



| | |
|---|-------------------------|
| <i>Title:</i> NEON User Guide to Soil inorganic nitrogen pools and transformations (NEON.DP1.10086) | <i>Date:</i> 03/08/2024 |
| <i>Author:</i> Samantha Weintraub-Leff | <i>Revision:</i> E |

NEON USER GUIDE TO SOIL INORGANIC NITROGEN POOLS AND TRANSFORMATIONS (NEON.DP1.10086)

| PREPARED BY | ORGANIZATION |
|-------------------------|---------------------|
| Samantha Weintraub-Leff | TOS |

CHANGE RECORD

| REVISION | DATE | DESCRIPTION OF CHANGE |
|----------|------------|---|
| A | 06/14/2018 | Initial Release |
| B | 09/01/2020 | Added text on publication of soil inorganic N data alongside field collection metadata. Included more details regarding net rate calculations and expanded the discussion of quality issues and special considerations |
| C | 03/02/2022 | Section 3.10: Updated info regarding the incubationPairID field and nitrite contamination in KCl solutions; minor clarifications and edits throughout; Section 4.3: Updated Data Revision with latest information regarding data release |
| D | 12/08/2022 | Section 3.10: Updated the info on negative sample concentrations due to nitrite contamination of KCl, added new findings of contamination from other consumables and steps to mitigate; minor text updates and clarifications throughout the document |
| E | 02/16/2024 | Section 3.10: Noted missing BARR 2022 net rate data, minor updates in 'Other sources of contamination' section. Section 4.4: Added new dataQF entry for resampled X,Y locations, expanded discussion of how to handle apparent negatives including stastical correction developed in Weintraub-Leff et al (2023) and integrated into the custom function. Appendix: Added 'noxCorrection' to variable list and equation to apply that correction presented in 4a. 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 | 4 |
| 3.2 | Temporal Sampling Design | 4 |
| 3.3 | Theory of Analytical Measurements | 5 |
| 3.4 | Variables Reported | 5 |
| 3.5 | Spatial Resolution and Extent | 6 |
| 3.6 | Temporal Resolution and Extent | 6 |
| 3.7 | Associated Data Streams | 6 |
| 3.8 | Product Instances | 6 |
| 3.9 | Data Relationships | 7 |
| 3.10 | Special Considerations | 7 |
| 4 | DATA QUALITY | 9 |
| 4.1 | Data Entry Constraint and Validation | 9 |
| 4.2 | Automated Data Processing Steps | 10 |
| 4.3 | Data Revision | 10 |
| 4.4 | Quality Flagging | 10 |
| 4.5 | Analytical Facility Data Quality | 11 |
| 5 | REFERENCES | 12 |
| 6 | APPENDIX | 12 |

LIST OF TABLES AND FIGURES

| | | |
|----------|--|---|
| Figure 1 | Simple overview diagram of the nitrogen cycle. Reproduced from University of California Cooperative Extension. | 3 |
| Figure 2 | Layout for paired samples used to measure soil net N transformation rates. | 4 |
| Figure 3 | Schematic of the application used by field technicians to enter KCl extraction data | 9 |

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 inorganic nitrogen pools and net N transformation rate components of the L1 data product Soil physical and chemical properties, periodic (DP1.10086.001). This includes the laboratory extraction procedures and chemical analyses involved in measuring soil extractable inorganic N concentrations and net N mineralization and nitrification rates. As of August 2020, the Soil inorganic nitrogen pools and transformations (DP1.10080.001), Soil chemical properties (DP1.10078.001), and Soil stable isotopes (DP1.10101.001) data products are being published as part of Soil physical and chemical properties, periodic (DP1.10086.001). However, this User Guide is focused on the tables surrounding soil inorganic N measurements while two related Data Product User Guides detail field collection metadata and total soil C and N 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 data collected and processed according to TOS Protocol and Procedure: SLS – Soil Biogeochemical and Microbial Sampling (AD[08]), or TOS Standard Operating Procedure: SLS – Wetland Soil Sampling (AD[09]) if the site is a wetland. The raw data that are processed in this document are detailed in the file NEON Raw Data Validation for Soil inorganic nitrogen pools and transformations, Level 0 (NEON.DP0.10080) (AD[04]), provided in the download package for this data product. Please note that raw data products (denoted by ‘DPO’) 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 Soil inorganic nitrogen pools and transformations, Level 0 (NEON.DP0.10080) |
| 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] | NEON.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: SLS – 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 |
|------------------------------|--------------------|
| KCl | potassium chloride |
| N | Nitrogen |
| NH ₄ ⁺ | ammonium |
| NO ₃ ⁻ | nitrate |
| NO ₂ ⁻ | nitrite |

3 DATA PRODUCT DESCRIPTION

Soil inorganic nitrogen data provide information on soil cores collected using TOS Protocol and Procedure: SLS – Soil Biogeochemical and Microbial Sampling (AD[08]), or TOS Standard Operating Procedure: SLS – Wetland Soil Sampling (AD[09]) if the site is a wetland. Inorganic N measurements implement the guidelines and requirements described in the TOS Science Design for Terrestrial Biogeochemistry (AD[07]). All inorganic N 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 subplot within a NEON plot. The temporal resolution for each sample is a single collection date.

The goal is to measure the two dominant forms of inorganic N in soil, ammonium (NH_4^+) and nitrate (NO_3^-), as well as quantify *in-situ* net N transformation rates, namely mineralization and nitrification (Figure 1). Net rates are estimated using field-based incubations of either covered cores (the majority of plots) or buried bags (wetlands). In either case, ‘initial’ soil cores are collected and extracted with 2M potassium chloride (KCl) along with a set of procedural blanks. At the same time, a covered core (2” diameter) or buried bag (4 mm thick) is deployed at each X,Y location within 0.2-0.5 m from the initial sample. After a two-to-four week incubation period, depending on site climate and time of year, ‘final’ samples are recovered, then extracted with 2M KCl along with their own procedural blanks. Blank-corrected differences in final versus initial inorganic N concentrations can be used to calculate net N mineralization, while blank-corrected difference in final versus initial NO_3^- is net nitrification (Figure 1). Note that laboratory extraction metadata will be available for download prior to the inorganic N analytical data.

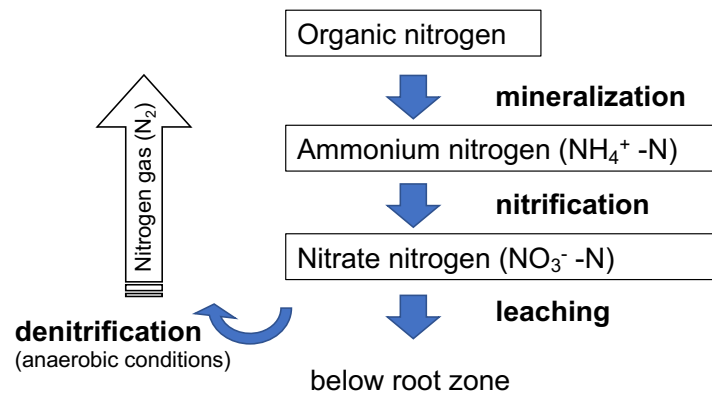


Figure 1: Simple overview diagram of the nitrogen cycle. Reproduced from University of California Cooperative Extension.

N is a key limiting nutrient in terrestrial ecosystems, and measurements of soil inorganic N pools and net N transformation rates help to reveal drivers of variation in nutrient availability and limitation, soil microbial activity, carbon-nitrogen interactions, and potential for nitrogen losses at the plot, site, and continental scales. They also provide essential data for understanding changes in soil and microbial N 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, one in each of three randomly assigned 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.

Two sub-locations at each X,Y are sampled - one for the initial core, which is subsampled for other chemistry and microbial measurements, and another in close proximity for the incubated sample (Figure 2).

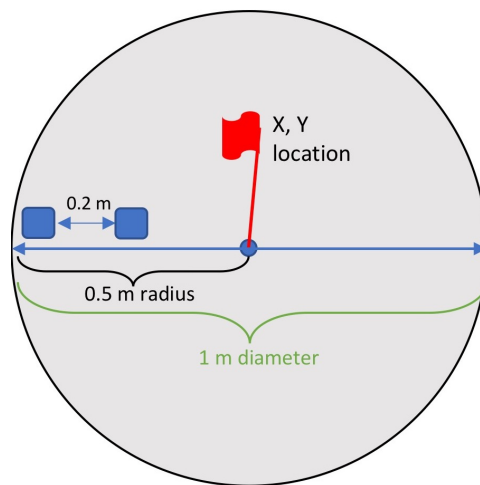


Figure 2: Layout for paired samples used to measure soil net N transformation rates.

As much as possible, sampling occurs in the same locations (for the soil protocol, plots) 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/endpoint/locations/>

3.2 Temporal Sampling Design

Analyses of inorganic N pools and transformation rates occur on an every-5-years interval at each terrestrial NEON site. During 'on' years, sampling occurs at least once at all sampling sites (8-12 Observatory-wide) during the local period of historic peak greenness. At nearly all sites, N transformation measurements occur two additional times, usually bracketing the peak greenness window and aimed to capture seasonal transitions in plant and microbial activity. Sampling frequency is reduced in arctic and boreal sites, which are only sampled during peak greenness.

Within 1 day of collection, initial and final soil samples are extracted with 2M KCl (along with a set of procedural blanks), filtered, and frozen. At the same time, soil subsamples are also measured for soil moisture. Once the initial/final set have both been processed, sample and blank extracts are shipped frozen to an external laboratory for analysis. Note that laboratory extraction metadata and soil moisture values will be available for download prior to inorganic N analytical data.

3.3 Theory of Analytical Measurements

Ammonium is measured using the salicylate-nitroferrocyanide method, and nitrate + nitrite is measured using cadmium reduction and sulfanilamide, both on an autoanalyzer. For the latter, NEON samples are not run with and without the reduction step, thus all measurements are actually nitrate + nitrite. That said, nitrite is often a minor constituent of natural soils.

NEON partners with external laboratories to conduct these analyses. The name of the lab that analyzed each sample is recorded in the data, and standard operating procedures (SOPs) from these external facilities can be found in the NEON Data Portal document library (<http://data.neonscience.org/documents>), in the External Lab Protocols > Terrestrial Biogeochemistry section. Many labs analyze a percentage of 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.

Both ammonium and nitrate + nitrite concentrations are reported in units of milligrams nitrogen per milliliter KCl. In order to blank-correct the data, then calculate micrograms N per gram dry soil (per day for net rates), users will need to combine data from several of the tables that come in the data product download package, then conduct a series of simple calculations. These calculations are described in the Appendix of this User Guide. Additionally, an R package to perform the calculations is available here: <https://github.com/NEONScience/NEON-Nitrogen-Transformations>.

3.4 Variables Reported

All variables reported from the NEON or external laboratory technician (L0 data) are listed in the file NEON Raw Data Validation for Soil inorganic nitrogen pools and transformations, Level 0 (NEON.DP0.10080) (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), the VegCore data dictionary (<https://projects.nceas.ucsb.edu/ncceas/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 unique X,Y sampling location within a NEON plot.

sampleID (unique ID given to the individual soil sampling location and horizon, X,Y coordinate locations are relative to the southwest corner of the plot) → **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 inorganic N 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**. The total number of sampling events per year per site will be 1-3. Sampling will be conducted once every 5 years per site. 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 is the linking variable that can be used to join inorganic N data to field collection metadata and measurements of soil moisture. Soils measured for inorganic N during peak greenness are also analyzed for total organic C and total N. These measurements are discussed in detail in the Data Product User Guide for Soil chemical properties and Soil stable isotopes, 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 inorganic N pool and transformation rate data 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 per initial or final collection event, with 2 events per

bout. Since sites will have 1-3 sampling bouts per year, this will result in 60-360 unique soil samples extracted for inorganic N per site. On average, 9-10 terrestrial NEON sites per year will be sampled for inorganic N pools and net transformation rates. Thus, we expect 540-3,240 data records (excluding procedural blanks) per year.

3.9 Data Relationships

TOS Protocol and Procedure: SLS – Soil Biogeochemical and Microbial Sampling dictates that each soil sample yields a unique **sampleID** per horizon per collectDate (day of year, local time). Each sample is then subsampled and extracted for inorganic N yielding a corresponding **kclSampleID**. Thus, a record from `sls_soilCoreCollection` may have zero or one child records in the `ntr_internalLab` table. The information needed to link procedural blanks with sample KCl extracts is provided in `ntr_externalLabBlanks`. Each child record from `ntr_internalLab` may appear from zero to two times in the `ntr_externalLab` table. Most will appear once, but some may appear twice if analytical replicates were conducted. The associated field collection metadata as well as measurements of soil moisture, pH, and bulk carbon and nitrogen appear in other tables that are described in related Data Product User Guides. Duplicates and/or missing data may exist where protocol or data entry aberrations have occurred; *users should check data carefully for anomalies before joining tables*.

`ntr_internalLab.csv` -> One record expected per **sampleID**, generates a single **kclSampleID** used to measure inorganic N concentrations

`ntr_internalLabBlanks.csv` -> One record expected per **kclReferenceID**, used to track which procedural blanks are associated with a batch of samples

`ntr_externalLab.csv` -> One record expected per **kclSampleID** x **analyticalRepNumber** combination, associated with ammonium and nitrate + nitrite concentration measurements in KCl extracts

`ntr_externalSummary.csv` -> One record expected per **analyte** x **sampleType** x **laboratoryName** x **qaReportingStartDate** combination, used to associate sample data with relevant uncertainty values. Available in the expanded data package.

Identifiers and barcodes are generated for each sample. Physical sample material is discarded once all analyses are complete.

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

1. *Linking variables*: Paired initial and final (incubated) cores can be linked through the **incubationPairID** field. Prior to 2022, this field was simply `sampleID` minus the collect date info. From 2022

onward, this field combines sampleID and eventID information. Samples should be associated with their corresponding procedural blanks using the **kclReferenceID** field, which then is linked to individual blank IDs in the ntr_internalLabBlanks table. The dates in the blank identifiers reflect the date of lab processing, not field collection, and hence often differ from the dates in associated samples because extractions occur the following day.

- Missing data:* Records from the sls_soilCoreCollection field metadata table may not appear in the downstream tables if samples were lost or discarded before they could be processed in the lab. For soil moisture, there is one site, TALL, where moisture data were not collected for any of the incubated samples from 2017. As such, net N transformation rates in micrograms N per gram dry soil per day cannot readily be calculated for these records. Additionally, the BARR site in 2022 experienced extremely poor weather during the summer season, as such it was not possible to incubate samples. This means soil extractable pool data are available but net N transformation rates are not.
- Alaska-specific consideration:* Beginning in 2018 the process for differentiating between organic soil and litter in arctic tundra and boreal sites was revised to more accurately reflect the slow rates of decomposition and extensive fibric organic soil layers. The earlier method may have thus resulted in a portion of the top-most fibric material being excluded from O-horizon soil samples, and could have impacted the top and bottom depths reported. TOOL was the only arctic site that measured inorganic N prior to 2018, but users should use caution when comparing these data from TOOL collected before and after 2018.
- Negative blank-corrected sample concentrations:* In some cases, blank-corrected concentrations appear negative, meaning more inorganic N was detected in procedural blanks compared to samples. Slightly negative values near method detection limits (- 0.01-0.02 mg/L) can be considered equivalent to zero, but significant negative values suggest an issue. Extensive troubleshooting of sample sets across the Observatory revealed that some batches of KCl used by NEON contained high levels of nitrite, which is chemodenitrified (abiotically converted to nitrogen gas and lost) in the presence of soil, but not in blanks (Homyak et al. 2015). This artifact is what leads to the significant negative values and greatly reduces sensitivity of the method in low-concentration sites. Starting with the 2022 sampling season and onward, all batches of reagent-grade potassium chloride purchased for soil extractions are tested for nitrate and nitrite concentrations. Nitrite concentrations of < 0.02 mg N/L are preferred, although powder with up to 0.05 mg N/L may be used if higher purity KCl cannot be found. Only batches that meet these purity levels are used for extractions. Thus far, this approach has been shown to greatly reduce, although not entirely eliminate, the issue with negative sample values. All samples with concentrations significantly lower than blanks are flagged, as noted in Quality Flagging (4.4), and users have several options for how to treat these data as discussed below.
- Other sources of contamination:* Over time, NEON has discovered other sources of contaminant N in the extraction procedure. For instance, the water source used to create KCl solutions, as discussed further in section 4.4. In 2022, all of the consumable items used in the lab extractions at two domains were tested for nitrite and nitrate. Scintillation vials were below detection limits, but GF/A filters and certain specimen cups used for extractions had significant nitrate levels. KCl leaching of filters prior to use was already part of the workflow; a double rinse of all specimen cups with ultrapure deionized water was adopted as of 2022-07-25 after testing showed this was sufficient to

remove contaminant N.

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 dropdown options, which reduces the number of processing steps necessary to prepare the raw data for publication. A schematic of the data entry application design for internal lab processing of KCI extracts is depicted in Figure 3. An additional set of constraints are implemented during the process of ingest into the NEON database.

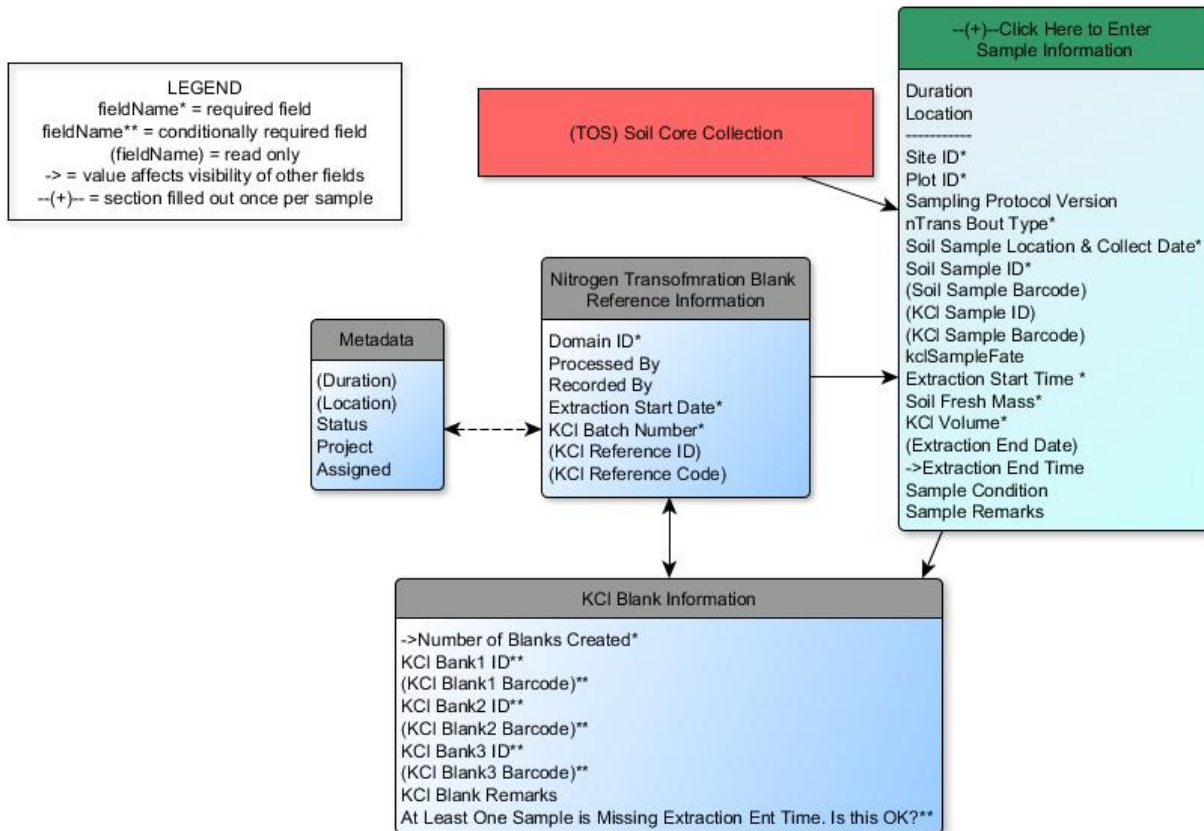


Figure 3: Schematic of the application used by field technicians to enter KCI extraction data

The product-specific data constraint and validation requirements built into data entry applications and database ingest are described in the document NEON Raw Data Validation for Soil inorganic nitrogen pools and transformations, Level 0 (NEON.DP0.10080), provided with every download of this data product. Contained within this file is a field named 'entryValidationRulesForm', which describes syntactically

the validation rules for each field built into the data entry application. There is also a field named ‘entry-ValidationRulesParser’, which described the validation rules for external labs that submit spreadsheets to the NEON database. Data entry constraints are described in Nici syntax in the validation file provided with every data download, and the Nici language is described in NEON’s Ingest Conversion Language (NICL) specifications ([AD[14]).

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 data record is a catch-all quality flag for known errors applying to the record, added by NEON Science upon data review. See below for an explanation of **dataQF** codes specific to inorganic N data. If more than one value applies, all are listed and separated by pipes (|).

1. In 2017, Type II (1 Mohm) water was used to create 2M KCl solutions and conduct extractions. However, this purity is not sufficient and Type I water (18.2 Mohm) has been used for all bouts conducted from 2018 onward. 2017 data was thus collected using a deprecated method and has been flagged accordingly.
2. Related to this issue, some sites that used Type II water in 2017 had extremely high N concentrations in blanks (> 1 mg/L), leading to unusable data. These values have been set to NA and the sample sets are flagged as described below.
3. There was a method change for Alaska soil sampling as noted above in Special Considerations (3.10), data have been flagged accordingly.
4. TOS Protocol and Procedure: SLS – Soil Biogeochemical and Microbial Sampling mandates that each unique X,Y location within a NEON soil plot is sampled only once during the lifetime of the project. However, occasional errors lead to X,Y locations being resampled months to years apart. When this occurs, a flag is added to the data records for instances other than the first (intended) sampling.

| fieldName | value | definition |
|-----------|------------------------|---|
| dataQF | deprecatedMethod | Data generated using deprecated procedures; Type II water |
| dataQF | concDataRemoved | Extremely high blank values from Type II water leading to non-sensical concentration data, values set to NA |
| 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 | location-resampled | The X,Y location where sample was collected has already been visited, deviates from NEON soil sampling design |

Additionally, several other condition and quality fields have been added to the data over time in order to communicate anomalous sample conditions, analytical results, or method deviations. Definitions for all of the categorical codes used in QF fields are included in the file NEON Categorical Codes for Soil physical and chemical properties, periodic (AD[06]), provided in the download package for this data product. Fields have been added over time and entries may be missing in older data.

In the fields **ammoniumNQF** and **nitrateNitriteNQF**, the entry ‘blank value outlier’ means a blank was flagged as an outlier compared to the other replicates and concentration values were removed. An entry of ‘blanks exceed sample value’ is used when concentration values blank-correct to significantly less than zero (< -0.02 mg/L), as explained in Section 3.10. Users have varied options for how to treat these apparent negatives, several of which were explored and tested in Weintraub-Leff et al (2023). One option is to leverage a statistical correction developed in that paper for pre-2022 data, which accounts for nitrite contamination in KCl. This correction is integrated into the [neonNTrans](#) package. See the package readme and function help files more detail on how the correction is applied.

4.5 Analytical Facility Data Quality

External laboratories that generate inorganic N data for NEON include standards run as unknowns alongside samples in order to gauge run acceptability. Long-term analytical precision and accuracy of these analyses are reported to allow users to interpret and analyze inorganic N data in the context of their uncertainty ranges. The data table `ntr_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

Homyak, P.M., Vasquez, K.T., Sickman, J.O., Parker, D.R. and Schimel, J.P. (2015). Improving Nitrite Analysis in Soils: Drawbacks of the Conventional 2 M KCl Extraction. *Soil Science Society of America Journal*, 79: 1237-1242. doi:10.2136/sssaj2015.02.0061n.

Weintraub-Leff, S.R., Hall, S.J., Craig, M.E., Sihi, D., Wang, Z. and Hart, S.C. (2023). Standardized data to improve understanding and modeling of soil nitrogen at continental scale. *Earth's Future*, 11, e2022EF003224. doi.org/10.1029/2022EF003224.

6 APPENDIX

In order to estimate blank-corrected N concentrations in KCl extracts, then convert to micrograms N per gram dry soil (per day for net rates), users will need to combine data from four tables, then conduct a series of simple calculations. *Note that this cannot be done until N concentration data are available, and sometimes this lags behind laboratory extraction metadata and soil moisture.* The steps entailed in the calculations are detailed below. For users comfortable with R, there is a simple package called [neon-NTrans](#) that will conduct these calculations for you, with installation and usage instructions available in the package readme on GitHub. The `def.calc.ntrans` function creates several new terms not contained in the file NEON Data Variables for Soil physical and chemical properties, periodic (DP1.10086.001) (AD[05]), thus they are defined below:

| fieldName | definition | units |
|----------------------------------|---|------------------------------------|
| blank1ID | identifier of the first blank associated with a set of samples | NA |
| blank2ID | identifier of the second blank associated with a set of samples | NA |
| blank3ID | identifier of the third blank associated with a set of samples | NA |
| blank1NH4 | ammonium - N concentration of the first blank associated with a set of samples | milligrams per liter |
| blank2NH4 | ammonium - N concentration of the second blank associated with a set of samples | milligrams per liter |
| blank3NH4 | ammonium - N concentration of the third blank associated with a set of samples | milligrams per liter |
| blank1NO3 | (nitrate + nitrite) - N concentration of the first blank associated with a set of samples | milligrams per liter |
| blank2NO3 | (nitrate + nitrite) - N concentration of the second blank associated with a set of samples | milligrams per liter |
| blank3NO3 | (nitrate + nitrite) - N concentration of the third blank associated with a set of samples | milligrams per liter |
| blankNH4mean | mean ammonium - N concentration of the blanks associated with a set of samples | milligrams per liter |
| blankNO3mean | mean (nitrate + nitrite) - N concentration of the blanks associated with a set of samples | milligrams per liter |
| kclAmmoniumN BlankCor | blank-corrected ammonium - N concentration of a sample, set to zero if negative | milligrams per liter |
| kclNitrateNitriteN BlankCor | blank-corrected (nitrate + nitrite) - N concentration of a sample, set to zero if negative | milligrams per liter |
| noxCorrection | was the quantile regression approach outlined in Weintraub-Leff et al (2023) used to correct for contaminant nitrite. Options for pre-2022 data are 'implemented' or 'not implemented'. For 2022 and onward always 'not applicable' | NA |
| soilDryMass | dry mass of soil used for KCl extractions | grams |
| soilAmmoniumNug PerGram | extractable ammonium - N content in sample | micrograms N per gram soil |
| soilNitrateNitriteNug PerGram | extractable (nitrate + nitrite) - N content in sample | micrograms N per gram soil |
| soilInorganicNug PerGram | extractable inorganic N content in sample | micrograms N per gram soil |
| netNminugPerGram PerDay | net N mineralized over the course of the incubation | micrograms N per gram soil per day |
| netNitugPerGram PerDay | net nitrification over the course of the incubation | micrograms N per gram soil per day |

For users who wish to conduct the N transformation calculations on their own, the following table lists the key variables needed as inputs and in which data product tables they are found.

| tableName | fieldName | units |
|-----------------------|------------------------|----------------------|
| ntr_externalLab | kclAmmoniumNConc | milligrams per liter |
| ntr_externalLab | kclNitrateNitriteNConc | milligrams per liter |
| ntr_internalLabBlanks | kclReferenceID | NA |
| ntr_internalLab | incubationLength | days |
| ntr_internalLab | kclVolume | milliliters |
| ntr_internalLab | soilFreshMass | grams |
| sls_soilMoisture | dryMassFraction | unitless |

Before conducting the calculations, users may wish to review the data product tables and decided whether they'd like to remove/ignore any values with anomolous **sampleCondition** or **dataQF** values. Additionally, they may want to exclude concentration values in ntr_externalLab with specific **ammonium-NQF** or **nitrateNitriteNQF** values, or apply the statistical correction for contaminant nitrite as described in Weintraub-Leff et al (2023). Note that the possibility to correct the data and/or exclude data for specific condition or QF values is part of the def.calc.ntrans function described above.

Step-by-step calculations:

1. Calculate mean blank ammounium - N concentration for each sample set using the 1-3 blanks associated with each **kclReferenceID**.

$$\overline{(NH_4^+ - N)}_B = \frac{(NH_4^+ - N)_{B1} + \dots + (NH_4^+ - N)_{Bn}}{n}$$

2. Calculate mean blank (nitrate + nitrite) - N concentration for each sample set using the 1-3 blanks associated with each **kclReferenceID**. (nitrate + nitrite) - N is abbreviated as NO₃⁻-N in the equations below.

$$\overline{(NO_3^- - N)}_B = \frac{(NO_3^- - N)_{B1} + \dots + (NO_3^- - N)_{Bn}}{n}$$

3. Calculate corrected ammonium - N concentrations for each sample by subtracting the mean value of the blanks for that sample set (use the **kclReferenceID** to associated samples with blanks). If analytical replicates were conducted, either take the mean value or keep only one of the reps.

$$(NH_4^+ - N)_{cor} = (NH_4^+ - N)_{Sam} - \overline{(NH_4^+ - N)}_B$$

4. Calculate corrected (nitrate + nitrite) - N concentrations for each sample by subtracting the mean value of the blanks for that sample set (use the **kclReferenceID** to associated samples with blanks). If analytical replicates were conducted, either take the mean value or keep only one of the reps.

$$(NO_3^- - N)_{cor} = (NO_3^- - N)_{Sam} - \overline{(NO_3^- - N)_B}$$

4a. In order to manually apply the noxCorrection, use this equation for 2021 and earlier data. -0.77 is the regression slope of the 0.01 quantile for 2021 and earlier data comparing blank nitrate+nitrite N concentration to blank-corrected sample nitrate+nitrite N concentration.

$$(NO_3^- - N)_{cor} = ((NO_3^- - N)_{Sam} - \overline{(NO_3^- - N)_B}) + 0.77 * \overline{(NO_3^- - N)_B}$$

5. Convert the **soilFreshMass** used in KCl extractions to soil dry mass by multiplying by the **dryMassFraction**.

$$soilDryMass = soilFreshMass * dryMassFraction$$

6. Convert to micrograms ammonium - N per gram soil by multiplying the corrected ammonium - N concentration by the **kclVolume** used in extractions and dividing by the soil dry mass calculated in the step above.

$$(NH_4^+ - N)_{soil} = \frac{(NH_4^+ - N)_{cor} * kclVolume}{soilDryMass}$$

7. Convert to micrograms (nitrate + nitrite) - N per gram soil by multiplying the corrected (nitrate + nitrite) - N concentration by the **kclVolume** used in extractions and dividing by the soil dry mass calculated above.

$$(NO_3^- - N)_{soil} = \frac{(NO_3^- - N)_{cor} * kclVolume}{soilDryMass}$$

8. Use the **incubationPairID** and **nTransBoutType** variables to arrange cores in initial and final pairs.

9. Calculate net N mineralization using a core pair as follows: total inorganic N in final core minus total inorganic N in initial core, divided by **incubationLength** of the final core.

$$netNmin = \frac{((NH_4^+ - N)_{cor} + (NO_3^- - N)_{cor})_{final} - ((NH_4^+ - N)_{cor} + (NO_3^- - N)_{cor})_{initial}}{incubationLength}$$

10. Calculate net nitrification using a core pair as follows: (nitrate + nitrate) - N in final core minus (nitrate + nitrate) - N in initial core, divided by **incubationLength** of the final core.

$$netNit = \frac{((NO_3^- - N)_{cor})_{final} - ((NO_3^- - N)_{cor})_{initial}}{incubationLength}$$