NEON ALGORITHM THEORETICAL BASIS DOCUMENT (ATBD): SOIL PHYSICAL PROPERTIES: MEGAPIT

PREPARED BY	ORGANIZATION	DATE
Edward Ayres	FIU	06/02/2015
Josh Roberti	FIU	06/02/2015
Janae Csavina	CVAL	06/02/2015

APPROVALS	ORGANIZATION	APPROVAL DATE
Andrea Thorpe	SCI	11/13/2015
Dave Tazik	SCI	11/02/2015
Hanne Buur	PSE	10/20/2015

RELEASED BY	ORGANIZATION	RELEASE DATE
Jennifer DeNicholas	CM	12/08/2015

See configuration management system for approval history.

The National Ecological Observatory Network is a project solely funded by the National Science Foundation and managed under cooperative agreement by NEON, Inc. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Change Record

REVISION	DATE	ECO #	DESCRIPTION OF CHANGE
A	12/08/2015	ECO-03235	Initial release

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

1 DESCRIPTION

Various physical and chemical properties will be measured on soil samples collected by soil horizon from the Terrestrial Instrument System (TIS) soil pit, known as the Megapit. This represents a one-time sampling effort, normally during site construction, at all NEON core and relocatable sites. This document details the processes necessary to convert “raw” measurements into meaningful scientific units and their associated uncertainties are described. The soil properties data will be generated by an external laboratory.

1.1 Purpose

This document details the algorithms used for creating the NEON Level 1 data products from Level 0 data, and ancillary data as defined in this document (such as calibration data), obtained from analyses at an external laboratory. It includes a detailed discussion of measurement theory and implementation, appropriate theoretical background, data product provenance, quality assurance and control methods used, approximations and/or assumptions made, and a detailed exposition of uncertainty resulting in a cumulative reported uncertainty for this product.

1.2 Scope

The theoretical background and entire algorithmic process used to derive Level 1 TIS Soil Properties data product from Level 0 data are describe in this document. The scope of the document is limited to soil physical properties measured on soil samples collected from the TIS Soil Pit (Megapit). This document assumes the soil samples will be analyzed in accordance with the USDA NRCS Soil Survey Laboratory Methods Manual (Soil Survey Staff 2004). This document does not provide computational implementation details, except for cases where these stem directly from algorithmic choices explained here.

2 RELATED DOCUMENTS AND ACRONYMS

2.1 Applicable Documents

AD[01]	NEON.DOC.000001 NEON Observatory Design (NOD) Requirements
AD[02]	NEON.DOC.005003 NEON Scientific Data Products Catalog
AD[03]	NEON.DOC.005004 NEON Level 1-3 Data Products Catalog
AD[04]	NEON.DOC.005005 NEON Level 0 Data Product Catalog
AD[05]	NEON.DOC.000782 ATBD QA/QC Data Consistency
AD[06]	NEON.DOC.011081 ATBD QA/QC plausibility tests
AD[07]	NEON.DOC.002839 NEON Data Publication Workbook for Soil Physical Properties (Megapit)
AD[08]	NEON.DOC.000746 Evaluating Uncertainty (CVAL)
AD[12]	NEON.DOC.000784 ATBD Profile Development
AD[13]	NEON.DOC.000785 TIS Level 1 Data Products Uncertainty Budget Estimation Plan
AD[14]	NEON.DOC.000751 CVAL Transfer of standard procedure
AD[15]	NEON.DOC.000747 Uncertainty Assessment Methodologies (CVAL)
AD[16]	NEON.DOC.003037 ATBD Soil Chemical Properties: Megapit
AD[17]	NEON.DOC.001307 TIS Soil Pit Sampling Protocol

2.2 Reference Documents

RD[01]	NEON.DOC.000008	NEON Acronym List
RD[02}	NEON.DOC.000243	NEON Glossary of Terms
RD[03]		
RD[04]		

2.3 Acronyms

Acronym	Explanation
ATBD	Algorithm Theoretical Basis Document
CI	Cyberinfrastructure Product Team
CVAL	NEON Calibration, Validation, and Audit Laboratory
DAS	Data Acquisition System
DP	Data Product
FIU	Fundamental Instrument Unit Product Team
GRAPE	Grouped Remote Analog Peripheral Equipment
L0	Level 0
L1	Level 1
UQ	Unquantifiable uncertainty
USDA	United States Department of Agriculture
NRCS	Natural Resource Conservation Service

2.4 Verb Convention

"Shall" is used whenever a specification expresses a provision that is binding. The verbs "should" and "may" express non-mandatory provisions. "Will" is used to express a declaration of purpose on the part of the design activity.

3 DATA PRODUCT DESCRIPTION

3.1 Variables Reported

This ATBD describes the steps needed to generate the L1 data products: Soil Physical Properties: Megapit (NEON.DOM.SIT.DP1.00096). Subproducts for this data product are listed below (Table 1). Detailed lists of the associated subproducts and metadata products are provided separately, along with example data in publication-ready spreadsheets in the NEON Data Publication Workbook for Soil Physical Properties (Megapit) (AD[07]).

Table 3-1. List of subproducts produced in this ATBD for the data product, Soil Physical Properties: Megapit (NEON.DOM.SIT.DP1.00096). The list is not exhaustive and a variety of supporting data will also be made available.

Number	Field Name	Description
NEON.DOM.SITE.DP1.00096.001.01179.001.001.001	soilSeries	Soil taxonomy at the series level
NEON.DOM.SITE.DP1.00096.001.01180.001.001.001	soilFamily	Soil taxonomy at the family level
NEON.DOM.SITE.DP1.00096.001.01181.001.001.001	soilSubgroup	Soil taxonomy at the subgroup level
NEON.DOM.SITE.DP1.00096.001.01182.001.001.001	soilGreatGroup	Soil taxonomy at the great group level
NEON.DOM.SITE.DP1.00096.001.01183.001.001.001	soilSuborder	Soil taxonomy at the suborder level
NEON.DOM.SITE.DP1.00096.001.01184.001.001.001	soilOrder	Soil taxonomy at the order level
NEON.DOM.SITE.DP1.00096.001.01187.002.001.001	horizonName	Soil horizon name
NEON.DOM.SITE.DP1.00096.001.01188.002.001.001	horizonTopDepth	Depth below the soil surface of the top of a soil horizon
NEON.DOM.SITE.DP1.00096.001.01189.002.001.001	horizonBottomDepth	Depth below the soil surface of the bottom of a soil horizon
NEON.DOM.SITE.DP1.00096.001.01336.003.001.001	bulkDensLT2mm	Bulk density of the <2 mm size fraction
NEON.DOM.SITE.DP1.00096.001.01340.004.001.001	coarseFrag5-20	Coarse fragment (2-5 mm) content of the <20 mm size fraction of the biogeochemistry soil sample
NEON.DOM.SITE.DP1.00096.001.01341.004.001.001	coarseFrag2-5	Coarse fragment (5-20 mm) content of the <20 mm size fraction of the biogeochemistry soil sample
NEON.DOM.SITE.DP1.00096.001.01275.004.001.001	sandTot	Total sand (0.047-2 mm) content of the <2 mm fraction
NEON.DOM.SITE.DP1.00096.001.01276.004.001.001	siltTot	Total silt (0.002-0.047 mm) content of the <2 mm fraction

NEON.DOM.SITE.DP1.00096.001.01277.004.001.001	clayTot	Total clay (<0.002 mm) content of the <2 mm fraction
NEON.DOM.SITE.DP1.00096.001.01279.004.001.001	co3Clay	Carbonate clay (<0.002 mm) content of the <2 mm fraction
NEON.DOM.SITE.DP1.00096.001.01280.004.001.001	siltFine	Fine silt (0.002-0.02 mm) content of the <2 mm fraction
NEON.DOM.SITE.DP1.00096.001.01281.004.001.001	siltCoarse	Coarse silt (0.02-0.047 mm) content of the <2 mm fraction
NEON.DOM.SITE.DP1.00096.001.01282.004.001.001	sandCoarse	Coarse sand (0.5-1 mm) content of the <2 mm fraction
NEON.DOM.SITE.DP1.00096.001.01283.004.001.001	sandFine	Fine sand (0.105-0.25 mm) content of the <2 mm fraction
NEON.DOM.SITE.DP1.00096.001.01284.004.001.001	sandMedium	Medium sand (0.25-0.5 mm) content of the <2 mm fraction
NEON.DOM.SITE.DP1.00096.001.01285.004.001.001	sandVeryCoarse	Very coarse sand (1-2 mm) content of the <2 mm fraction
NEON.DOM.SITE.DP1.00096.001.01286.004.001.001	sandVeryFine	Very fine sand (0.047-0.105 mm) content of the <2 mm fraction
NEON.DOM.SITE.DP1.00096.001.01337.003.001.001	porosity	Fraction of soil volume comprised of voids
NEON.DOM.SITE.DP1.00096.001.01289.005.001.001	archiveID	An identifier for the archive sample

3.2 Product Instances

Soil Physical Properties: Megapit data products will be produced for each soil horizon identified in the TIS soil pit (Megapit), unless this is not appropriate because of methodological constraints (see Theory of Measurement/Observation section). 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.

3.3 Temporal Resolution and Extent

The finest temporal resolution that Soil Physical Properties: Megapit data will be tracked is at the level of a day. The Soil Physical 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).

3.4 Spatial Resolution and Extent

The finest spatial resolution at which Soil Physical Properties: Megapit will be tracked is per sample from a defined depth increment within a soil horizion 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 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:

spatial hierarchy =

bulkDensID (Identifier for bulk density sample) (bulkDensTopDepth, bulkDensBottomDepth)	*biogeoID (Identifier for the biogeochemistry sample) (biogeoTopDepth, biogeoBottomDepth)*

horizonName (Soil horizon name) →

pitID (ID of Megapit soil pit within site) (*decimalLatitude, decimalLongitude*) →

siteID (ID of NEON site) → *domainID* (ID of a NEON domain)

3.5 Associated Data Streams

The Soil Physical Properties: Megapit data product is directly linked to the TIS Soil Chemical Properties: Megapit (NEON.DOM.SIT.DP1.00097) data products, as well as 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. The Soil Physical Properties: Megapit data product (excluding the bulk density, soil taxonomy, and horizon name subproducts) is derived from the same parent sample as the Soil Chemical 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 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.

4 SCIENTIFIC CONTEXT

4.1 Theory of Measurement/Observation

The methods used to quantify the laboratory-based data products included in this ATBD are described in the USDA NRCS Soil Survey Laboratory Methods Manual (Soil Survey Staff 2004), while the methods used to generate the soil taxonomy, soil horizon name, and soil horizon depths are described in the Soil Survey Manual (Soil Survey Division Staff 1993). Method codes from the Soil Survey Laboratory Methods Manual for each data product are listed in Table 3 in the Appendix. See the USDA NRCS Soil Survey Laboratory Methods Manual for a description of each method.

Because the primary users of the data products included in this ATBD 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.

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 (order, suborder, great group, subgroup, family, and series), horizon name, and horizon depths 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.

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), total silt (0.002-0.047 mm), total clay (<0.002), fine silt (0.002-0.02 mm), coarse silt (0.02-0.05 mm), very fine sand (0.047-0.105 mm), fine sand (0.105-0.25 mm), medium sand (0.25-0.5 mm), coarse sand (0.5-1 mm), and very coarse sand (1-2 mm). This analysis is performed on samples that were from all soil horizons except Oi.

Carbonate clay content of the <2 mm soil fraction is determined by adding hydrochloric acid to the residue from the particle-size distribution analysis and measuring the amount of CO_2 evolved. This analysis is performed on samples that have effervescence greater than or equal to “slight” as defined by NRCS.

The proportion of rock fragments in 2-5 mm and 5-20 mm size classes 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.

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. This analysis is performed on all samples.

The density of >2 mm fragments from the bulk density sample is calculated by dividing the oven dry weight by the volume of the fragments determined using a gas pycnometer. This analysis is not performed on the bulk density samples collected from the first Megapits, in which case a particle density of 2.65 g cm^{-3} for the >2 mm fragments is assumed. This data product is produced for all bulk density samples that contained >2 mm rock fragments.

4.2 Theory of Algorithm

4.2.1 Summary of algorithm for the mgp_permegapit data

1. Generate **uid**

4.2.2 Summary of algorithm for the mgp_perhorizon data

1. Generate **uid**

4.2.3 Summary of algorithm for the mgp_perbulksample data

1. Generate **uid**
2. Calculate **bulkDensLT2mm** using algorithm described below
3. Perform range test on **bulkDensLT2mm** and flag data accordingly
4. Calculate **porosity** using algorithm described below
5. Perform range test on **porosity** and flag data accordingly

4.2.4 Summary of algorithm for the mgp_perbiogeosample data

1. Generate **uid**
2. Calculate **coarseFrag5-20**, **coarseFrag2-5**, **sandTot**, **siltTot**, **clayTot**, **co3Clay**, **siltFine**, **siltCoarse**, **sandCoarse**, **sandFine**, **sandMedium**, **sandVeryCoarse**, and **sandVeryFine** using the algorithms described below
3. Perform range test on **coarseFrag5-20**, **coarseFrag2-5**, **sandTot**, **siltTot**, **clayTot**, **co3Clay**, **siltFine**, **siltCoarse**, **sandCoarse**, **sandFine**, **sandMedium**, **sandVeryCoarse**, and **sandVeryFine** and flag data accordingly
4. Perform sum tests on **sandTot**, **siltTot**, **clayTot**, **siltFine**, **siltCoarse**, **sandCoarse**, **sandFine**, **sandMedium**, **sandVeryCoarse**, **sandVeryFine** and flag data accordingly

4.2.5 Summary of algorithm for mgp_perarchivesample data

1. Generate **uid**

5 ALGORITHM IMPLEMENTAT ION

Unless otherwise specified below, all variables that appear in tables mgp_permegapit_pub, mgp_perhorizon_pub, mgp_perbulksample_pub, mgp_perbiogeosample_pub, and mgp_perarchivesample_pub can be passed directly from the L0 or CI datastore to the L1 variable with the same name.

5.1 Run the following processing steps for NEON Raw Data Ingest Workbook for Soil Physical Properties (Megapit) (mgp_permegapit)

None

5.2 Run the following processing steps for NEON Raw Data Ingest Workbook for Soil Physical Properties (Megapit) (mgp_perhorizon)

None

5.3 Run the following processing steps for NEON Raw Data Ingest Workbook for Soil Physical Properties (Megapit) (mgp_perbulksample)

1. Calculate **bulkDensLT2mm**
   - Do not calculate **bulkDensLT2mm** if **bulkDensSampleType** = Audit or **biogeoSampleType** = Audit for the same **pitID** and **horizonName**.
   - If **bulkDensHorizonProportion** = 1 for the same **pitID** and **horizonName**, calculate **bulkDensLT2mm** as follows:

$$\rho_b = \frac{\left(BD_{weight} - BD_{>2weight}\right)}{\left(BD_{volume} - \left(\frac{BD_{>2weight}}{\rho_m}\right)\right)} \quad (1)$$

where

ρ_b is **bulkDensLT2mm**
BD_{weight} is **bulkDensDryWeight**
$BD_{>2weight}$ is **bulkDensCoarseFragWeight**,
BD_{volume} is **bulkDensVolume**,
ρ_m is **bulkDensCoarseFragDens**.

- If **bulkDensHorizonProportion** < 1 for the same **pitID** and **horizonName**, calculate **bulkDensLT2mm** as follows:

$$\rho_b = P_i \frac{\left(BD_{weight_i} - BD_{>2weight_i}\right)}{\left(BD_{volume_i} - \left(\frac{BD_{>2weight_i}}{\rho_{m_i}}\right)\right)} + P_j \frac{\left(BD_{weight_j} - BD_{>2weight_j}\right)}{\left(BD_{volume_j} - \left(\frac{BD_{>2weight_j}}{\rho_{m_j}}\right)\right)} \quad (2)$$

where

ρ_b is **bulkDensLT2mm**
BD_{weight} is **bulkDensDryWeight**
$BD_{>2weight}$ is **bulkDensCoarseFragWeight**,
BD_{volume} is **bulkDensVolume**,
ρ_m is **bulkDensCoarseFragDens**.
P is **bulkDensHorizonProportion**
i corresponds to the first soil sample within the same **pitID** and **horizonName**
j corresponds to the second soil sample within the same **pitID** and **horizonName**.

2. Perform range test on **bulkDensLT2mm**
   - Create quality flag field **bulkDensLT2mmQF** and populate with zeroes
     Check whether **bulkDensLT2mm** is within the range the range specified below and assign the quality flag accordingly.
     - For samples with a **horizonName** where the first letter, not necessarily the first character, is "O" or "Litter" set **bulkDensLT2mmQF** to 1 if **bulkDensLT2mm** is outside the range 0.01-1.20.
     - For samples with a **horizonName** where the first letter is "A", "B", "E", or "C" set **bulkDensLT2mmQF** to 1 if **bulkDensLT2mm** is outside the range 0.15-2.30.
     - If **bulkDensLT2mm** is within the range specified above, leave **bulkDensLT2mmQF** set to zero.
3. Calculate **porosity**
   - Do not calculate **porosity** if **bulkDensSampleType** = Audit or **biogeoSampleType** = Audit for the same **pitID** and **horizonName**.
   - If **bulkDensHorizonProportion** = 1 for the same **pitID** and **horizonName**, calculate **porosity** as follows using an assumed particle density of 2.65 g cm^{-3}, which is typical of most soils (Soil Survey Staff 2004):

$$\phi = 1 - \left(\frac{\rho_b}{2.65}\right) \quad (3)$$

where

ϕ is **porosity**
ρ_b is **bulkDensLT2mm**
2.65 is the assumed particle density of the <2 mm soil fraction

4. Perform range test on **porosity**

- Create quality flag field **porosityQF** and populate with zeroes
  Check whether **porosity** is within the range the range specified below and assign the quality flag accordingly.
  - For samples with a **horizonName** where the first letter, not necessarily the first character, is "O" or "Litter" set **porosityQF** to 1 if **porosity** is outside the range 0.547-0.997.
  - For samples with a **horizonName** where the first letter is "A", "B", "E", or "C" set **porosityQF** to 1 if **porosity** is outside the range 0.132-0.943.
  - If **porosity** is within the range specified above, leave **porosityQF** set to zero.

5.4 Run the following processing steps for NEON Raw Data Ingest Workbook for Soil Physical Properties (Megapit) (mgp_perbiogeosample)

1. Calculate **sandTot**
   - Do not calculate **sandTot** if **biogeoSampleType** = Audit or **bulkDensSampleType** = Audit for the same **pitID** and **horizonName**.
   - If **biogeoHorizonProportion** = 1 for the same **pitID** and **horizonName**, calculate **sandTot** as follows:

$$Z_{L1} \equiv Z_{L0} \quad (4)$$

where

Z_{L0} is Level 0 **sandTot**
Z_{L1} is Level 1 **sandTot**

   - If **biogeoHorizonProportion** < 1 for the same **pitID** and **horizonName**, calculate **sandTot** as follows:

$$Z_{L1} = Z_{L0_i}P_i + Z_{L0_j}P_j \quad (5)$$

where

Z_{L0} is Level 0 **sandTot**
P is **biogeoHorizonProportion**
Z_{L1} is Level 1 **sandTot**
i corresponds to the first soil sample within the same **pitID** and **horizonName**
j corresponds to the second soil sample within the same **pitID** and **horizonName**.

2. Calculate **siltTot**, **clayTot**, **sandVeryCoarse**, **sandCoarse**, **sandMedium**, **sandFine**, **sandVeryFine**, **siltCoarse**, **siltFine**, and **co3Clay**
   - Use the **sandTot** algorithm described above, replacing **sandTot** with the relevant parameter, to calculate each of the following Level 1 data products: **siltTot**, **clayTot**, **sandVeryCoarse**, **sandCoarse**, **sandMedium**, **sandFine**, **sandVeryFine**, **siltCoarse**, **siltFine**, and **co3Clay**
3. Perform sum test on Level 1 **sandTot**, **siltTot**, and **clayTot**
   - Create quality flag field **textureQF** and populate with zeroes
   - Set **textureQF** to 1 if:

$$99.9 > sandTot + siltTot + clayTot$$
or
$$100.1 < sandTot + siltTot + clayTot \quad (6)$$

4. Perform sum test on **sandVeryCoarse**, **sandCoarse**, **sandMedium**, **sandFine**, and **sandVeryFine**
   - Create quality flag field **sandQF** and populate with zeroes
   - Set **sandQF** to 1 if:

$$sandTot - 0.1 > sandVeryCoarse + sandCoarse + sandMedium + sandFine + sandVeryFine$$
or
$$sandTot + 0.1 < sandVeryCoarse + sandCoarse + sandMedium + sandFine + sandVeryFine \quad (7)$$

5. Perform sum test on Level 1 **siltCoarse** and **siltFine**
   - Create quality flag field **siltQF** and populate with zeroes
   - Set **siltQF** to 1 if:

$$siltTot - 0.1 > siltFine + siltCoarse$$
or
$$siltTot + 0.1 < siltFine + siltCoarse \quad (8)$$

6. Calculate **coarseFrag5To20**
   - Do not calculate **coarseFrag5To20** if **bulkDensSampleType** = Audit or **biogeoSampleType** = Audit for the same **pitID** and **horizonName**.
   - If **biogeoHorizonProportion** = 1 for the same **pitID** and **horizonName**, calculate **coarseFrag5To20** as follows:

$$R_{5-20} = \frac{R_{5-20weight}}{(B_{weight} - R_{20-75weight})/1000} \quad (9)$$

where

R_{5-20} is **coarseFrag5To20**,
$R_{5-20weight}$ is **biogeoTotWeight5To20**
B_{weight} is **biogeoTotWeight**
$R_{20-75weight}$ is **biogeoTotWeight20To75**

   - If **biogeoHorizonProportion** < 1 for the same **pitID** and **horizonName**, calculate **coarseFrag5To20** as follows:

$$R_{5-20} = P_i\left(\frac{R_{5-20weight_i}}{\left(B_{weight_i} - R_{20-75weight_i}\right)/1000}\right) + P_j\left(\frac{R_{5-20weight_j}}{\left(B_{weight_j} - R_{20-75weight_j}\right)/1000}\right) \quad (10)$$

where

R_{5-20} is **coarseFrag5To20**
$R_{5-20weight}$ is **biogeoTotWeight5To20**

B_{weight} is **biogeoTotWeight**
$R_{20-75weight}$ is **biogeoTotWeight20To75**
P is **biogeoHorizonProportion**
i corresponds to the first soil sample within the same **pitID** and **horizonName**
j corresponds to the second soil sample within the same **pitID** and **horizonName**.

7. Calculate **coarseFrag2To5**
   - Do not calculate **coarseFrag2To5** if **bulkDensSampleType** = Audit or **biogeoSampleType** = Audit for the same **pitID** and **horizonName**.
   - If **biogeoHorizonProportion** = 1 for the same **pitID** and **horizonName**, calculate **coarseFrag2To5** as follows:

$$R_{2-5} = \frac{R_{2-5weight}}{\left(B_{weight} - R_{20-75weight}\right)/1000} \quad (11)$$

where

R_{2-5} is **coarseFrag2To5**,
$R_{2-5weight}$ is **biogeoTotWeight2To5**
B_{weight} is **biogeoTotWeight**
$R_{20-75weight}$ is **biogeoTotWeight20To75**

   - If **biogeoHorizonProportion** < 1 for the same **pitID** and **horizonName**, calculate **coarseFrag2To5** as follows:

$$R_{2-5} = P_i\left(\frac{R_{2-5weight_i}}{\left(B_{weight_i} - R_{20-75weight_i}\right)/1000}\right) + P_j\left(\frac{R_{2-5weight_j}}{\left(B_{weight_j} - R_{20-75weight_j}\right)/1000}\right) \quad (12)$$

where

R_{2-5} is **coarseFrag2To5**
$R_{2-5weight}$ is **biogeoTotWeight2To5**
B_{weight} is **biogeoTotWeight**
$R_{20-75weight}$ is **biogeoTotWeight20To75**
P is **biogeoHorizonProportion**
i corresponds to the first soil sample within the same **pitID** and **horizonName**
j corresponds to the second soil sample within the same **pitID** and **horizonName**.

5.5 Run the following processing steps for NEON Raw Data Ingest Workbook for Soil Physical Properties (Megapit) (mgp_perarchivesample)

None

6 UNCERTAINTY

The uncertainty section of this ATBD is not being implemented during the initial release of data derived using this ATBD.

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 soil property DPs. It is a reflection of the information described in AD[13], and is explicitly described for the various soil properties detailed in the following sections.

6.1 Uncertainty of Soil Sampling and laboratory analyses

Because of the subjective nature of humans and thus, the identification, characterization, and sampling of soil horizons, many sources of uncertainties can be associated with soil properties. These uncertainties introduced by humans can be easily identified but properly quantifying them can be problematic. The aim of the following sections is to identify all known uncertainties inherent in the previously mentioned processes, and if possible, quantify each.

6.1.1 Reported Uncertainty

It should be noted that reported measurement uncertainties outlined in this document comprise *only quantifiable measurement uncertainties*. Quantifiable uncertainties will be computed annually by FIU via the equations in the Appendix (Section 9.2). Once computed, these values will pass from FIU to CI and accompany the measurements discussed in Sections 6.1.4 through 6.1.6. It is urged that the end-user(s) acknowledge that many uncertainties, such as those mentioned in Sections 6.1.2 through 6.1.3 are unquantifiable at this time, and reported uncertainties are most likely underestimated as a result.

6.1.2 Horizon Identification and characterization

Soil horizons at all NEON soil pits are identified and characterized by local soil scientists, usually from USDA NRCS (for more information, please refer to Schoeneberger *et al.* (2012A) and AD[17]). 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. One way to mitigate these issues and *quantify* this uncertainty would be to gather a group ($n \geq 30$) of soil scientists, and have each characterize the horizons of the same soil pit independently of one another. Uncertainty (in the form of *reproducibility*) could then be obtained through this approach. However, such an approach is time and labor intensive and is therefore not completed.

Another identifiable source of uncertainty is spatial representativeness. Given that *only one* 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.

Both examples presented in this section represent identification and characterization uncertainties (of soil horizons) that can be identified but cannot be quantified at this time.

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). 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.

6.1.3 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 (see Section 6.1.1) 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 (e.g., bulk density and biogeochemistry) 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.

6.1.4 Particle-Size Distribution Analyses

The procedures and equations associated with particle-size distribution analysis (PSDA) can be found in the NRCS Soil Survey Laboratory Methods Manual (Soil Survey Staff 2004). Because traceable Particle-size distribution standards do not exist, the trueness of the sample cannot be quantified. Uncertainty evaluations for PSDA procedures and equations are given *only* in the form of repeatability and reproducibility.

6.1.5 Bulk Density Analysis

Bulk density samples are collected at each NEON site primarily via soil corer (please refer to AD[17] for more information). Contrary to the clod method where samples are equilibrated at a constant water tension, the water tension of coring samples cannot be controlled and represent the soil moisture at the time of sampling, and thus, the soil corer method and subsequent analyses are prone to additional uncertainties (Rich Ferguson and Ellis Benham, NRCS-USDA, personal contact 2013) Additionally, because only *one* sample is collected at each horizon the reproducibility of moisture variation cannot be quantified.

A traceable reference material is not available for this analysis and therefore the trueness of the sample cannot be quantified. The only uncertainty metric reported for this data product is *repeatability*.

6.2 Uncertainty Budget

The uncertainty budget is a visual aid detailing i) quantifiable sources of uncertainty, ii) means by which they are derived, and iii) the order of their propagation. Individual uncertainty values denoted in this budget are either provided here (within this document) or will be provided by other NEON teams (e.g., FIU) and stored in the CI data store.

Table 6-1. Uncertainty budget for L1 Soil Property DPs. Shaded rows denote the order of uncertainty propagation (from lightest to darkest).

Source of uncertainty	Standard uncertainty component $u(x_i)$	Value of standard uncertainty	$c_i \equiv \frac{\partial f}{\partial x_i}$	$u_i(Y) \equiv \lvert c_i \rvert u(x_i)$
Particle-size distribution	$u_c(PSDA)$	Eq. (1)	--	--
Repeatability	u_{repeat}	Eq. (3) or (4)	1	Eq. (3) or (4)
Reproducibility	u_{reprod}	Eq. (5) or (6)	1	Eq. (5) or (6)

Bulk Density	$u_c(Bulk)$	Eq. (1)	--	--
Repeatability	u_{repeat}	Eq. (3) or (4)	1	Eq. (3) or (4)

7 FUTURE MODIFICATIONS AND PLANS

The uncertainty budget will be implemented in the future. The approach for quantifying the uncertainty for the data products in this ATBD may be revised in the future based on the amount and type of data received from the analytical laboratory. Consistency tests and flags may be added to the QA/QC procedures for some or all of the data products in this ATBD. Thresholds for the QA/QC procedures may be changed and additional QA/QC procedures may be added.

An Audit Test may be added. This will consist of a routine audit with reference material and/or analysis of lab protocols performed on the external lab providing data by NEON's CVAL team to determine whether data passes or fails this test. If the lab is found to be out of compliance with stated uncertainty values or NEON protocols, incoming data will be flagged until lab proves restored compliance.

8 BIBLIOGRAPHY

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

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 Division Staff. 1993. Soil Survey Manual. Soil Conservation Service. U.S. Department of Agriculture Handbook 18.

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.

9 APPENDIX

9.1 Validation rules for ingesting Level 0 data

No data outcome for **date** in mgp_perbulksample_in conditional: \pass\\ if **horizonName**\\\\=\\\\"R\\\\" and **bulkDensID** = null or if \\\\"**biogeoSampleType**\\\\"=\\\\"Audit\\\\" and **bulkDensID** = null; otherwise \\\\"fail\\\\"\\\"\\"\""

No data outcome for **bulkDensID** in mgp_perbulksample_in
conditional: \pass\\ if **horizonName**\\\\=\\\\"R\\\\" and **date** = null or if \\\\"**biogeoSampleType**\\\\"=\\\\"Audit\\\\" and **date** = null; otherwise \\\\"fail\\\\"\\\"\\"\""

No data outcome for **date** in mgp_perbiogeosample_in and mgc_perbiogeosample_in
conditional: \pass\\ if **horizonName**\\\\=\\\\"R\\\\" and **biogeoID** = null or if \\\\"**bulkDensSampleType**\\\\"=\\\\"Audit\\\\" and **biogeoID** = null; otherwise \\\\"fail\\\\"\\\"\\"\""

No data outcome for **biogeoID** in mgp_perbiogeosample_in and mgc_perbiogeosample_in
conditional: \pass\\ if **horizonName**\\\\=\\\\"R\\\\" and **date** = null or if \\\\"**bulkDensSampleType**\\\\"=\\\\"Audit\\\\" and **date** = null; otherwise \\\\"fail\\\\"\\\"\\"\""

No data outcome for **date** in mgp_perarchivesample_in and mgc_perarchivesample_in
conditional: pass if **horizonName**=R and **archiveID** = null or if **biogeoSampleType**=Audit and **archiveID** = null or if **bulkDensSampleType**=Audit and **archiveID** = null; otherwise fail

No data outcome for **archiveID** in mgp_perarchivesample_in and mgc_perarchivesample_in
conditional: pass if **horizonName**=R and **date** = null or if **biogeoSampleType**=Audit and **date** = null or if **bulkDensSampleType**=Audit and **date** = null; otherwise fail

9.2 USDA NRCS method codes

Table 9-1. Method codes for each data product.

Field Name	**Method code***
sandTot	3A1a1b
siltTot	3A1a1b
clayTot	3A1a1b
sandVeryCoarse	3A1a1b
sandCoarse	3A1a1b
sandMedium	3A1a1b
sandFine	3A1a1b
sandVeryFine	3A1a1b
siltCoarse	3A1a1b
siltFine	3A1a1b
co3Clay	3A1a1b
clayFine	3A1a1b

biogeoTotWeight	3A2a1
biogeoTotWeight2-5	3A2a1
biogeoTotWeight5-20	3A2a1
biogeoTotWeight20-75	3A2a1
bulkDensVolume	3B6a
bulkDensDryWeight	3B6a
bulkDensCoarseFragWeight	3B6a
bulkDensCoarseFragDens	3G1a2

*See USDA NRCS Soil Survey Laboratory Methods Manual (Soil Survey Staff 2004) for method descriptions.

9.3 Measurement Uncertainty Equations

9.3.1 Measurement Uncertainty

The following subsections present the uncertainties associated with *individual pressure observations*. It is important to note that the uncertainties presented in the following subsections are *measurement uncertainties*, that is, they reflect the uncertainty of an *individual* measurement. These uncertainties should not be confused with those presented in Section 6.1.2. We urge the reader to refer to AD[11] for further details concerning the discrepancies between quantification of measurement uncertainties and L1 uncertainties.

NEON calculates measurement uncertainties according to recommendations of the Joint Committee for Guides in Metrology (JCGM) 2008. In essence, if a measurand *y* is a function of *n* input quantities $x_i\ (i = 1, \dots, n),\ \ i.e., y = f(x_1, x_2, \dots, x_n)$, the combined measurement uncertainty of *y*, assuming the inputs are independent, can be calculated as follows:

$$u_c(y) = \left(\sum_{i=1}^{N} \left(\frac{\partial f}{\partial x_i} \right)^2 u^2(x_i) \right)^{\frac{1}{2}} \tag{1}$$

where

$\frac{\partial f}{\partial x_i}$ = partial derivative of *y* with respect to x_i

$u(x_i)$ = combined standard uncertainty of x_i.

Thus, the uncertainty of the measurand can be found be summing the input uncertainties in quadrature. For temperature measurements, the sources of uncertainty are depicted in Figure 1. The calculation of these input uncertainties is discussed below.

Repeatability:

$$\bar{X}_i = \frac{1}{n} \sum_{k=1}^{n} X_{i,k} \tag{2}$$

AND:

$$u_{repeat}(measurand\ units) = \left[\frac{1}{(n-1)}\sum_{k=1}^{n}(X_{i,k} - \bar{X}_i)^2\right]^{\frac{1}{2}} \quad (3)$$

OR:

$$u_{repeat}\ (relative) = \frac{u_{repeat}(x_i)}{|\bar{X}_i|} \quad (4)$$

Where the input quantity X_i is estimated from n independent observations $X_{i,k}$ obtained under identical measurement conditions.

Reproducibility:

$$u_{reprod}(measurand\ units) = \left[\frac{1}{(n-1)}\sum_{k=1}^{n}(X_{1,k} - X_{2,k})^2\right]^{\frac{1}{2}} \quad (5)$$

OR:

$$u_{reprod}(relative) = \left[\frac{1}{(n-1)}\sum_{k=1}^{n}\frac{(X_{1,k} - X_{2,k})^2}{\left(\frac{X_{1,k} + X_{2,k}}{2}\right)^2}\right]^{\frac{1}{2}} \quad (6)$$

Where $X_{1,}$ *and* X_2 are replicate measurements with from *n* independent observations for X_k having varying conditions of measurement. Laboratory data will be provided by FIU and stored in the CI data store to calculate both repeatability and reproducibility.

Trueness:

$$u_{true}(measurand\ units) = \left[\frac{1}{(n-1)}\sum_{i=1}^{n}(X_i - X_{i,truth})^2\right]^{\frac{1}{2}} \quad (7)$$

Where X_i is estimated from the n independent observations of the *truth (e.g., reference standard)* $X_{i,truth}$ obtained under identical measurement conditions. A reference standard is a sensor or material used for the calibration of other standards / materials (JCGM 2012).

Truth (uncertainty of):

This metric is taken directly from calibration certificates or manufacturer specifications. It represents the uncertainty of the reference material or standard and may be presented as a combined or expanded uncertainty. If given as a combined uncertainty it can be plugged directly into the final combined uncertainty equation. If this uncertainty is provided at an expanded confidence level, it must be divided by the appropriate expansion factor to convert it to a combined uncertainty prior to propagation:

$$u_{truth}(measurand\ units) = \frac{u_{certified\ value}\ (\%\ CL)}{k_p} \tag{8}$$

OR:

$$u_{truth}(relative) = \frac{u_{truth}}{X_{truth}} \tag{9}$$

Where, k_p is the expansion factor at the designated confidence level and X_{truth} is the value of the traceable reference material.

10 CHANGELOG