

<i>Title:</i> NEON User Guide to Soil Chemical Properties: Megapit (NEON.DP1.10097)	<i>Date:</i> 02/05/2018
<i>Author:</i> Ed Ayres	<i>Revision:</i> B

## NEON USER GUIDE TO SOIL CHEMICAL PROPERTIES: MEGAPIT (NEON.DP1.10097)

<b>PREPARED BY</b>	<b>ORGANIZATION</b>	<b>DATE</b>
Ed Ayres	FIU	02/05/2018
Robert Lee	FIU	02/05/2018
Josh Roberti	FIU	02/05/2018

<i>Title:</i> NEON User Guide to Soil Chemical Properties: Megapit (NEON.DP1.10097)	<i>Date:</i> 02/05/2018
<i>Author:</i> Ed Ayres	<i>Revision:</i> B

## CHANGE RECORD

<b>REVISION</b>	<b>DATE</b>	<b>DESCRIPTION OF CHANGE</b>
A	06/30/2017	Initial Release
B	02/02/2018	Added data quality and uncertainty sections, as well as other minor updates to the text.

## TABLE OF CONTENTS

<b>1 DESCRIPTION</b>	<b>1</b>
1.1 Purpose . . . . .	1
1.2 Scope . . . . .	1
<b>2 RELATED DOCUMENTS AND ACRONYMS</b>	<b>2</b>
2.1 Associated Documents . . . . .	2
2.2 Acronyms . . . . .	2
<b>3 DATA PRODUCT DESCRIPTION</b>	<b>3</b>
3.1 Spatial Sampling Design . . . . .	3
3.2 Temporal Sampling Design . . . . .	3
3.3 Variables Reported . . . . .	3
3.4 Spatial Resolution and Extent . . . . .	4
3.5 Temporal Resolution and Extent . . . . .	4
3.6 Associated Data Streams . . . . .	4
3.7 Product Instances . . . . .	5
3.8 Data Relationships . . . . .	5
<b>4 DATA QUALITY</b>	<b>5</b>
4.1 Data Entry Constraint and Validation . . . . .	5
4.2 Data Processing Steps . . . . .	6
4.3 Data Revision . . . . .	6
4.4 Quality Flagging . . . . .	6
4.5 Analytical Facility Data Quality . . . . .	6
<b>5 UNCERTAINTY</b>	<b>6</b>
5.1 Unquantified Uncertainty . . . . .	7
5.1.1 Horizon Identification and characterization . . . . .	7
5.1.2 Sample Collection . . . . .	7
5.2 Quantified Uncertainty . . . . .	8
<b>6 REFERENCES</b>	<b>10</b>

## 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, the dry weights of litter functional groups 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 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 file, NEON Data Variables for Soil Chemical Properties: Megapit (NEON.DP1.00097.001) (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 TIS Soil Pit Sampling Protocol (AD[06]). The raw data that are processed in this document are detailed in the file, NEON Raw Data Validation for Soil Chemical Properties: Megapit (NEON.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 Level 1, Level 2 and Level 3 Data Products Catalog
AD[04]	NEON.DOC.002837	NEON Raw Data Validation for Soil Chemical Properties: Megapit (NEON.DP0.00096.001)
AD[05]	NEON.DOC.002840	NEON Data Variables for Soil Chemical Properties: Megapit (NEON.DP1.00097.001)
AD[06]	NEON.DOC.001307	TIS Soil Pit Sampling Protocol
AD[07]	NEON.DOC.000008	NEON Acronym List
AD[08]	NEON.DOC.000243	NEON Glossary of Terms
AD[09]	NEON.DOC.001020	NEON Algorithm Theoretical Basis Document: Soil Physical Properties: Megapits
AD[10]	NA	NEON's Ingest Conversion Language (NICL) specifications

### 2.2 Acronyms

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

### 3 DATA PRODUCT DESCRIPTION

The Soil Chemical Properties: Megapits data product includes data on the total content of a range of chemical elements, pH, and electrical conductivity in the  $\leq 2$  mm soil fraction for each soil horizon sampled in a TIS soil pit. Data were derived from a sampling location expected to be representative of the area where the five Instrumented Soil Plots per site are located and were collected once during site construction. Soil Physical Properties: Megapits data are derived from the same sampling location; NEON's other soil data products (Soil physical properties (Distributed periodic), Soil chemical properties (Distributed periodic), Soil stable isotopes (Distributed periodic), Soil physical properties (Distributed initial characterization), and Soil chemical properties (Distributed initial characterization)) represent distributed sampling across many soil pits, and the products designated as periodic are repeated over the lifetime of the observatory. Soil from most soil pit horizons are archived in the Megapit Soil Archive and are available upon request (<http://www.neonscience.org/request-megapit-soil-samples>).

#### 3.1 Spatial Sampling Design

The Soil Chemical Properties: Megapit data product is available at each terrestrial site. The soil pit location was chosen to be representative of the 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 soil pit is 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 TIS 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

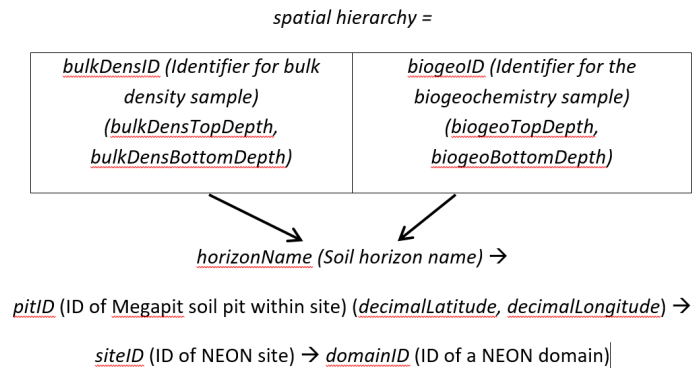
All variables reported from the field or laboratory technician (L0 data) are listed in the file, NEON Raw Data Validation for Soil Chemical Properties: Megapit (NEON.DP0.00096.001) (AD[04]). All variables reported in the published data (L1 data) are also provided separately in the file, NEON Data Variables for Soil Chemical Properties: Megapit (NEON.DP1.00097.001) (AD[05]).

NEON TIS 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 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.

### 3.4 Spatial Resolution and Extent

The finest spatial resolution at which Soil 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 Chemical Properties: Megapit will be generated for a single horizontal location at each NEON terrestrial core and relocatable site (i.e., the location of the Megapit). Separate data products will be generated for each soil horizon identified in the megapit; therefore, the data products will be generated at several depth increments with the number of depth increments depending on the number of soil horizons. The data products 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 and plot width.

### 3.5 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 Chemical Properties: Megapit data products 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 currently 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 here: <https://github.com/NEONScience/NEON-utilities>

### 3.6 Associated Data Streams

**pitID**, **horizonID**, **biogeoID**, and **archiveID** are linking variables that tie specific samples and associated metadata to the Soil Chemical Properties: Megapit: data product (NEON.DP1.00097) and the Soil physical properties:

<i>Title:</i> NEON User Guide to Soil Chemical Properties: Megapit (NEON.DP1.10097)	<i>Date:</i> 02/05/2018
<i>Author:</i> Ed Ayres	<i>Revision:</i> B

Megapit data product (NEON.DP1.00096).

### 3.7 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.

### 3.8 Data Relationships

Data from both Soil Chemical Properties: Megapits and Soil Physical Properties: Megapits are used to inform the data generated by other soil-based sensors at TIS sites, such as the soil water content sensor calibration equation and data generated from analysis of Megapit Soil Archive samples by archive users. The Soil Chemical Properties: Megapit data product is derived from the same parent sample as the Soil Physical Properties derived data product. The parent sample of the Soil Chemical Properties: Megapit and Soil Physical Properties: Megapit (excluding the bulk density, soil taxonomy, and horizon name subproducts) is subsampled from the same parent sample as the Soil Archive: Megapit sample. The Soil Physical Properties: Megapit bulk density subproduct is derived from a soil sample collected from the same soil pit and soil horizon as the Soil Chemical Properties: Megapit sample and the sample that produces the remaining Soil Physical Properties: Megapit data product subproducts. The soil taxonomy and horizons subproducts, and the soil water content sensor calibration equation, relate to the soil pit where the samples for Soil Chemical Properties: Megapit and Soil Physical Properties: Megapit data products were collected.

mgc\_permegapit.csv - > One record expected per **pitID**

mgc\_perhorizon.csv - > One record expected per **pitID** per **horizonID** combination.

mgc\_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.

mgc\_perarchivesample.csv - > One record expected per **horizonID** per **archiveID** combination.

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

## 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 Chemical Properties: Megapit (NEON.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.

## 4.5 Analytical Facility Data Quality

The analytical labs that generate soil chemical properties (Megapit) data include standards or secondary reference materials run alongside NEON samples in order to gauge run acceptability. In general, an entry of 0 in a quality flag field means there is no issue to report.

For further information about individual laboratory QA procedures, refer to the lab-specific SOPs found in the NEON Data Portal document library (<http://data.neonscience.org/documents>), External Lab Protocols section.

## 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 than 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 12 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). 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 variation (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 horizon 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 1).

**Table 1.** Relative uncertainty (1 standard deviation) of soil chemical 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 (%)
carbonTot	±10.9
nitrogenTot	±48.77
sulfurTot	±34.78
estimatedOC	±16.53
airDryOvenDry	±0.07
alMjelm	±4.51
caMjelm	±11.21
feMjelm	±3.54
kMjelm	±2.88
mgMjelm	±6.48
mnMjelm	±6.79
naMjelm	±3.49
pMjelm	±21.91
siMjelm	±2.4
srMjelm	±3.93
tiMjelm	±2.91

Variable	Relative uncertainty (%)
zrMjelm	±4.18
phCacl2	±0.78
phH2o	±0.76
ec12pre	±48.49
gypsumConc	±0
caco3Conc	±0.43
caNh4d	±34.56
kNh4d	±27.25
mgNh4d	±55.38
naNh4d	±0.94
cecdNh4	±4.03
alSatCecd33	±23.7
baseSumCecd10	±23.64
bsesatCecd10	±20.06
ececCecd33	±10.2
alKcl	±22.29
feKcl	±NA
mnKcl	±59.73
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

<i>Title:</i> NEON User Guide to Soil Chemical Properties: Megapit (NEON.DP1.10097)	<i>Date:</i> 02/05/2018
<i>Author:</i> Ed Ayres	<i>Revision:</i> B

<b>Variable</b>	<b>Relative uncertainty (%)</b>
so4Satx	±3.96

## 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.