NEON USER GUIDE TO PARTICULATE MASS (NEON.DP1.00101)

PREPARED BY	ORGANIZATION	DATE
Robert Lee	FIU	04/20/2018
		04/20/2018

CHANGE RECORD

REVISION	DATE	DESCRIPTION OF CHANGE
A	07/31/2017	Initial Release

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, the dry weights of litter functional groups from a single collection event are considered the lowest level (Level 0). Raw data that have been quality checked via the steps detailed herein, as well as simple metrics that emerge from the raw data are considered Level 1 data products.

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

1.2 Scope

This document describes the steps needed to generate the L1 data product Particulate Mass - the pre- and post-deployment masses of PM10 quartz microfiber filters - and associated metadata from input data. This document also provides details relevant to the publication of the data products via the NEON data portal, with additional detail available in the file, NEON Data Variables for Particulate Mass (NEON.DP1.00101) (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 data collected in the field pertaining to TIS Protocol and Procedure: Particulate Analyzer - Mass (AD[05]). The raw data that are processed in this document are detailed in the file, NEON Raw Data Validation for Particulate Mass (NEON.DP0.00101) (AD[03]), 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.002652	NEON Level 1, Level 2 and Level 3 Data Products Catalog
AD[03]	NEON.DP0.00101.001 _dataValidation.csv	NEON Raw Data Validation for Particulate Mass (NEON.DP0.00101)
AD[04]	NEON.DP1.00101.001 _variables.csv	NEON Data Variables for Particulate Mass (NEON.DP1.00101)
AD[05]	NEON.DOC.00XXXX	TIS Protocol and Procedure: Particulate Analyzer - Mass
AD[06]	NEON.DOC.000008	NEON Acronym List
AD[07]	NEON.DOC.000243	NEON Glossary of Terms
AD[08]		NEON's Ingest Conversion Language (NICL) specifications
AD[09]		NEON Algorithm Theoretical Basis Document: OS Generic Transitions

3 DATA PRODUCT DESCRIPTION

The mass of atmospheric particulate matter (PM), size fraction 10 m, is measured by a particulate mass analyzer using high volume sampling method (typically sampling 1 cubic meter or more of air per second). The high volume sampling method collects particulate matter on quartz microfiber filters, which are archived and available upon request to users.

Particulate matter is collected continuously at the tower top of selected NEON terrestrial towers using an EcoTech Hivol 3000 collector, which pumps air through a filter at a rate of approximately 68 cubic meters per minute. The filters used are quartz microfiber filters with a porosity of 10 m. The filters are conditioned according to the EPA Compendium Method IO-2.1 (ER[01]) at an external analytical facility. After conditioning, a pre-deployment weight is determined by the analytical facility, and filters are sent to NEON for field deployment. Filters are deployed in the instrument for two weeks, after which they are collected and returned to the analytical facility for re-conditioning and post-deployment weighing. These weights along with sensor data on pressure, air volumes, and temperature are reported as the L1 data product.

3.1 Spatial Sampling Design

Particulate mass sampling is executed at 6 of NEON's terrestrial sites, located in Domains 10, 13, and 15. The subset of sites included for sampling are those in the Basin and Range, Eastern and Western slopes of the Rocky Mountains, and the Eastern plains of Colorado. This selection of sites enables focus on transportation of particulate matter from the Great Basin and the Colorado Plateau by prevailing westerly winds over the Colorado Rocky Mountains, to receptor sites in the Rockies and Great Plains.

3.2 Temporal Sampling Design

Technicians service the instrument on a bi-weekly basis, and retrieve sample during instrument service. Sample retrieval is intended to occur every 14 days, however the schedules of field technicians can deviate from that schedule due to factors such as safety concerns or temporary lack of personnel. Additionally, technicians are instructed not to retrieve samples during precipitation events, which can delay sample collection. The maximium expected number of samples per site per year is 26.

3.3 Variables Reported

All variables reported from the field collection and laboratory analysis (L0 data) are listed in the file, NEON Raw Data Validation for Particulate Mass (NEON.DP0.00101) (AD[03]). All variables reported in the published data (L1 data) are also provided separately in the file, NEON Data Variables for Particulate Mass (NEON.DP1.00101) (AD[04]).

NEON TIS spatial data employs the World Geodetic System 1984 (WGS84) for its fundamental reference datum and Earth Gravitational Model 96 (EGM96) for its reference gravitational ellipsoid. Latitudes and longitudes are denoted in decimal notation to 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.

3.4 Spatial Resolution and Extent

The finest resolution at which spatial data are reported is the point location of the collector.

The basic spatial data in the data download package include the latitude, longitude, and elevation of the collector, plus associated uncertainty due to GPS error. Sampling at terrestrial sites always occurs at the tower top. Sensor geolocation data are given as x, y and z offsets in meters, relative to the base of the tower CD leg. The elevation of the collector can be calculated by adding the z offset to the elevation of the tower leg. The center of the collector air intake is given with the x and y offsets. An absolute location of the center of the collector air intake can be calculated by trigonometric addition of the x and y offsets to the location of the CD leg of the tower. More details on this procedure will be provided once location metadata has been published for NEON terrestrial towers.

Additional spatial data can be accessed via the NEON API or the geoNEON package: https://github.com/NEONScience/NEON-geolocation

During NEON construction, spatial data may not be available yet for some locations, and the spatial fields in downloaded data may be blank.

3.5 Temporal Resolution and Extent

The finest resolution at which temporal data are reported is the approximately bi-weekly range between **setDate** and **collectDate**.

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 stack files across months is available here: https://github.com/NEONScience/NEON-utilities

3.6 Associated Data Streams

The Dust and Particulate Size Distribution data product (DP1.00017.001) is closely related to Particulate Mass, as Dust and Particulate Size Distribution measurements are taken in parallel with Particulate Mass sampling at the same six NEON terrestrial sites. Data from this intstrument may be correlated with Particulate Mass sensor data (in dpm_sensor, in the expanded data package) based on the timestamps from both instruments.

3.7 Product Instances

Collection of samples occurs on a bi-weekly basis, with an estimated maximum of 26 sampling events per year per site, for a total of 156 records across the observatory each year. Factors such as precipitation or other delays in sample collection may reduce the number of collection events.

3.8 Data Relationships

The protocol dictates that each sample collection event corresponds to one unique **sampleID** in dpm_field. A record from dpm_field will have a corresponding (and identical) **sampleID** in dpm_lab. Laboratory weighing data will be returned for a **sampleID** only if sampling and lab weighing are successful; partial records caused by lost or damaged filters will be reported in dpm_field but will not have corresponding records in dpm_lab.

dpm_field -> One record per **sampleID** for all time.

dpm_lab -> One record per **sampleID** for all time.

In each table, a **filterID** may also be present. This represents the manufacturer-produced ID stamped on each filter, and may be used to identify an individual sample, though these IDs may not always be present in a record.

Sensor data is provided in the dpm_sensor table. Sensor data coresponding to the filter deployment period can be extracted by confining sensor data timestamps to those between the **setDate** and **collectDate** values given in dpm_field.

3.9 Special Considerations

Many users will want to calculate the volume of air that passed through the filter during its deployment, in order to derive the flux rate of particulate deposition. This can be calculated using **corrAirVolume** in the dpm_sensor table, and the time range indicated by **setDate** and **collectDate** in the dpm_field table. Simple subtraction of of the **corrAirVolume** values corresponding to **setDate** and **collectDate** will yield the total volume sampled during the filter deployment, corrected to standard temperature and pressure.

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. 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 applications and database ingest are described in the document NEON Raw Data Validation for Particulate Mass (NEON.DP0.00101), 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. 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[08]).

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[09]).

4.3 Data Revision

All data are provisional until a numbered version is released; the first release of a static version of NEON data, annotated with a globally unique identifier, is planned to take place in 2020. During the provisional period, QA/QC is an active process, as opposed to a discrete activity performed once, and records are updated on a rolling basis as a result of scheduled tests or feedback from data users. The Change Log section of the data product readme, provided with every data download, contains a history of major known errors and revisions.

4.4 Quality Flagging

The **dataQF** field in each data record is a quality flag for known errors applying to the record. There are currently no **dataQF** codes in use in this data product.

Additionally, there are multiple fields comment on data quality. In the **dpm_lab** table, both **externalRemarks** and **labQARemarks** are used to convey remarks about filter and measurement conditions, while **fieldFilterCondition** and **fieldFilterDamage** in the table are used to convey infromation on sample/filter health. The **equipmentCondition** field in **dmp_field** is also used to provide information on the health of the sampling equipment inte field.