTIES, MEGAPIT (DP1.00096.001)

PREPARED BY	ORGANIZATION
Ed Ayres	TIS
Robert Lee	TIS
Josh Roberti	TIS

CHANGE RECORD

REVISION	DATE	DESCRIPTION OF CHANGE
A	06/30/2017	Initial Release
B	06/30/2020	Details the combined delivery of physical and chemical measurements

TABLE OF CONTENTS

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 bottom 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 product Soil physical and chemical properties, Megapit - and associated metadata from input data. 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 Soil physical and chemical properties, Megapit (DP1.00096.001) (AD[05]) and NEON Categorical Codes for Soil physical and chemical properties, Megapit (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 TIS Soil Pit Sampling Protocol (AD[07]). The raw data that are processed in this document are detailed in the file, NEON Raw Data Validation for Soil properties (megapit) (DP0.00096.001) (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 Data Products Catalog
AD[04]	Available with data download	NEON Raw Data Validation for Soil properties (megapit) (DP0.00096.001)
AD[05]	Available with data download	NEON Data Variables for Soil physical and chemical properties, Megapit (DP1.00096.001)
AD[06]	Available with data download	NEON Categorical Codes for Soil physical and chemical properties, Megapit
AD[07]	NEON.DOC.001307	TIS Soil Pit Sampling Protocol
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

2.2 Acronyms

Acronym	**Definition**
NRCS	Natural Resource Conservation Service
TIS	Terrestrial Instrument System
TOS	Terrestrial Observation System
USDA	United States Department of Agriculture

3 DATA PRODUCT DESCRIPTION

The Soil physical and chemical properties, Megapit data product includes data on the soil taxonomy, horizon names, horizon depths, soil bulk density, coarse fragment content, texture (sand, silt, and clay content), pH, electrical conductivity, and total content of a range of chemical elements in the <2 mm soil fraction for each soil horizon sampled in a TIS soil pit. This work was conducted as a collaboration between the NEON Terrestrial Instrument System (TIS) and the UDSA Natural Resources Conservation Service (NRCS). As of 2020, physical and chemical properties of Megapit soils are reported together and DP1.10097.001 is published as part of DP1.10096.001. Prior to this, the two types of measurements were reported in separate data downloads.

Data were derived from a sampling location expected to be representative of the area where the TIS instrumented soil plots per site are located and were collected once during site construction. NEON's other soil physical and chemical data products (Soil physical and chemical properties, distributed initial characterization; Soil physical and chemical properties, distributed periodic) represent distributed sampling across many soil pits/plots, and the periodic product involves repeated sampling over the lifetime of the observatory. Soil from most soil Megapit horizons are archived in the Megapit Soil Archive (https://www.neonscience.org/data/samples-specimens/megapit-soil-archive) and are available upon request for additional analyses. The Megapit Information page (https://data.neonscience.org/megapit-info) contains pit descriptions and photos.

3.1 Spatial Sampling Design

The Soil physical and chemical properties, Megapit data product is available at each terrestrial site. The soil pit location was chosen to be representative of the five sensor-based soil plots based on soil type, vegetation and topography, as well as being accessible by a backhoe and outside the main measurement zone of other sensors. The Megapit was usually within a few hundred meters of the sensor-based soil plots and NEON tower. Soil samples were collected by soil horizon up to 200 cm at non-permafrost sites (up to 300 cm at Alaskan sites). The sampling procedure is described in AD[06].

3.2 Temporal Sampling Design

Soil pits are sampled once during construction of a NEON terrestrial site, including at future locations of relocatable sites. They represent a point in time, although many of the soil properties typically change fairly slowly (e.g., over decades). The sampling procedure is described in AD[06].

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 properties (megapit) (DP0.00096.001) (AD[04]). All variables reported in the published data (L1 data) are also provided separately in the file, NEON Data Variables for Soil physical and chemical properties, Megapit (DP1.00096.001) (AD[05]).

NEON TIS 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 five 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.

Because the primary users of this data product are expected to be US-based ecologists, soil scientists, and other environmental science researchers, the USDA NRCS soil taxonomic classification system is used since it is expected to be the classification system that is most familiar to the users.

3.4 Theory of Field and Laboratory Measurements

A brief description of each method is included below. See the Soil Survey Manual (Soil Survey Division Staff 1993) and the USDA NRCS Soil Survey Laboratory Methods Manual (Soil Survey Staff 2004) for details on the field- and laboratory-based methods, respectively.

Soil taxonomy, horizon name, and horizon depths (**soilSeries**, **soilFamily**, **soilSubgroup**, **soilGreatGroup**, **soilSuborder**, **soilOrder**, **horizonName**, **horizonTopDepth** and **horizonBottomDepth**) are described in the field by local soil scientist, typically from the USDA NRCS, following the Soil Survey Manual guidelines (Soil Survey Division Staff 1993). This description is performed for every soil Megapit and for every horizon within the pit.

Bulk density sample volume is measured at the time of sample collection, while the weight of the bulk density sample and of the >2 mm rock fragments from the sample are determined in the laboratory after oven drying and sieving in order to calculate bulk density (**bulkDensExclCoarseFrag**). A particle density of 2.65 g cm-3 for the >2 mm fragments is assumed. This analysis is performed on all samples.

The proportion of rock fragments in 2-5 mm and 5-20 mm size classes (**coarseFrag2To5** and **coarseFrag5To20**) are determined by weighing the entire biogeochemistry soil sample and then sieving the sample to separate it into the size classes and weighing each class. Particles larger than 20 mm cannot be reliably estimated due to limitations in the quantity of soil that can be sent for analysis. This analysis is performed on all samples.

Soil particle-size distribution (i.e., soil texture) is determined on the <2 mm soil fraction by wet sieving and dry sieving (sand fractions) and sedimentation (silt and clay fractions). The particle-size categories are: total sand (0.047-2 mm; **sandTotal**), total silt (0.002-0.047 mm; **siltTotal**), total clay (<0.002; **clayTotal**), fine silt (0.002-0.02 mm; **siltFineContent**), coarse silt (0.02-0.05 mm; **siltCoarseContent**), very fine sand (0.047-0.105 mm; **sandVeryFineContent**), fine sand (0.105-0.25 mm; **sandFineContent**), medium sand (0.25-0.5 mm; **sandMediumContent**), coarse sand (0.5-1 mm; **sandCoarseContent**), and very coarse sand (1-2 mm; **sandVeryCoarseContent**). This analysis is performed on samples that were from all soil horizons except Oi.

Carbonate clay content of the <2 mm soil fraction (**carbonateClay**) is determined by adding hydrochloric acid to the residue from the particle-size distribution analysis and measuring the amount of CO2 evolved. This analysis is performed on samples that have effervescence greater than or equal to “slight” as defined by NRCS.

The total concentration of C, N, and S (**carbonTot**, **nitrogenTot**, and **sulfurTot**) are assessed using an elemental analyzer. This analysis is performed on all samples.

The total concentration of other elements (Al, Ca, Fe, K, Mg, Mn, Na, P, Si, Sr, Ti, Zr; **alMjelm**, **caMjelm**, **feMjelm**, **kMjelm**, **mgMjelm**, **mnMjelm**, **naMjelm**, **pMjelm**, **siMjelm**, **srMjelm**, **tiMjelm** and **zrMjelm**) are assessed via hydrofluoric acid, nitric acid, and hydrochloric acid digestion followed by analysis on an inductively coupled plasma atomic emission spectrometer (ICP-AES). This analysis is performed on all samples.

Soil pH is measured in soil-deionized water (1:1; **phH2o**) and soil-salt solution (1:2 CaCl2; **phCacl2**) using an electrode. This analysis was performed on all samples.

Electrical conductivity is determined by mixing soil with deionized water (1:2 weight basis; **ec12pre**), leaving the sample to equilibrate for several hours, and measured using a conductivity meter. This analysis is performed on all samples except some samples high in plant material.

Gypsum content (**gypsumConc**) is measured by dissolving it on water, extracting a subsample of the liquid and adding acetone to precipitate out the gypsum. The gypsum is then dissolved in water again and the concentration is determined by measuring the electrical conductivity of the solution. This analysis is performed on samples that had an electrical conductivity of $\geq 0.5dSm^{-1}$.

Carbonate content (**caco3Conc**) is measured by determining the amount of CO2 released following the addition of hydrochloric acid (HCl) to the soil. Carbonate content was expressed as calcium carbonate. This analysis is performed on samples that had a pH measured in CaCl2 $\geq\ 6.95$ and/or effervescence of greater than “none” in the NRCS soil description.

Ammonium acetate extractable Ca, K, Mg, Na, the sum of extractable bases, and ammonium acetate cation exchange capacity (**caNh4d**, **kNh4d**, **mgNh4d**, **naNh4d**, **baseSumCecd10** and **cecdNh4**) are measured by leaching the soil with ammonium acetate buffered at a pH of 7 to saturate the cation exchange sites in the soil. The leachate is collected and analyzed for Ca, K, Mg, and Na content using an atomic absorption spectrophotometer. The sum of extractable bases represents the sum of Ca, K, Mg, and Na content in the leachate. The ammonium acetate cation exchange capacity (cecdNh4) is determined by rinsing the ammonium acetate saturated soil with ethanol to remove excess ammonium, extracting the remaining ammonium with potassium chloride solution, and determining the concentration of ammonium via steam distillation and titration. Ammonium acetate extractable Ca, K, Mg, Na and ammonium acetate cation exchange capacity are measured on all samples. However, for samples that required grinding to pass through an 80-mesh sieve in order to perform the analysis (usually organic soil samples) the sum of extractable bases is not calculated as it may not be representative of field condition (NRCS Kellogg Soil Survey Laboratory comment).

Potassium chloride (KCl) extractable Al and Mn (**alKcl** and **mnKcl**) are measured by leaching the soil with potassium chloride solution, adding hydrochloric acid to the leachate, and analyzing it on an inductively coupled plasma atomic emission spectrophotometer (ICP-AES). This analysis ia performed on samples that had a pH measured in CaCl2 ≤ 5.05.

Effective cation exchange capacity (**ececCecd33**) is calculated by adding the sum of ammonium acetate extractable bases to the potassium chloride extractable Al. This calculation is performed on samples that are not ground to pass through an 80-mesh sieve.

Base saturation (**bsesatCecd10**) is determined by dividing the sum of ammonium acetate extractable bases (**baseSumCecd10**) by the ammonium acetate cation exchange capacity (**cecdNh4**) and multiplying by 100 to express it as a percentage. This calculation is performed on samples that are not ground to pass through an 80-mesh sieve.

Aluminum saturation (**alSatCecd33**) is calculated by dividing potassium chloride extractable Al (**alKcl**) by the effective cation exchange capacity (**ececCecd33**) and multiplying by 100 to express it as a percentage. This calculation is performed on samples where both these measurements were available.

Electrical conductivity of saturated paste and water content of saturated paste (**ecSatp** and **waterSatx**) are determined by adding reverse osmosis water to soil to achieve a saturated paste. A saturated paste has the following subjective characteristics: 1) glistens as it reflects light; 2) flows slightly when the container it is in is tilted; and 3) slides freely and cleanly from a spatula (except soils with high clay content). The paste is covered and left overnight and the characteristics are rechecked and additional water is added if necessary. Since the amount of water that needs to be added to achieve a saturated paste is somewhat subjective, the water content of the paste is determined by weighing a subsample followed by oven-dying and reweighing, and reported as a percentage of soil dry weight. The electrical conductivity of the saturated paste is determined using a conductivity meter. This analysis is performed on samples that had an electrical conductivity of $\geq 0.25 dSm^{-1}$.

Ca, K, Mg, and Na in the saturation extract (**caSatx**, **kSatx**, **mgSatx** and **naSatx**) are determined by extracting the water from a saturated paste under a vacuum. An ionization suppressant (lanthanum oxide) is added to the extract and Ca, K, Mg, and Na concentration is measured using an atomic absorption spectrophotometer. This analysis is performed on samples that had an electrical conductivity of $\geq 0.25 dSm^{-1}$.

Bromide, chloride, fluoride, nitrite, nitrate, phosphate, and sulfate in the saturation extract (**brSatx**, **clSatx**, **flSatx**, **no2Satx**, **no3Satx**, **pSatx** and **so4Satx**) are determined by extracting the water from a saturated paste under a vacuum. The extract is diluted with reverse osmosis water and injected into an ion chromatograph to separate the anions. A conductivity detector is used to identify the anions and measure their concentration. This analysis is performed on samples that had an electrical conductivity of $\geq 0.25 dSm^{-1}$.

Carbonate and bicarbonate in the saturation extract (**co3Satx** and **hco3Sx**) are determined by extracting the water from a saturated paste under a vacuum. A subsample of the extract is diluted with reverse osmosis water and the concentration of carbonate and bicarbonate are determined via titration. This analysis is performed on samples that had an electrical conductivity of $\geq 0.25 dSm^{-1}$.

The pH of the saturated extract (**phSp**) is determined on samples that had an electrical conductivity of $\geq 0.25 dSm^{-1}$.

Estimated organic carbon content (**estimatedOC**) is calculated by subtracting the carbon content of calcium carbonate (% gravimetric) measured in the sample from the total carbon content (% gravimetric). The carbon content of the calcium carbonate is determined by dividing weight of calcium carbonate (% gravimetric) by the molecular mass of calcium carbonate (100.087), and multiplying the result by the molecular mass of carbon (12.011). This calculation is performed on all samples.

Table 1. USDA NRCS method code (Soil Survey Staff 2004) for soil physical properties (Megapit) analytical variables. Variable corresponds to the values in the **fieldName** column in the data product.

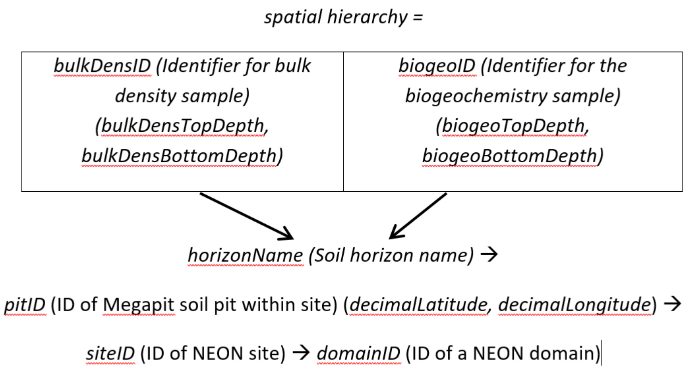
Variable	**Method code**
bulkDensExclCoarseFrag	3B6a
coarseFrag2To5	3A2a1
coarseFrag5To20	3A2a1
sandTotal	3A1a1b
siltTotal	3A1a1b
clayTotal	3A1a1b
carbonateClay	3A1a1b
clayFineContent	3A1a1b

Variable	Method code
siltFineContent	3A1a1b
siltCoarseContent	3A1a1b
sandVeryFineContent	3A1a1b
sandFineContent	3A1a1b
sandMediumContent	3A1a1b
sandCoarseContent	3A1a1b
sandVeryCoarseContent	3A1a1b
cTot	4H2a1
nTot	4H2a2
sTot	4H2a3
airDryOvenDry	3D1
alMjelm	4H1b1a1a1
caMjelm	4H1b1a1a2
feMjelm	4H1b1a1a3
kMjelm	4H1b1a1a4
mgMjelm	4H1b1a1a5
mnMjelm	4H1b1a1a6
naMjelm	4H1b1a1a7
pMjelm	4H1b1a1a8
siMjelm	4H1b1a1a9
srMjelm	4H1b1a1a12
tiMjelm	4H1b1a1a11
zrMjelm	4H1b1a1a10
phCacl2	4C1a2a2b1
phH2o	4C1a2a1b1
ec12pre	4F1a1a
gypsum	4E2a1a1
caco3	4E1a1a1
caNh4d	4B1a1b1
kNh4d	4B1a1b3
mgNh4d	4B1a1b2
naNh4d	4B1a1b4
cecdNh4	4B1a1a1a1
alSatCecd33	4B4d1
baseSumCecd10	4B4a1

Variable	Method code
bsesatCecd10	4B4c1
ececCecd33	4B4b2
alKcl	4B3a1a1
feKcl	Not analyzed
mnKcl	4B3a1a2
bSatx	Not analyzed
brSatx	4F2c1b1a1
caSatx	4F2c1a1
clSatx	4F2c1b1a3
co3Satx	4F2c1c1a1
ecSatp	4F2b1
flSatx	4F2c1b1a4
h2oSatx	4F2a1
hco3Sx	4F2c1c1a2
kSatx	4F2c1a3
mgSatx	4F2c1a2
naSatx	4F2c1a4
no2Satx	4F2c1b1a6
no3Satx	4F2c1b1a5
pSatx	4F2c1b1a7
phSp	4C1a1a2
resist	Not analyzed
so4Satx	4F2c1b1a8
estimatedOC	Not applicable

3.5 Spatial Resolution and Extent

The finest spatial resolution at which Soil physical and chemical properties, Megapit will be tracked is per sample from a defined depth increment within a soil horizon identified within a Megapit soil pit (one Megapit soil pit per NEON terrestrial core and relocatable site, number of soil horizons per Megapit varies among sites). The Soil physical and chemical properties, Megapit data will be generated for a single horizontal location at each NEON terrestrial core and relocatable site (i.e., the location of the Megapit). Data records will be generated for each soil horizon identified in the Megapit; therefore, the data records will be generated at several depth increments with the number of depth increments depending on the number of soil horizons. The data produced are assumed to be representative of the entire depth increment of the soil horizon from which the soil sample was collected. The data products represent a point in horizontal space. Overall, this results in a spatial hierarchy of:



The basic spatial data included in the data downloaded include the latitude, longitude, and elevation of the pit where sampling occurred + associated uncertainty due to GPS error.

3.6 Temporal Resolution and Extent

The finest resolution at which temporal data are reported is the **collectDate**, the date when the sample was collected. The Soil physical and chemical properties, Megapit data product will be generated once for each soil horizon at each NEON core and relocatable site. They represent a point in time, although many of the soil properties typically change fairly slowly (e.g., over decades).

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. For code resources to work with these files, see Data Relationships (3.9).

3.7 Associated Data Streams

The Soil physical and chemical properties, Megapit data product is directly linked by the variable **pitID** to the Root biomass and chemistry, Megapit (DP1.00066.001) data product, as these samples are collected from the same soil Megapit over the same time period. However, linking the two types of data by depth increment requires some discretion as sub-sampling for soil chemical and physical properties is carried out per soil horizon while roots are sampled at fixed depth intervals.

Data from the soil Megapits can be used to inform observations from the TIS soil sensors as well as other ecological and biogeochemistry observational data. To explore any such linkages, use the **siteID** variable to join across data products.

Users may wish to compare soil physical and chemical properties from Megapits to those measured in a distributed initial soil characterization effort. The latter are reported as part of the Soil physical and chemical properties, distributed initial characterization (DP1.10047.001) data product. To help cross-reference the different term names and measurements across these two products, there is a mapping file located in the NEON Data Portal document library (http://data.neonscience.org/documents), in the Data Product User Guides folder.

3.8 Product Instances

There is one Megapit at each NEON terrestrial core and relocatable site. The number of soil horizons identified at each pit will vary among the sites. As relocatable TIS sites are moved, new Megapit data will be generated.

3.9 Data Relationships

Each soil pit sampled yields a unique **pitID** in the mgp_permegapit table. A record from mgp_permegapit then has several child records, one for each horizon in the pit, in mgp_perhorizon. Each **horizonID** record from mgp_perhorizon will then have one to several records in mgp_perbulksample and mgp_perbiogeosample. More than one record is expected per horizon if audit samples were taken. Additionally, each horizon record from mgp_perhorizon should have one record in the mgp_perarchivesample table. 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*.

Data from Soil physical and chemical properties, Megapit can be linked to other data generated from samples collected from the same soil pit, such as the soil water content sensor calibration equation and data generated from analysis of Megapit Soil Archive samples by archive users, using the variable names outlined below.

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

mgp_perhorizon.csv - > One record expected per **pitID** per **horizonID** combination.

mgp_perbulksample.csv - > One record expected per **horizonID** per **bulkDensID** combination. Multiple records will exist for a horizon where bulk density audit samples were taken. For horizons where no samples were taken, there will still be a record for the horizon, but the remaining portion of the record will be left blank.

mgp_perbiogeosample.csv - > One record expected per **horizonID** per **biogeoID** combination. Multiple records will exist for a horizon where biogeochemistry audit samples were taken. For horizons where no samples were taken, there will still be a record for the horizon, but the remaining portion of the record will be left blank.

mgp_perarchivesample.csv - > One record expected per **horizonID** per **archiveID** combination.

Sample identifiers (**bulkDensID**, **biogeoID** and **archiveID**) will be generated for each collection event and analysis group within a sample. Only samples collected for archive will be retained; in all other cases the physical sample will be discarded following analysis.

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

4 DATA QUALITY

4.1 Data Entry Constraint and Validation

A 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 properties (megapit) (DP0.00096.001), provided with every download of this data product. 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[10]).

4.2 Data Processing Steps

Following data entry 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[09]).

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 record is a quality flag for known issues applying to the record, added by NEON Science upon data review. At present, there are no known dataQF entries for this data product.

4.5 Analytical Facility Data Quality

The analytical labs that generate Soil physical and chemical properties, Megapit include standards or reference materials run alongside NEON samples in order to gauge run acceptability. For further information about laboratory QA procedures, refer to the USDA NRCS Soil Survey Laboratory Methods Manual. No external lab data quality flags are reported in this data product, but information on the uncertainty of measurements is given below.

5 UNCERTAINTY

Uncertainty of measurement is inevitable (JCGM 2008, 2012; Taylor 1997). It is crucial that measurement uncertainties are identified and quantified to determine statistical interpretations about mean quantity and variance structure; both are important when constructing higher level data products (e.g., L1 DP) and modeled processes. This portion of the document serves to identify, evaluate, and quantify sources of uncertainty relating to L1 megapit soil property DPs.

5.1 Unquantified Uncertainty

5.1.1 Horizon Identification and characterization

Soil horizons at all NEON Megapit soil pits are identified and characterized by USDA soil scientists (for more information, please refer to Schoeneberger et al. (2012A)). As a result of the subjective nature of this process, identifiable, but currently unquantifiable uncertainties are introduced. For instance, soil horizons and associated characteristics (e.g., depth, thickness, etc.) may be quantified differently among a group of soil scientists; an identifiable uncertainty is introduced.

Another identifiable source of uncertainty is spatial representativeness. Given that only one Megapit soil pit is dug at each NEON site, the spatial variance of horizon depth and thickness cannot be quantified at multiple points throughout the site.

Some aspects of this uncertainty may be gleaned from the soil horizon description, which includes a section describing the boundary between that horizon and the horizon below it (Soil Survey Division Staff 1993). Soil descriptions for the Megapits can be found at: http://data.neonscience.org/megapit-images. The boundary description consists of two components: distinctness and topography.

Distinctness is the depth over which the boundary occurs and is separated into four categories:

- Abrupt: Less than 2 cm thick
- Clear: 2 to 5 cm thick
- Gradual: 5 to 15 cm thick
- Diffuse: More than 15 cm thick

Topography is defined as “the irregularities of the surface that divides the horizons” (Soil Survey Division Staff 1993), and is also separated into four categories:

- Smooth: The boundary is a plane with few or no irregularities.
- Wavy: The boundary has undulations in which depressions are wider then they are deep.
- Irregular: The boundary has pockets that are deeper than they are wide.
- Broken: One or both of the horizons or layers separated by the boundary are discontinuous and the boundary is interrupted.

5.1.2 Sample Collection

Upon completion of the horizon identification and characterization process, each horizon is sampled (collected) independently following the regulations set forth by the National Soil Survey Center (NSSC). Specifically, a sample is collected across a horizon’s full depth and breadth, a method known as Horizon Sampling (Schoeneberger et al. 2012B). Although collecting a sample in such a manner fosters quantification of variability within each horizon, uncertainties from the identification and characterization processes propagate into the collection process and can cause horizons to be improperly sampled. For example, if horizon depths and thicknesses are incorrectly quantified by the soil scientist, the NEON scientist or technician gathering samples may unknowingly combine soils from neighboring horizons into a sample that represents a single horizon. Resulting laboratory analyses may therefore be unrepresentative of the specified horizon of interest. As with the identification and characterization process, this type of uncertainty can only be identified at current date.

5.2 Quantified Uncertainty

Replicate soil samples (audit samples) were collected at a subset of sites across the Observatory to determine the reproducibility of soil property measurements AD[06]. JCGM (2008) defines reproducibility of measurement as repeated measurements of the same or similar objects obtained out a set of conditions, e.g., locations and operators, and repeated measurements of the same or similar objects. Here, we use the phrase measurement reproducibility to represent the collection of replicate samples because multiple soil types were collected under different conditions, e.g., different scientists and geographic locations, but the method of physically collecting the soil with the soil core sampler remained the same.

In addition to the regular sample, audit samples from an individual soil horizon were collected at 13 sites (site code horizon name: SERC Bt3, GUAN C, MAME Bk2, UNDE A, TREE 2E & Bt, KONA Bt2, KONZ Bt2, MLBS 2Cg2, ORNL 2Bt2, NOGP C, MOAB C2, TEAK A1, and WREF Ac). These samples were analyzed to provide a preliminary assessment of soil property reproducibility. Note that the MAME site was descoped after the megapit samples were collected, as a result it could be included in this analysis but no data is available from the NEON Data Portal for this site. The coefficient of variavtion (i.e., standard deviation / mean) was calculated for each soil horizon and each soil property. Since varying numbers of audit samples were collected from the soil horzon at the different sites, the weighted average coefficient of variation was calculated for each soil property (weighted by the number of samples per horizon; Table 2).

Table 2. Relative uncertainty (1 standard deviation) of soil physical properties (Megapit) variables. Variable corresponds to the values in the **fieldName** column in the data product. NA indicates that an analysis was not performed for any of the horizons evaluated.

Variable	Relative uncertainty (%)
bulkDensExclCoarseFrag*	+/-4.82
coarseFrag2To5	+/-42.58
coarseFrag5To20	+/-49.95
sandTotal	+/-5.35
siltTotal	+/-2.68
clayTotal	+/-9.13
carbonateClay	+/-35.88
clayFineContent	+/-NA
siltFineContent	+/-10
siltCoarseContent	+/-12.27

Variable	Relative uncertainty (%)
sandVeryFineContent	+/-7.35
sandFineContent	+/-11.11
sandMediumContent	+/-30.54
sandCoarseContent	+/-59.75
sandVeryCoarseContent	+/-32.34
carbonTot	+/-10
nitrogenTot	+/-43.34
sulfurTot	+/-35.35
estimatedOC	+/-14.97
airDryOvenDry	+/-0.09
alMjelm	+/-5.07
caMjelm	+/-10.38
feMjelm	+/-3.7
kMjelm	+/-2.96
mgMjelm	+/-6.85
mnMjelm	+/-6.52
naMjelm	+/-3.42
pMjelm	+/-20.58
siMjelm	+/-2.46
srMjelm	+/-3.97
tiMjelm	+/-2.91
zrMjelm	+/-4.37
phCacl2	+/-0.83
phH2o	+/-0.77
ec12pre	+/-44.79
gypsumConc	+/-0
caco3Conc	+/-0.43
caNh4d	+/-30.01
kNh4d	+/-23.66
mgNh4d	+/-48.09
naNh4d	+/-0.82
cecdNh4	+/-5.17
alSatCecd33	+/-19
baseSumCecd10	+/-20.53
bsesatCecd10	+/-20.69

Variable	Relative uncertainty (%)
ececCecd33	+/-10.18
alKcl	+/-20
feKcl	+/-NA
mnKcl	+/-48.1
bSatx	+/-NA
brSatx	+/-47.14
caSatx	+/-4.69
clSatx	+/-5.77
co3Satx	+/-0
ecSatp	+/-5.78
flSatx	+/-0
waterSatx	+/-5.81
hco3Sx	+/-49.85
kSatx	+/-15.71
mgSatx	+/-9.98
naSatx	+/-5.37
no2Satx	+/-47.14
no3Satx	+/-47.14
pSatx	+/-0
phSp	+/-0.29
resist	+/-NA
so4Satx	+/-3.96

**Sample D17_TEAK_PIT1_BULKD_BULKD1 was excluded due to unusually low bulk density possibly caused by its higher rock content impacting sample collection.*

6 REFERENCES

Joint Committee for Guides in Metrology (JCGM) (2008) Evaluation of measurement data - Guide to the expression of uncertainty in measurement. pp. 120.

Joint Committee for Guides in Metrology (JCGM) (2012) International vocabulary of metrology - Basic and general concepts and associated terms (VIM). 3rd Edition. pp. 92.

Schoeneberger, P. J., Wysocki D. A., and Chiaretti, J. V., 2012A. Soil Taxonomy. 4.1 - 4.13 in National Resources Conservation Service's National Soil Survey Center, editor. Field book for describing and sampling soils. 3, United States Department of Agriculture, Lincoln, Nebraska, USA.

Schoeneberger, P. J., Wysocki D. A., and Chiaretti, J. V., 2012B. Soil Sampling. 8.1 - 8.7 in National Resources Conservation Service's National Soil Survey Center, editor. Field book for describing and sampling soils. 3, United States Department of Agriculture, Lincoln, Nebraska, USA.

Soil Survey Staff. 2004. Soil Survey Laboratory Methods Manual - Soil Survey Investigations Report No. 42, Version 4.0. USDA Natural Resource Conservation Service.

Taylor, J. R. (1997) An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements. University Science Books, Mill Valley, California. 2nd Ed. pp. 327.