NEON USER GUIDE TO SOIL CHEMICAL PROPERTIES AND SOIL STABLE ISOTOPES (DP1.10086.001)

PREPARED BY	ORGANIZATION
Samantha Weintraub-Leff	TOS

CHANGE RECORD

REVISION	DATE	DESCRIPTION OF CHANGE
A	03/12/2018	Initial Release
B	08/27/2020	Added text on publication of soil chemistry and stable isotope data alongside field collection metadata
C	04/08/2022	Updated section 4.3 Data Revision with latest information regarding data release
D	03/18/2024	Section 3.4, added details on the use of CO2 trapping and how the C:N ratio value is generated. Section 3.9, added the CO2 trapping variable to the primary key list for the soil chemistry table. Minor text clarifications throughout.

TABLE OF CONTENTS

1 DESCRIPTION **1**
1.1 Purpose . . . 1
1.2 Scope . . . 1

2 RELATED DOCUMENTS AND ACRONYMS **2**
2.1 Associated Documents . . . 2
2.2 Acronyms . . . 2

3 DATA PRODUCT DESCRIPTION **3**
3.1 Spatial Sampling Design . . . 3
3.2 Temporal Sampling Design . . . 3
3.3 Theory of Laboratory Measurements . . . 3
3.4 Variables Reported . . . 4
3.5 Spatial Resolution and Extent . . . 5
3.6 Temporal Resolution and Extent . . . 5
3.7 Associated Data Streams . . . 5
3.8 Product Instances . . . 6
3.9 Data Relationships . . . 6
3.10 Special Considerations . . . 6
3.10.1 Sample drying issue . . . 6
3.10.2 Alaska-specific method changes . . . 7

4 DATA QUALITY **7**
4.1 Data Entry Constraint and Validation . . . 7
4.2 Automated Data Processing Steps . . . 7
4.3 Data Revision . . . 8
4.4 Quality Flagging . . . 8
4.5 Analytical Facility Data Quality . . . 8

5 REFERENCES **9**

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 chemistry and stable isotope components of the L1 data product Soil physical and chemical properties, periodic (DP1.10086.001). This includes total organic carbon (C) and total nitrogen (N) concentrations and stable isotopes in surface soils (0-30 cm). As of August 2020, the Soil chemical properties (DP1.10078.001), Soil stable isotopes (DP1.10101.001), and Soil inorganic nitrogen pools and transformations (DP1.10080.001) data products are published as part of Soil physical and chemical properties, periodic (DP1.10086.001). However, this User Guide is focused on the tables surrounding bulk C and N concentrations and stable isotopes, while two related Data Product User Guides detail field collection metadata and inorganic N pools and net N transformation rate measurements.

This document also provides details relevant to the publication of the data products via the NEON data portal, with additional detail available in the files NEON Data Variables for Soil physical and chemical properties, periodic (DP1.10086.001) (AD[05]) and NEON Categorical Codes for Soil physical and chemical properties, periodic (AD[06]), provided in the download package for this data product.

This document describes the process for ingesting and performing automated quality assurance and control procedures on C and N concentration and stable isotope values measured in NEON soil samples. How the Level 0 data are processed is detailed in the file, NEON Raw Data Validation for Carbon and nitrogen concentrations and stable isotopes in plants and soil (DP0.10103.001) (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]	Available with data download	NEON Raw Data Validation for Carbon and nitrogen concentrations and stable isotopes in plants and soil (DP0.10103.001)
AD[05]	Available with data download	NEON Data Variables for Soil physical and chemical properties, periodic (DP1.10086.001)
AD[06]	Available with data download	NEON Categorical Codes for Soil physical and chemical properties, periodic
AD[07]	DOC.000906	TOS Science Design for Terrestrial Biogeochemistry
AD[08]	NEON.DOC.014048	TOS Protocol and Procedure: SLS - Soil Biogeochemical and Microbial Sampling
AD[09]	NEON.DOC.004130	TOS Standard Operating Procedure: Wetland Soil Sampling
AD[10]	NEON.DOC.000008	NEON Acronym List
AD[11]	NEON.DOC.000243	NEON Glossary of Terms
AD[12]	NEON.DOC.004825	NEON Algorithm Theoretical Basis Document: OS Generic Transitions
AD[13]	Available on NEON data portal	NEON Ingest Conversion Language Function Library
AD[14]	Available on NEON data portal	NEON Ingest Conversion Language

2.2 Acronyms

Acronym	Definition
δ13C	delta 13C, the stable carbon isotope ratio (13C:12C) in a sample compared to a reference material, reported in parts per thousand
δ15N	delta 15N, the stable nitrogen isotope ratio (15N:14N) in a sample compared to a reference material, reported in parts per thousand
C	Carbon
N	Nitrogen

3 DATA PRODUCT DESCRIPTION

Soil chemistry and stable isotope data provide information about soils collected using TOS Protocol and Procedure: SLS - Soil Biogeochemical and Microbial Sampling (AD[08]), or TOS Standard Operating Procedure: Wetland Soil Sampling (AD[09]) if the site is a wetland. These procedures implement the guidelines and requirements described in the TOS Science Design for Terrestrial Biogeochemistry (AD[07]). All laboratory data are reported at the spatial resolution of a single soil sample, which is collected from a unique X,Y coordinate (+/- 0.5 meters) and horizon (mineral or organic) within a NEON plot. The temporal resolution is that of a single collection date.

Measurements of soil organic carbon and total soil nitrogen help to reveal drivers of variation in belowground element storage and stoichiometry at the plot, site, and continental scales. Measurements of stable isotopes can shed light on nutrient sources and transformations and serve as biogeochemical integrators of plant-soil-microbial dynamics. Repeat measurements of concentrations and stable isotope ratios will provide essential data for understanding change in soil biogeochemical dynamics over time.

3.1 Spatial Sampling Design

Briefly, soils are sampled at all terrestrial NEON sites from three pre-determined, randomly assigned X,Y locations in 3 of 4 randomly selected subplots per 40 x 40 meter plot. Soils are sampled to a maximum depth of 30 cm and separated by horizon type (mineral vs organic). Ten plots per site are sampled, four within the Tower airshed and six others distributed across the landscape and located in dominant vegetation types. See AD[02] for further details on the NEON spatial design and the associated Data Product User Guide for Soil physical and chemical properties, periodic for more information on the soil sampling spatial design.

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

Soils are analyzed for chemistry and stable isotope composition every 5 years at each site, using samples collected during the period of peak greenness. Soil sampling for other parameters occurs more frequently, but does not result in chemical or stable isotopic measurements. When soils are analyzed for total organic C and total N, an air-dried subsample is also created for archive purposes. Archive samples are housed in the NEON Biorepository and are available for community use.

3.3 Theory of Laboratory Measurements

Concentrations and stable isotope ratios of soil C and N are measured simultaneously using elemental analysis coupled to isotope ratio mass spectrometry (EA-IRMS). As DP1.10086.001 reports on total soil *organic* carbon, samples from dryland sites that contain soil carbonates must be analyzed twice - once using an acidified subsample where inorganic C has been removed, and again with a non-acidified subsample

to get reliable N values. These two analytical runs per dryland sample are reported separately in the data. In other sites, it is sometimes necessary to trap CO_2 in order to get good N data, most often required for high carbon to nitrogen samples. This also requires two analytical runs and yields C and N data reported on separate rows. In cases of acid treatment and CO_2 trapping, the C:N ratio field is left empty. Otherwise, the ratio of C:N is reported using % data that is not rounded, whereas the percent data are reported after rounding to 2 decimal places. This can lead to subtle differences in the C:N ratio variable in the data compared to a user generated value.

Isotopes are measured as the abundance ratio of a heavy, rare isotope (H) to a light, more common isotope (L), normalized by those same ratios in a standard reference material.

$$\delta = [(R_{sample}/R_{standard} - 1)] \times 1000$$

where R = H/L. For all NEON stable isotopic data, δ15N values are normalized to atmospheric N2 and δ13C values are normalized to Vienna Pee Dee Belemite.

NEON partners with external laboratories to conduct these analyses. Standard operating procedures (SOPs) from external labs can be found in the NEON Data Portal document library (http://data.neonscience.org/documents), in the External Lab Protocols > Terrestrial Biogeochemistry section.

Most labs analyze a percentage of EA-IRMS samples in duplicate in order to monitor internal consistency and repeatability. In these instances, NEON passes along replicate analyses in case the uncertainty information is of interest. Before using the data, end users may wish to remove or average the information from analytical replicates, by taking 'mean' or 'first' of all measurements of a given analyte associated with the same sampleID.

3.4 Variables Reported

All variables reported from the laboratory (L0 data) are listed in the file, NEON Raw Data Validation for Carbon and nitrogen concentrations and stable isotopes in plants and soil (DP0.10103.001) (AD[04]). All variables reported in the published data (L1 data) are also provided separately in the file NEON Data Variables for Soil physical and chemical properties, periodic (DP1.10086.001) (AD[05]).

Field names have been standardized with Darwin Core terms (http://rs.tdwg.org/dwc/; accessed 16 February 2014), the Global Biodiversity Information Facility vocabularies (http://rs.gbif.org/vocabulary/gbif/; accessed 16 February 2014), and the VegCore data dictionary (https://projects.nceas.ucsb.edu/nceas/projects/bien/wiki/VegCore; accessed 16 February 2014), where applicable. NEON TOS spatial data employs the World Geodetic System 1984 (WGS84) for its fundamental reference datum and 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.

3.5 Spatial Resolution and Extent

The finest resolution at which spatial data are reported is a single X,Y sampling location within a NEON plot.

cnSampleID (unique ID given to the individual soil sample analyzed for carbon and nitogen) → **subplotID** (ID of subplot within plot) → **plotID** (ID of plot within site) → **siteID** (ID of NEON site) → **domainID** (ID of a NEON domain.

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

To derive a more precise estimate of the location of each X,Y soil sampling location, users can join the chemistry and stable isotope data to field collection metadata for associated samples and then follow the instructions contained in the associated Data Product User Guide for Soil physical and chemical properties, periodic.

3.6 Temporal Resolution and Extent

The finest resolution at which temporal data are reported is the **collectDate**. All samples associated with a sampling event have collect dates within a ~14-day window. The total number of sampling events will be one, once every 5 years during the period of historic peak greenness. The exception is for 'legacy data' that was collected prior to 2017, which may have been sampled more frequently. 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. For code resources to work with these files, see Data Relationships (3.9).

3.7 Associated Data Streams

sampleID and **cnSampleID** are the linking variables that can be used to join soil C and N data to field collection metadata and measurements of soil moisture and pH. Soils measured for total organic C and total N are also extracted with 2M potassium chloride (KCl) to determine inorganic N concentrations and net N transformation rates. These measurements are discussed in detail in the Data Product User Guide for Soil inorganic nitrogen pools and transformations, and **sampleID** is the linking variable to join across tables. Moreover, soils are also analyzed to produce several microbial data products. These are discussed in detail in the associated Data Product User Guide for Soil physical and chemical properties, periodic.

Users may wish to link soil chemistry and stable isotope measurements to other biogeochemical pools and fluxes measured at the site during the same year (litter chemistry, foliar chemistry, etc). In some cases, measurements will be coincident at the plot level, and different measurement streams can be joined using the **plotID** variable. In other cases, data will not overlap in space and users may need to take site-level means and join using **siteID**, or use other spatially explicit approaches to align the data.

3.8 Product Instances

A maximum of 10 plots will be sampled per site. For each soil horizon present (maximum of 2, organic and mineral), 3 samples will be collected per plot. Thus on average, this will result in 30-60 unique soil samples per year per site sampled for carbon and nitrogen. Approximately 8-12 terrestrial NEON sites per year will be sampled for carbon and nitrogen. Thus, we expect 240-720 unique data records per year.

3.9 Data Relationships

TOS Protocol and Procedure: SLS - Soil Biogeochemical and Microbial Sampling dictates that each X,Y location sampled for soil yields a unique **sampleID** per horizon per collectDate. This sample is then subsampled for chemistry and stable isotope analyses, yielding a corresponding **cnSampleID** in the sls_bgcSubsampling table. Each of these child records may then appear from zero to four times in the sls_carbonNitrogen table. Most **cnSampleID**s will appear once, but some may appear more than once if analytical replicates were conducted, if samples were acidified to remove carbonates, or if CO_2 trapping was required (see Section 3.4). The associated field metadata as well as measurements of moisture, pH, and inorganic N appear in other tables that are described in related Data Product User Guides. Duplicates and/or missing data may exist where protocol and/or data entry abberations have occurred; *users should check data carefully for anomalies before joining tables*.

sls_soilChemistry.csv - > One record expected per **cnSampleID** x **analyticalRepNumber** x **acidTreatment** x **co2Trapped** combination.

bgc_CNiso_externalSummary.csv - > One record expected per **analyte** x **sampleType** x **laboratoryName** x **qaReportingStartDate** combination, used to associate sample data with relevant C and N uncertainty values.

Identifiers and barcodes are generated for each sample. Only soil material designated for long-term archive (**bgcArchiveID** and **bgcArchiveCode**) will be retained; otherwise excess physical soil material is discarded following a short-term holding period, generally one year from collection.

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

3.10 Special Considerations

3.10.1 Sample drying issue

Due to a miscommunication, prior to 2020-07-09 soil samples analyzed for carbon (C) and nitrogen (N) concentrations and stable isotopes were not re-dried prior to weighing and analysis at the external lab. All NEON soil samples are dried at 65C in the domain labs, but they are often then stored for weeks to

months before being shipped to or analyzed by external labs. During this time they may accumulate moisture, especially in humid areas.

Subsequent testing revealed that for organic horizons, %C data measured prior to 2020-07-09 are likely underestimated by 1.5-2.5% due to this lack of re-drying prior to analysis. As organic horizon soil samples tend to have high %C (> 20%), this bias may have only minor impacts on many analyses, but is something for end users to keep in mind. For the other parameters (%N, C:N, d15N, d13C) as well as all parameters in mineral horizons, testing suggests there were no detectable differences between re-dried samples and originals. All affected records have been flagged, see section 4.4 for more details. All samples collected after 2020-07-09 are re-dried (both horizon types) prior to external analysis.

3.10.2 Alaska-specific method changes

Regarding Alaska collections (TOOL, BARR, DEJU, BONA, and HEAL), beginning in 2018 the process for differentiating between organic soil and litter in tundra and boreal sites was revised. Prior to 2018, instructions to technicians may have resulted in a portion of the top-most fibric material being included in measurements of litter depth, and excluded from O-horizon soil samples. This likely means that top and bottom depths for Alaska samples were recorded differently before and after the protocol change. Comparison of chemistry values from samples collected pre and post 2018 (inclusive) should thus be undertaken with caution.

4 DATA QUALITY

4.1 Data Entry Constraint and Validation

Many quality control measures are implemented at the point of data entry within a mobile data entry application or web user interface (UI). For example, data formats are constrained and data values controlled through the provision of list-of-value options, which reduce the number of processing steps necessary to prepare the raw data for publication. An additional set of constraints are implemented during the process of ingest into the NEON database. The product-specific data constraint and validation requirements built into data entry and database ingest are described in the document NEON Raw Data Validation for Carbon and nitrogen concentrations and stable isotopes in plants and soil (DP0.10103.001), provided with every download of this data product. Contained within this file is a field named 'entryValidationRulesParser', which describes the validation rules for external labs that submit spreadsheets to the NEON database. Data entry constraints are described in Nicl syntax in the validation file provided with every data download, and the Nicl language is described in NEON's Ingest Conversion Language (NICL) specifications (AD[14]).

Note that field data collected prior to 2017 were processed using a paper-based workflow that did not implement the full suite of quality control features associated with the interactive digital workflow. Moreover, external laboratory data were also not subject to the same full suite of quality controls.

4.2 Automated Data Processing Steps

Following data entry into a mobile application or web user interface, the steps used to process the data through to publication on the NEON Data Portal are detailed in the NEON Algorithm Theoretical Basis

Document: OS Generic Transitions (AD[12]).

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 record is a quality flag for known issues applying to the record, added by NEON Science upon data review. See below for an explanation of **dataQF** codes specific to soil chemistry and stable isotope data. The issues with sample re-drying and Alaska-specific methods are detailed further in the *Special Considerations* section above.

fieldName	value	definition
dataQF	legacyData	Data collected using a paper-based workflow that did not implement the full suite of quality control features associated with the interactive digital workflow
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
dataQF	dryingProtocolError	Samples were not re-dried prior to external lab analysis, weight percent C values for organic horizons are likely underestimated by 1.5-2.5 percent

4.5 Analytical Facility Data Quality

Primary reference materials are used to calibrate each EA-IRMS run and secondary reference materials are analyzed in order to gauge run acceptability. Labs communicate run-level issues with the accuracy of secondary reference materials or standards, as well as record-level issues with samples or measurements, using a suite of quality flags. Definitions for the categorical codes used for these QF fields are included in the file NEON Categorical Codes for Soil physical and chemical properties, periodic (AD[06]). In general, an entry of OK in a quality flag field means there is no issue to report. Fields have been added over time and entries may be missing in older data.

In addition, long-term analytical precision and accuracy of secondary reference material analyses are reported per lab to allow users to interpret and analyze chemistry and stable isotope data in the context of their uncertainty ranges. The data table bgc_CNiso_externalSummary, which is available in the data product expanded package, contains the long-term precision and accuracy of lab analyses.

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 > Terrestrial Biogeochemistry section. NEON's Calibration/Validation department has regular procedures for auditing the quality assurance of external laboratories and their reports are also available to data users.

5 REFERENCES