NEON USER GUIDE TO SOIL MICROBE BIOMASS (DP1.10104.001)

PREPARED BY	ORGANIZATION
Lee Stanish	TOS
Samantha Weintraub-Leff	TOS
Sam Simkin	TOS

CHANGE RECORD

REVISION	DATE	DESCRIPTION OF CHANGE
A	3/9/2018	Initial Release
B	8/27/2019	Section 3: added information on laboratory methods and calculating lipid concentrations; Section 3.9: added information on mapping lipid terms and calculating total lipid concentration; Section 3.11: new data quality flagging information added; Updated References section.
C	12/23/2020	Included general statement about usage of neonUtilities R package and statement about possible location changes; Section 3: Corrected calculation for correcting lipid concentrations by extraction efficiency; Section 3.3: Added Sampling Design Changes section and included changes to sampling frequency; Section 3.7: Updated descriptions of Associated Data Streams for bundled soil physical and chemical properties and the microbe community composition data products; Section 4.4: Updated list of values for sample quality flagging.
D	03/02/2022	Updated link to table of lipids for mapping to other nomenclatures; minor clarifications throughout; Section 4.3: Updated Data Revision with latest information regarding data release.
E	12/08/2022	Added Theory of Measurements section; additional content in Data Quality section; several other clarifications.
F	2/15/2024	Updates to Sections 3.4, 3.10, 4.4, and 4.5 to include new sme_scaledMicrobialBiomass table and revised sme_batchResults table.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

1 DESCRIPTION

1.1 Purpose

This document provides an overview of the data included in this NEON Level 1 data product, the quality controlled product generated from raw Level 0 data, and associated metadata. In the NEON data products framework, the raw data collected in the field - for example, soil temperature from a single collection event - are considered the lowest level (Level 0). Raw data that have been quality checked via the steps detailed herein, as well as simple metrics that emerge from the raw data are considered Level 1 data products.

The text herein provides a discussion of measurement theory and implementation, data product provenance, quality assurance and control methods used, and approximations and/or assumptions made during L1 data creation.

1.2 Scope

This document describes the steps needed to generate the L1 data product for Soil microbe biomass, and associated metadata, from input data on terrestrial samples. 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 Microbe Biomass (DP1.10104) (AD[04]), 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 laboratory data from samples generated by the field sampling protocols TOS Protocol and Procedure: Soil Biogeochemical and Microbial Sampling (AD[06]) for upland soil samples, and with TOS Standard Operating Procedure: Wetland Soil Sampling (AD[07]) for wetland soil samples. The raw data that are processed as described in this document are detailed in the file, NEON Raw Data Validation for Microbe Biomass (DP0.10104) (AD[03]), provided in the download package for this 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.002652	NEON Data Products Catalog
AD[03]	Available with data download	Validation csv
AD[04]	Available with data download	Variables csv
AD[05]	NEON.DOC.000908	TOS Science Design for Microbial Diversity
AD[06]	NEON.DOC.014048	TOS Protocol and Procedure: Soil Biogeochemical and Microbial Sampling
AD[07]	NEON.DOC.004130	TOS Standard Operating Procedure: Wetland Soil Sampling
AD[08]	NEON.DOC.000913	TOS Science Design for Spatial Sampling
AD[09]	NEON.DOC.000008	NEON Acronym List
AD[10]	NEON.DOC.000243	NEON Glossary of Terms
AD[11]	NEON.DOC.004825	NEON Algorithm Theoretical Basis Document: OS Generic Transitions
AD[12]	Available on NEON data portal	NEON Ingest Conversion Language Function Library
AD[13]	Available on NEON data portal	NEON Ingest Conversion Language
AD[14]	Available with data download	Categorical Codes csv

2.2 Acronyms

Acronym	Definition
PLFA	Phospholipid Fatty Acid
qPCR	Quantitative Polymerase Chain Reaction

3 DATA PRODUCT DESCRIPTION

The Soil Microbe Biomass data product provides quantitative estimates of total microbial biomass in soil samples. NEON measures the abundances of numerous lipid biomarkers that are found in soil microbiota. Data are generated using the high-throughput phospholipid fatty acid (PLFA) analysis, in which the total phospholipid content of a soil sample is extracted, and discrete lipid molecules are quantified using Gas Chromatography and Mass Spectrometry (2018 and earlier) or Gas Chromatography and Flame Ionization Detection (2019 and later) (Buyer and Sasser 2012, Gomez et al. 2014). While there is no perfect method for quantifying microbial biomass in soils, PLFA analysis is widely considered to be a reliable (Zelles 1999) and sensitive (Allison and Martiny 2008) proxy of the viable microbial community (Zhang et al., 2019). There are numerous applications for PLFA data (e.g. Willers et al., 2015); the NEON PLFA data product is designed for quantification of total soil microbial biomass, although other applications may be pursued, such as determination of fungal:bacterial ratios or use of lipid biomarker ratios to understand microbial response to stress (Norris et al. 2023). Data users should understand how to correctly apply these data to their specific research questions as well as the limitations.

The sample plan implements the guidelines and requirements in the Science Designs for TOS Terrestrial Microbial Diversity (AD[05]). Information on sample collection methods such as frequencies per sample type can be found in the Soil Sampling Protocol (AD[06]) and Wetland SOP for wetland soils (AD[07]), and in the NEON User Guide to Soil Physical and Chemical Properties, Periodic (DP1.10086.001).

Microbial biomass samples are a subset of the homogenized soil sample collected as part of the soil microbial diversity and biogeochemistry sampling. After field collection, bulk soil is stored on wet ice and transported to the NEON field laboratory. Within 1 day, the field-moist, bulk soil is passed through a 2 mm sieve (for mineral horizons) or picked of rocks, roots and coarse debris (for organic horizons), and then a representative subsample (5-10 grams) is placed into a tube and stored at -80°C. Samples are shipped to an analytical laboratory where sample processing and analysis occurs.

3.1 Spatial Sampling Design

Sampling for soil microbe biomass analysis is executed at all NEON terrestrial sites, with data reported at the resolution of a single sampling location. This equates to a randomly-assigned X,Y coordinate ($\pm$ 0.5 meters) within a subplot of a NEON plot. Ten plots are sampled, with 3 of 4 subplots randomly selected for sampling per plot per bout (Figure 1). For most bouts, only the surface horizon is sampled to a maximum depth of 30cm, and horizons are broadly defined as either organic (O) or mineral (M).

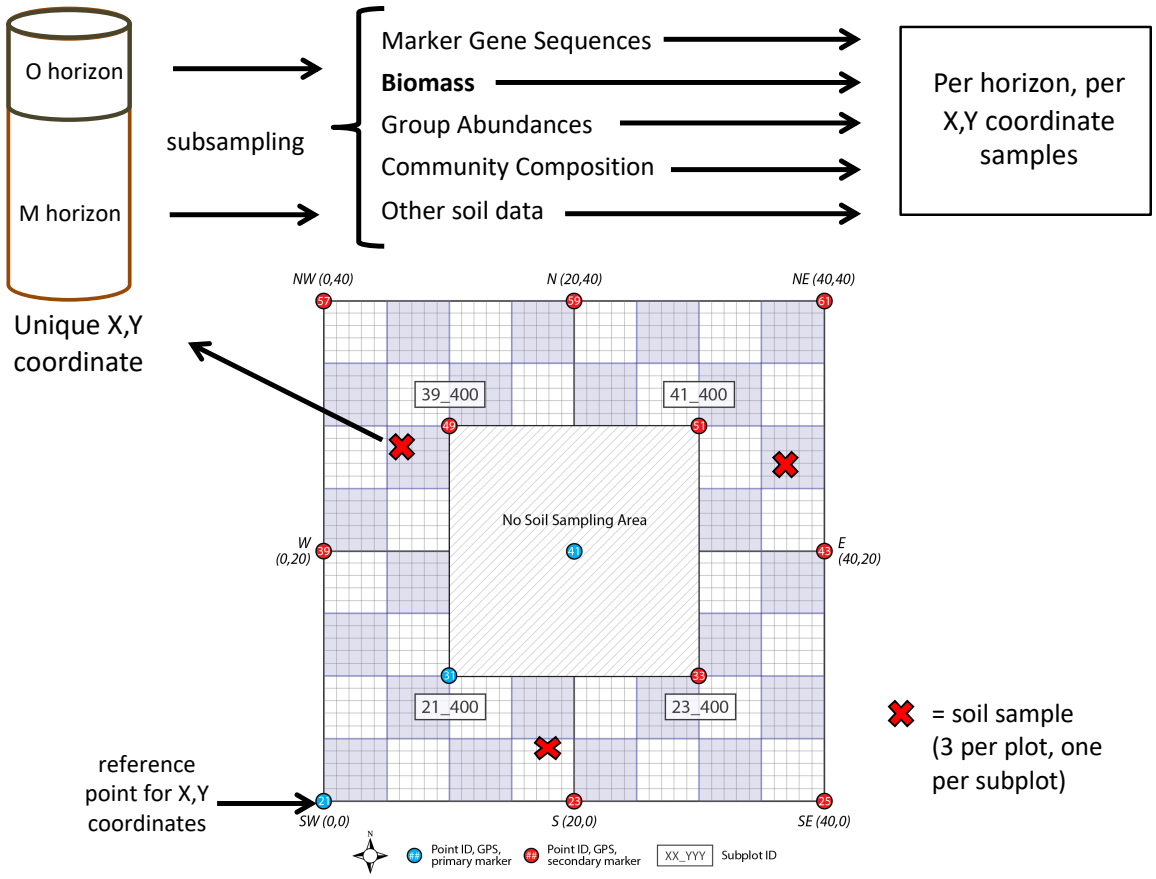

Figure 1: Overview of soil microbial field sampling, spatial design, and analysis workflow.

The spatial design for the microbial biomass data product is described in more detail in the Data Product User Guide for Soil Physical Properties (DP1.10086.001). For a description of the methods used in terrestrial plot selection, refer to the TOS Science Design for Spatial Sampling (AD[08]).

As much as possible, sampling occurs in the same locations over the lifetime of the Observatory. However, over time some sampling locations may become impossible to sample, due to disturbance or other local changes. When this occurs, the location and its location ID are retired. A location may also shift to slightly different coordinates. Refer to the locations endpoint of the NEON API for details about locations that have been moved or retired: https://data.neonscience.org/data-api/endpoints/locations/

3.2 Temporal Sampling Design

Soil sampling for microbial biomass analysis occurs during all bouts at core sites, and 'coordinated' bouts at both core and gradient sites in which additional biogeochemical measurements are made. Coordinated bouts occur at a site once every five years. At most terrestrial sites, sampling occurs 3 times per year in conjunction with the soil physical properties data product (DP1.10086). Two sampling bouts occur during periods of seasonal transitions (e.g. winter-spring or wet-dry), and one during the period of peak green-

ness (as measured by remote sensing data). Only one sampling bout takes place at sites with short growing seasons (e.g. tundra and taiga), during peak greenness.

Up to 2 soil horizons (organic and mineral) are sampled for microbial biomass analysis to a maximum depth of 30 cm.

For all samples, the temporal resolution is that of a single collection date. For a comprehensive description of field methods, refer to TOS Protocol and Procedure: Soil Biogeochemical and Microbial Sampling (AD[06]). Descriptions of the upstream field data for soil sampling (DP1.10086.001) can be found in the Data Product User Guide for Soil Physical Properties.

3.3 Sampling Design Changes

Over the course of early operations, the design for soil sampling has changed. Below is a list of previous sampling strategies that differ from the current design, with applicable years indicated.

- 2013-2020: Subsamples were collected for microbial biomass analysis during 'coordinated' bouts only for all site types. Accordingly, microbial biomass was only measured every 5 years per site.
- 2020-current: Subsamples are collected for microbial biomass analysis during every soil sampling bout at the domain core site, such that each domain analyzes core site soils for microbial biomass annually. Microbial biomass is still analyzed at all sites conducting 'coordinated' bouts as before.

3.4 Theory of Measurements

The NEON Soil Microbe Biomass data provides lipid concentrations that are estimated using phospholipid fatty acid (PLFA) analysis. Data analysis is performed by the analytical laboratory according to the methods described in the laboratory SOP in use at the time (available in the External Lab Protocols > Microbial Analyses section of the NEON Data Portal Document Library, https://data.neonscience.org/documents). Briefly, lipid compounds are identified by comparison to a reference standard, and the areas under each lipid peak are quantified. Concentrations are calculated by comparison to the reference standard containing dozens of lipid compounds with known concentrations.

There are several closely related PLFA analytical methods, including GC-MS (gas chromatography - mass spectrometry) and GC-FID (gas chromatography - flame ionization detection), that have been used for NEON data. Both methods use gas chromatography to separate lipids. For the GC-MS method only, lipids that are not present in the reference solution can also be quantified, with concentrations for lipids that are not part of the reference standard determined based on the response peak of the nearest lipid standard in the chromatogram. For the lipids that are returned, the GC-FID method can arguably return more accurate lipid concentrations than GC-MS. Lipids that have been returned by GC-MS but not GC-FID include: c8To0Concentration, lipid2OH12To0Concentration, lipid3OH12To0Concentration, lipid2OH14To0Concentration, lipid3OH14To0Concentration, trans18To1n9Concentration, trans18To2n912Concentration, c18To3n3Concentration, c18To0Concentration, and c20To3n3Concentration.

The raw results are converted by the analytical laboratory from micrograms/ml to nanomoles/gram soil:

$$\frac{\left(\frac{conc\ (micrograms/ml) * chloroform\ correction * extraction\ volume}{FAME\ formula\ weight (grams/mole)}\right)}{Soil\ freeze{-}dried\ mass (g)}$$

The sum of the individual lipids in the sme_microbialBiomass table is reported in the totalLipidConcentration field.

In order to account for differences across soils in the effectiveness of lipid extraction, the high-throughput phospholipid fatty acid (PLFA) currently used by NEON (Buyer and Sasser, 2012) specifies that a known quantity of an internal standard be added to each sample. Calculated lipid concentrations are then scaled by the recovery or extraction efficiency of that internal standard and those scaled lipid concentrations are reported in sme_scaledMicrobialBiomass. The extraction efficiency is not directly provided in sme_scaledMicrobialBiomass, but for records collected on or after November 1, 2021 users can estimate it by comparing recovery of the internal standard (in raw GC units) for each sample in sme_scaledMicrobialBiomass to the recovery of the internal standard (in raw GC units) in standards and blanks in the sme_batchResults table.

Prior to November 1, 2021: Unscaled lipid concentrations were reported in sme_microbialBiomass, along with lipid extraction efficiency (%) values calculated by the analytical laboratory as:

$$\frac{concentration\ of\ spiked\ internal\ standard * chloroform\ correction * final\ extraction\ volume * 100}{(expected\ mass\ (ng)\ of\ standard),}$$

where the expected mass in the denominator varied by laboratory SOP.

The expectation was that end users would perform scaling corrections themselves as follows:

$$\frac{Unscaled\ lipid\ concentration}{Extraction\ efficiency} * 100$$

Since the community best practice is to use scaled data, NEON stopped ingesting unscaled data for samples collected after November 1, 2021. Samples collected before November 1, 2021 that had an extraction efficiency of 33% or greater in the unscaled sme_microbialBiomass table were converted to scaled values using the extraction efficiency and equation above and then included in the sme_scaledMicrobialBiomass table.

3.5 Variables Reported

All variables reported from the field or laboratory technician (L0 data) are listed in the file, NEON Raw Data Validation for Microbe Biomass (DP0.10104) (AD[03]). All variables reported in the published data (L1 data) are also provided separately in the files within NEON Data Variables for Soil Microbe Biomass (DP1.10104) (AD[04]).

Field names have been standardized with Darwin Core terms (http://rs.tdwg.org/dwc/; accessed 16 February 2014), the Global Biodiversity Information Facility vocabularies (http://rs.gbif.org/vocabulary/gbif/; accessed 16 February 2014), the VegCore data dictionary (https://projects.nceas.ucsb.edu/nceas/projects/bien/wiki/VegCore; accessed 16 February 2014), where applicable.

There are many conventions used for naming lipid compounds, and NEON does not conform to any particular convention with its lipid terms. However, a table exists to help map NEON lipid field names to other existing naming conventions, including Alpha Notation, Omega convention, CAS number, Formula, ChemSpider ID, PubChem ID, and other names. This file is a available here or by navigating to the External Lab Protocols > Microbial Analyses section of the NEON Data Portal Document Library (https://data.neonscience.org/documents). Lipid names typically follow the common lipid nomenclature, with the definition including both the common and scientific names of the compound.

NEON TOS spatial data employs the World Geodetic System 1984 (WGS84) for its fundamental reference datum and GEOID09 for its reference gravitational ellipsoid. Latitudes and longitudes are denoted in decimal notation to six decimal places, with longitudes indicated as negative west of the Greenwich meridian.

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

3.6 Spatial Resolution and Extent

The finest resolution at which spatial data are reported is a single sampling location. This corresponds to a single X,Y coordinate location within a subplot within a plot (Figure 1). The spatial hierarchy is as follows:

sampleID (unique ID given to the individual soil sampling location and horizon) → **subplotID** (ID of subplot within plot) → **plotID** (ID of plot within site) → **siteID** (ID of NEON site) → **domainID** (ID of a NEON domain).

The spatial data are located in the data product Soil Physical Properties, distributed periodic (DP1.10086), in the table ***sls_soilCoreCollection***. The spatial data provided in the download are measured at the plot *centroid*, and have an accuracy of $\pm$ 20 m. However, a more precise measurement may be determined by calculating the offset from the plot centroid using the variables **coreCoordinateX** and **coreCoordinateY**. Refer to the User Guide for Soil Physical Properties, distributed periodic, for more information and instructions.

3.7 Temporal Resolution and Extent

The finest resolution at which temporal data are reported is the **collectDate**, the date and time of day when the sample was collected in the field.

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. Code to combine ("stack") files across months is available here: https://github.com/NEONScience/NEON-utilities

3.8 Associated Data Streams

This section describes the data products that are directly linked or closely related to the soil microbe biomass data product.

Soil data are derived from subsamples collected during soil biogeochemical and microbial sampling and include numerous related data products:

- Soil physical and chemical properties, periodic (DP1.10086.001) - This data product includes field data, soil moisture and pH, laboratory measurements of soil carbon and nitrogen concentrations and stable isotopes, and inorganic nitrogen pools and net transformation rates derived from field incubations. Note that not all measurements are made on every corresponding sample measured for microbial biomass, and vice-versa. Data tables from DP1.10086.001, as well as all downstream laboratory data in the data products referenced below, can be joined to the linking table, ***sls_soilCoreCollection*** by the sample identifier defined for each data product. For DP1.10086.001

referenced here, the linking identifier is the **sampleID**. After joining the other data tables to the ***sls_soilCoreCollection*** table, join to the biomass data using the **biomassID**.

- Soil microbe marker gene sequences (DP1.10108.001) - Microbial 16S and ITS sequence data and metadata. The linking identifier is the **geneticSampleID**, which is present in the table ***mmg_soilDnaExtraction***. First, join the marker gene sequencing tables ***mmg_soilPcrAmplification_16S(or ITS)***, ***mmg_soilMarkerGeneSequencing_16S(or ITS)***, and ***mmg_soilRawDataFiles*** by the **dnaSampleID**, then join to the ***mmg_soilDnaExtraction*** table, which contains the **geneticSampleID**. Next, join to the table ***sls_soilCoreCollection*** by the **geneticSampleID**. Table ***sls_soilCoreCollection*** contains the **biomassID**, which links to the microbial biomass data product.
- Soil microbe community composition (DP1.10081.001) - Microbial community composition data derived from marker gene sequencing. As for the marker gene sequences data, the **geneticSampleID** variable in the tables ***mcc_soilSeqVariantMetadata_16S*** and ***mcc_soilSeqVariantMetadata_ITS*** may be used to link data in this product to the microbial biomass data.
- Soil microbe group abundances (DP1.10109.001): Bacterial/archaeal and fungal abundances as measured by qPCR. The **geneticSampleID** variable in the table ***mga_soilGroupAbundances*** can be used to link data in this product to the microbial biomass data.

3.9 Product Instances

A maximum of 10 plots will be sampled at every site one to three times per year. Up to 2 soil horizons will be collected as separate samples. For each soil horizon sampled, 3 unique subplots are sampled per plot, for up to 6 samples per plot and per sampling event. Thus, there will be 30-180 product instances generated per site per year.

3.10 Data Relationships

The protocol dictates that each X,Y location sampled within a randomly assigned subplot yields a unique **sampleID** per horizon per collectDate (day of year, local time) in the table ***sls_soilCoreCollection*** in DP1.10086.001. Every bout type that includes biomass (e.g. the variable **boutType** includes the string 'biomass') should sample for microbial biomass analysis. A record from ***sls_soilCoreCollection*** may have zero or one child records in the ***sme_scaledMicrobialBiomass*** table (and ***sme_microbialBiomass*** prior to November 1, 2021) of this data product.

Each **biomassID** is a subsample of the parent **sampleID** in the table ***sls_soilCoreCollection***, and is sent for microbial biomass analysis. The PLFA results data appear in the table ***sme_scaledMicrobialBiomass*** (and also in unscaled ***sme_microbialBiomass*** prior to November 1, 2021), and are linked by the **biomassID**. One **biomassID** is expected per record. Duplicate records for an individual **biomassID** should not exist.

For each batch of samples recorded in ***sme_scaledMicrobialBiomass***, at least one record is expected in the batch-level table ***sme_batchResults***. The ***sme_batchResults*** table includes data on standard reference samples and blanks for a batch of samples, linked to ***sme_scaledMicrobialBiomass*** by the **batchID** field, and is available in the expanded data package.

The table ***sme_labSummary*** is expected to have one record annually or when an update to the analytical methods occurs. This table includes the long-term average of the ***sme_batchResults*** precision and accuracy data for lipid analytes used as analytical standards and is available in the expanded data package.

Duplicates and/or missing data may exist where protocol and/or data entry aberrations have occurred; *users should check data carefully for anomalies before joining tables*.

Soil Physical Properties (NEON DP1.10086)

sls_soilCoreCollection.csv - > One record expected per **sampleID**. Generates samples used in Soil microbe biomass (DP1.10104.001), Soil microbe marker gene sequences (DP1.10108.001), Soil microbe community composition (DP1.10081.001), and Soil microbe group abundances (DP1.10109.001).

Soil Microbe Biomass (DP1.10104.001)

sme_scaledMicrobialBiomass.csv - > One record expected per **biomassID**. A biomassID will represent one sample per plot/horizon/X,Y coordinate combination and per collectDate (day of year, local time). There will be only one biomass sample per **biomassID**. For each batch of samples run, a **batchID** is generated.

sme_microbialBiomass.csv - > One record expected per **biomassID** prior to November 1, 2021 (none after that date). A biomassID will represent one sample per plot/horizon/X,Y coordinate combination and per collectDate (day of year, local time). There will be only one biomass sample per **biomassID**. For each batch of samples run, a **batchID** is generated.

sme_batchResults.csv - > At least one record is expected per **batchID**, which corresponds to the **batchID**, in the table ***sme_scaledMicrobialBiomass***.

sme_labSummary.csv - > The laboratory reports long-term precision and accuracy in known analytes every 6 months to 1 year, or when analytes, methods or instrumentation changes. Thus, at least 1 record is expected from a unique laboratory per year. The **labSpecificStartDate** and **labSpecificEndDate** can be used to apply the lab summary data to individual sets of data.

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

3.11 Special Considerations

1. The total lipid concentrations reported in this data product represent the sum total of all lipids measured within a sample. The total number of lipids capable of being measured can vary by analytical laboratory. While there is a baseline set of dominant lipids that all laboratories must measure, there are many additional lipids that a lab may measure depending on the capabilities of that particular laboratory (e.g. using GC-MS vs GC only). These differences may influence total lipid con-

centrations reported in the field **totalLipidConcentration**. If this is a concern, we recommend users calculate total lipids for each sample by summing the lipid concentrations from a subset of the reported lipids that are being quantified consistently across all samples. Lipids that were measured within a sample will contain a numeric value (including 0, when there was no quantifiable amount of target lipid), while lipids that were not measured within a sample will be blank.

4 DATA QUALITY

4.1 Data Entry Constraint and Validation

Many quality control measures are implemented on the laboratory data at the point of data ingest into the NEON database. For example, data formats are constrained and data values are controlled through the provision of controlled list of values (LOV's), which reduces the number of processing steps necessary to prepare the raw data for publication. An additional set of constraints is 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 Microbe Biomass (DP0.10104). This document is provided with every download of this data product. Contained within this file is a field named 'entryValidationRulesParser', which describes syntactically the validation rules for each field built into the data ingest validation. Data entry constraints are described in Nicl syntax in the validation file provided with every data download, and the Nicl language is described in NEON's Ingest Conversion Language (NICL) specifications (AD[12]).

4.2 Automated Data Processing Steps

Following laboratory submission of metadata into the NEON automated data ingest process, 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[11]).

4.3 Data Revision

All data are provisional until a numbered version is released. Annually, NEON releases a static version of all or almost all data products, annotated with digital object identifiers (DOIs). The first data Release was made in 2021. 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 Issue Log section of the data product landing page contains a history of major known errors and revisions.

4.4 Quality Flagging

The **dataQF** field in each data record is a catch-all quality flag for known errors applying to the record. The dataQF codes specific to the four tables belonging to this data product are detailed below.

Table 1: Descriptions of the dataQF codes for quality flagging

fieldName	value	description
dataQF	alaskaDeprecatedMethod	Different methods used for measuring litter depth and the boundaries between soil horizons prior to 2018, use caution when comparing measurements to data collected in 2018 and later

Additionally, several other quality fields have been added over time in order to communicate anomalous sampling conditions or method deviations. These include flags of sample-level data when QA values are out of the accepted range (field **analysisResultsQF**) and flags of batch-level issues with QA standards or blanks in **analyteStandardQF**. Definitions for the categorical codes used for these fields are included in the file NEON Categorical Codes for Soil microbe biomass (AD[14]), provided in the download package for this data product. Fields have been added over time and entries may be missing in older data. Note that the threshold for flagging records with low extraction efficiency in **analysisResultsQF** differed among the two external labs that reported it: extraction efficiencies < 10% were flagged for the "EcoCore_CSU" lab (which used GC-MS) and extraction efficiencies < 45% were flagged for the "Microbial ID, Inc." lab.

Records of land management activities, disturbances, and other incidents of ecological note that may have a potential impact are found in the Site Management and Event Reporting data product (DP1.10111.001)

4.5 Analytical Facility Data Quality

All analytical labs that generate microbial biomass data include standards run as unknowns alongside NEON samples in order to gauge run acceptability. The raw values of standards and blanks analyzed within each sample batch are returned in the ***sme_batchResults*** table, while the ***sme_labSummary*** table contains the long-term analytical precision (**analyteStandardDeviation**) and accuracy (**analyteAccuracy**) of these lab analyses of standards. The long-term analytical precision and accuracy of these standard analyses are reported for each lab to allow users to interpret and analyze lipid concentrations in the context of their uncertainty ranges. Both the ***sme_batchResults*** and ***sme_labSummary*** tables are available in the expanded package.

Within the ***sme_batchResults*** table the lipid (or lipids if a ratio of lipids or total lipids) analyzed in the standard is in the lipidID field, the expected value is in the **analyteKnownValue** field, the observed value and units are in the **analyteObservedValue** and **analyteUnits** fields, respectively, and the full description of the standard material (including manufacturer's catalog number if applicable) and lot number (if applicable) are in the **analysStandardID** and **lotNumber** fields. Starting on November 1, 2021, a change was made to report specific lipid quantities in analytical blanks, total lipids measured from extracting an in-house soil standard, and the recovery of the internal standard (e.g. C19:0 internalStandard) in both types of QC materials in the ***sme_batchResults*** table, with these QA values linked to groups of samples by the **batchID** field. For soils collected prior to November 1, 2021, it is not possible to blank-correct the data, but analytical labs analyzing samples at that time felt their blanks were not high enough to warrant reporting and/or the labs did not report lipids for a known contamination issue (for example c18:0 not reported by MIDI lab).

In addition, labs communicate record-level issues with samples or measurements using the quality flags described in the Quality Flagging section. In general, a null entry in a quality flag field means there is no issue to report.

For further information about individual laboratory QA procedures, refer to the lab SOPs, found on the NEON Data Portal (http://data.neonscience.org/home) in the Resources > Document Library > External Lab Protocols section.

5 REFERENCES

1. Allison, S.D. and J.B.H. Martiny. 2008. Resistance, resilience, and redundancy in microbial communities. Proceedings of the National Academy of Sciences, USA **105**:11512-11519.
2. Buyer, J.S. and M. Sasser. 2012. High throughput phospholipid fatty acid analysis of soils. Applied Soil Ecology **61**:127-130.
3. Gomez, J.D., K. Denef, C.E. Stewart, J. Zheng, and M.F. Cortrufo. 2014. Biochar addition rate influences soil microbial abundance and activity in temperate soils. European Journal of Soil Science **65**:28-39.
4. Norris, C.E., M.J.B. Swallow, D. Liptzin, M. Cope, G. Mac Bean, S.B. Cappellazzi, K.L.H. Greub, E.L. Rieke, P.W. Tracy, C.L.S. Morgan, and C.W. Honeycutt. 2023. Use of phospholipid fatty acid analysis as phenotypic biomarkers for soil health and the influence of management practices. Applied Soil Ecology **185**: 104793.
5. Willers, C., P.J. Jansen van Rensberg, and S. Claassens. 2015. Phospholipid fatty acid profiling of microbial communities: A review of interpretations and recent applications. Journal of Applied Microbiology **119**:1207-1218.
6. Zelles, L. 1999. Fatty acid patterns of phospholipids and lipopolysaccharides in the characterisation of microbial communities in soil: a review. Biological Fertility in Soils **29**:111-129.
7. Zhang, Y., N. Zheng, J. Wang, H. Yao, Q. Qiu, and S.J. Chapman. 2019. High Turnover Rate of Free Phospholipids in Soil Confirms the Classic Hypothesis of PLFA Methodology. Soil Biology and Biochemistry **135**: 323–330.