NEON Algorithm Theoretical Basis Document for AOS Secchi Depth and Depth Profile Sampling: QA/QC of Raw Field Data

PREPARED BY	ORGANIZATION	DATE
Stephanie Parker	AQU	5/30/14

APPROVALS (Name)	ORGANIZATION	APPROVAL DATE
Charlotte Roehm	AQU	06/10/2014
Francois Pradeau	CYI	02/05/2015
Jennifer DeNicholas	CM	06/02/2014
Mike Stewart	SYS	08/28/2014

RELEASED BY (Name)	ORGANIZATION	RELEASE DATE
Judy Salazar	CM	02/12/2015

See Configuration Management System for approval history.

Change Record

REVISION	DATE	ECO #	DESCRIPTION OF CHANGE
A	02/12/2015	ECO-01959	Initial release

TABLE OF CONTENTS

1 DESCRIPTION

1.1 Purpose

This document details the algorithms used for creating a subset of NEON Level 1 data products that are the quality controlled products generated from raw Level 0 data, and associated metadata. In the NEON data products framework, the raw data collected in the field, for example, hind foot length of an individual small mammal, are considered the lowest level (Level 0). Raw data that have been quality checked via the algorithms detailed herein, as well as simple metrics that emerge from the raw data, such as total species richness of small mammals at a particular site, are considered Level 1 data products. This document relates only to the former group of L1 data products, the quality controlled pass-through products from the Level 0 data products.]

It includes a detailed discussion of measurement theory and implementation, appropriate theoretical background, data product provenance, quality assurance and control methods used, approximations and/or assumptions made, and a detailed exposition of uncertainty resulting in a cumulative reported uncertainty for this product.

1.2 Scope

This document describes the theoretical background and entire algorithmic process for creating a subset of quality controlled and calibrated L1 data products and associated metadata from input data. These data products include lake Secchi depth, non-wadeable stream Secchi depth, lake depth profile, and non-wadeable stream depth profile (Tables 1 - 2). It does not provide computational implementation details, except for cases where these stem directly from algorithmic choices explained here. This document also provides details relevant to the publication of the data products via the NEON data portal (AD[16]).

This document describes the algorithms for ingesting and performing automated quality assurance and control procedures on the data collected in the field pertaining to Algae sampling in lakes and non-wadeable streams (AD[08]), Zooplankton sampling in lakes (AD[09]), Microbes in lakes and non-wadeable streams (AD[10]), Surface water chemistry sampling in lakes and non-wadeable streams (AD[11]), and Sediment chemistry in lakes and non-wadeable streams (AD[12]). The raw data that are processed in this document are detailed in the NEON Raw Data Ingest Workbook for AD[15].

2 RELATED DOCUMENTS AND ACRONYMS

2.1 Applicable Documents

AD[01]	NEON.DOC.000001	NEON Observatory Design (NOD) Requirements
AD[02]	NEON.DOC.005003	NEON Scientific Data Products Catalog
AD[03]	NEON.DOC.005004	NEON Level 1-3 Data Products Catalog
AD[04]	NEON.DOC.005005	NEON Level 0 Data Product Catalog
AD[05]	NEON.DOC.005011	NEON Coordinate Systems Specification
AD[06]	NEON.DOC.001247	NEON ATBD for QA/QC Plausibility Testing of Organismal and Observation Based Field and Lab Data
AD[07]	NEON.DOC.004309	NEON Field Site Information
AD[08]	NEON.DOC.001203	Algae sampling in lakes and non-wadeable streams
AD[09]	NEON.DOC.001194	Zooplankton sampling in lakes
AD[10]	NEON.DOC.001200	Microbes in lakes and non-wadeable streams
AD[11]	NEON.DOC.001190	Surface water chemistry sampling in lakes and non-wadeable streams
AD[12]	NEON.DOC.001191	Sediment chemistry in lakes and non-wadeable streams
AD[13]	NEON.DOC.001626	Algorithm Basis Theoretical Document: QA/QC of Aquatic General Field Metadata
AD[14]	NEON.DOC.001627	Raw Data Ingest Workbook for AQU General Field Metadata
AD[15]	NEON.DOC.001695	Raw Data Ingest Workbook for AOS Secchi Depth and Depth Profile Sampling
AD[16]	NEON.DOC.001696	Data Publication Workbook for AOS Secchi Depth and Depth Profile Sampling

2.2 Reference Documents

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

2.3 Acronyms

Acronym	Definition
ATBD	Algorithm Theoretical Basis Document
DO	Dissolved oxygen
GLEON	Global Lake Ecological Observatory Network

LTER	Long Term Ecological Research
m^{-2}	Per square meter
mm	Millimeter
NAWQA	National Water-Quality Assessment (USGS)
NLA	National Lakes Assessment (USEPA)
NRSA	National Rivers and Streams Assessment (USEPA)
RBP	Rapid Bioassessment Protocol (USEPA)
USEPA	US Environmental Protection Agency
USGS	US Geological Survey

3 DATA PRODUCT DESCRIPTION

Several aquatic data products rely on Secchi depth and depth profile data, which help to determine the depth of the euphotic zone and whether or not the lake or river is stratified.

A Secchi disk is a black and white disk that is used to determine the transparency of water. For NEON sampling, two Secchi disk measurements will be taken at each location, and the mean of the two measurements will be used as the final data product.

Depth profiles determine whether or not the lake or river is stratified, where the body of water is separated into two or more layers based on temperature. Typically, in a stratified system, the top layer of water is warmer (epilimnion) while the lower layer is colder (hypolimnion). The area separating the two layers is known as the thermocline. If the lake or river is stratified, different sampling procedures take places in the associated protocols (AD[08-12]).

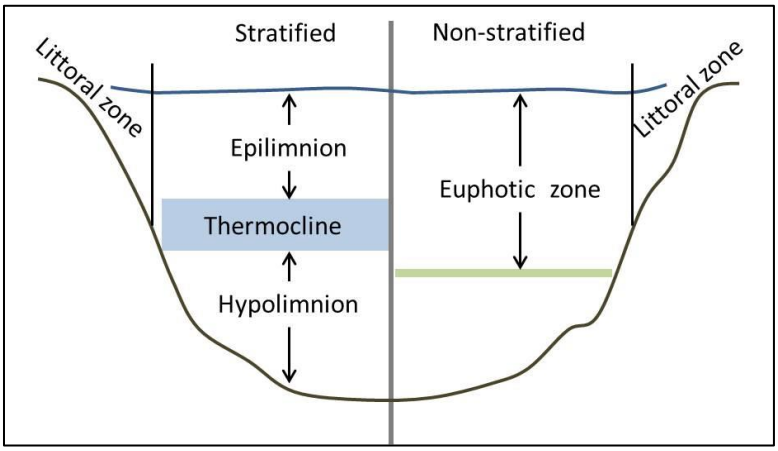

Figure 1. Diagram representing the zones in a stratified and non-stratified lake

3.1 Variables Reported

This ATBD describes the steps needed to generate the L1 data products lake Secchi depth, non-wadeable stream Secchi depth, lake depth profile, and non-wadeable stream depth profile.

Subproducts for these data products are listed below (Table 1 – 2). Detailed lists of the associated subproducts and metadata products are provided separately, along with example data in publication-ready spreadsheets (AD[16]). 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), and with the Bird Monitoring Data Exchange standards (http://www.avianknowledge.net; accessed 16 February 2014), where applicable. Geospatial data shall conform to the standards set forth in the NEON Coordinate Systems Specification (AD[05]).

Table 1. List of subproducts produced in this ATBD for the dataproducts lake Secchi depth and non-wadeable stream Secchi depth. The list is not exhaustive and a variety of supporting data will also be made available.

Number	Subproduct (Field Name)	Description
	secchi1Depth	First Secchi disk depth at sampling location (m)
	secchi2Depth	Second Secchi disk depth at sampling location (m)
	secchiMeanDepth	Mean of two Secchi disck depths at sampling location (m)
	euphoticDepth	Depth of euphotic zone (m)
	totalDepth	Depth of lake or river at sampling location (m)

Table 2. List of subproducts produced in this ATBD for the dataproducts lake depth profile and non-wadeable stream depth profile. The list is not exhaustive and a variety of supporting data will also be made available.

Number	Subproduct (Field Name)	Description
	depth	Depth of record
	DO	Dissolved oxygen at given depth (%)
	waterTemperature	Water temperature at given depth (^{o}C)
	specificConductivity	Specific conductivity at given depth (uS/cm)
	thermalStratification	Indication of whether the water column is stratified or non-stratified
	epilimnionDepth	Depth of epilimnion (m)

3.2 Temporal Resolution and Extent

The finest temporal resolution that Secchi depth and depth profiles will be tracked is per sampling bout. Each sampling bout at a site typically occurs within one day. Algae (AD[08]), zooplankton (AD[09]), and sediment chemistry (AD[12]) protocols are sampled during three bouts per year. Surface water microbes (AD[10]) and surface water chemistry (AD[11]) occur during 6 sampling bouts per year.

3.3 Spatial Resolution and Extent

Each set of Secchi measurements or depth profile measurements represents the water column at one location in the lake or non-wadeable stream. These measurements will be recorded at up to three locations per site per sampling bout. Overall, this results in a spatial hierarchy of:

decimalLatitude, decimalLongitude → locationID → habitatType → siteID → domainID

3.4 Associated Data Streams

The Secchi depth and depth profile field measurement data are linked to the ATBD for Aquatic General Field Metadata (AD[13], AD[14]). Database tables will be linked to the Secchi depth and depth profile data publication workbook (AD[16]) directly through **siteID** and **collectDate**.

Secchi depth and depth profile field measurement data are also linked to the sampling protocol that occurs at the same time, which will be linked through **siteID** and **collectDate**:

Algae sampling in lakes and non-wadeable streams (AD[08])
Zooplankton sampling in lakes (AD[09])
Microbes in lakes and non-wadeable streams (AD[10])
Surface water chemistry sampling in lakes and non-wadeable streams (AD[11])
Sediment chemistry in lakes and non-wadeable streams (AD[12])

3.5 Product Instances

The data collected from Secchi depth and depth profile measurements are used to inform the sampling depths of five field collection protocols on each sampling date (AD[08-12]). Data will be collected three to six times per year per protocol, depending on NEON Field Operations Schedules.

4 SCIENTIFIC CONTEXT

4.1 Theory of Measurement/Observation

Secchi depth is a quick way to determine the depth of the euphotic zone in a lake or non-wadeable stream, allowing the observer to determine a coarse estimate of the depth of transparency in the water column. Transparency of water depends on the turbidity and density of organisms in the water column. Phytoplankton are present in higher density above the metalimnion in stratified lakes (lakes with a thermocline) and within the euphotic zone (the region through which light penetrates) in non-stratified lakes (Wehr and Sheath 2003). Turbid waters where sediments are suspended have shallow Secchi depths, as do waters that have phytoplankton blooms. The limit of visibility in the water column is the depth to which ~5% of light penetrates, although the bottom of the euphotic zone is the depth that 1% of light penetrates, which is typically 2-3 times the Secchi depth. NEON will multiply the Secchi depth depth by 2.5 to estimate the bottom of the euphotic zone (Mischke et al. 2012)

Depth profiles determine the thermal stratification of a lake or non-wadeable stream. A thermocline develops in bodies of water where there is little mixing. These water bodies tend to be deeper (>6 m) and tend to have a period of times where the upper waters are heated quickly, such as in higher-latitude lakes in early summer (Wetzel 2001). In a stratified lake or non-wadeable stream, there will be a noticeable change in temperature at the depth of the thermocline, which determines where chemistry, microbe, or phytoplankton samples are collected in the water column.

4.2 Theory of Algorithm

This document describes the algorithms for assessing the integrity of the L0 data stream generated by the field measurements of Secchi depth and depth profiles. The approaches described below are necessary components of quality control and quality assurance, including defining the range of possible values for each data product, specifying the lookup tables that contain accepted values for particular data products

4.3 Special Considerations

Secchi depth and depth profile data are are collected in the field and used as metadata to determine the depth of water chemistry, microbe, phytoplankton, and zooplankton sample collectionsin NEON lake and non-wadeable stream sites. Although used by NEON as metadata, these data products are provided as L1 data products via the web portal because they are common measurements used by limnologists taken at many sites across the world.

5 ALGORITHM IMPLEMENTATION

Automated Processing Steps for Field Collected Data

5.1 Run the following processing steps for all data in the Secchi depth and depth profile data ingest workbook (AD[15])

1. Assign data values, in the secchi depth and depth profile database, from L0 data values
    a. Run **Assign: L1 Data from L0 Data**, in AD[06], where:
        a. List of data ingest sheets = AD[15]:*dep_secchi_in* and AD[15]:*dep_depthprofile_in*
        b. List of database tables = *dep_secchi_db* and *dep_depthprofile_db*
2. Use AD[15]:*dep_datasummary_in* to link controlled vocabulary field names to data entered by technicians
    a. For each values in the table field of AD[15]:*dep_datasummary_in*:
        a. Locate the database table that is named according to the value, but with '_db' on the end as opposed to '_in'
            i. Example: where value in table field == *'dep_secchi_in'*, database table of interest == *'dep_secchi_db'*
        b. For each value in the **dataEntry** field of the database table of interest:
            i. Use the mapping in AD[15]:*dep_tablesummary_in* to replace fieldnames (entries found in **dataEntry** field of AD[15]:*dep_tablesummary_in*) in the database table of interest with the correct field names (entries found in **fieldName** field of AD[15]:*dep_tablesummary_in*
3. Verify that all records are complete
    a. Run **Validation Test: Complete Records**, in AD[06], where:
        a. List of data ingest sheets = AD[15]:*dep_secchi_in,* AD[15]:*dep_depthprofile_in*
        b. List of database tables = *dep_secchi_db, dep_depthprofile_db*
4. Verify that all records of all fields are of the correct data type

a. Run **Validation Test: Data Type**, in AD[06], where:
    a. List of data ingest sheets = AD[15]:*dep_secchi_in,* AD[15]:*dep_depthprofile_in*
    b. List of database tables = *dep_secchi_db, dep_depthprofile_db*

5. Convert **date** to correct format and verify that this is within acceptable range
    a. Run **Validation Test: Date**, in AD[06], where:
        a. List of data ingest sheets = AD[15]:*dep_secchi_in,* AD[15]:*dep_depthprofile_in*
        b. List of date fieldNames = (**date**)
        c. List of database tables = *dep_secchi_db, dep_depthprofile_db*
6. Verify that data values are valid, as specified by validation rules
    a. Run **Validation Test: Validation Rules**, in AD[06], where:
        a. List of data ingest sheets= AD[15]:*dep_secchi_in,* AD[15]:*dep_depthprofile_in*
        b. List of date fieldnames= (**date**)
        c. List of database tables= *dep_secchi_db, dep_depthprofile_db*
7. Verify that **siteID** values are valid for the site and domain in which data were collected
    a. Run **Validation Test: Location**, in AD[06], where:
        a. List of data ingest sheets= AD[15]:*dep_secchi_in,* AD[15]:*dep_depthprofile_in*
        b. List of database tables= *dep_secchi_db, dep_depthprofile_db*
8. Generate a unique ID (**uid**) for each record
    a. Run **Generate: Unique ID**, in AD[06]
        a. List of data ingest sheets= AD[15]:*dep_secchi_in,* AD[15]:*dep_depthprofile_in*
        b. List of database tables= *dep_secchi_db, dep_depthprofile_db*
9. Assign **domainID** for each record
    a. Run **Assign: Location IDs**, in AD[06], where
        a. List of database tables= *dep_secchi_db, dep_depthprofile_db*
10. Generate spatial uncertainty information for data publication
    a. Run **Generate: Spatial Information and Uncertainty** in AD[06], where:
        i. Database tables = *dep_secchi_db, dep_depthprofile_db*
        ii. Fieldname = **decimalLongitude**, **decimalLatitude**
        iii. Spatial table = **AOS Point-Level Spatial Data**
        iv. Subtype = **locationID**
11. Populate new fields from *dep_secchi_db* and *dep_depthprofile_db*
    a. Populate fieldnames **coordinateUncertainty** in *dep_secchi*_pub and *dep_depthprofile_pub* from *dep_secchi_db* and *dep_depthprofile_db*
        i. Link fieldnames from *dep_secchi*_pub and *dep_depthprofile_pub* to *dep_secchi_db* and *dep_depthprofile_db* using **siteID** and **date**
12. Generate quality flag summary
    a. Run **Generate: Quality Flag Summary**, in AD[06], where
        a. List of database tables= *dep_secchi_db, dep_depthprofile_db*

5.2 Run the following processing steps for the *dep_secchi_in* sheet in the Secchi depth and depth profile data ingest workbook (AD[15])

1. Verify **secchiMeanDepth**
    a. Generate **secchiMeanDepthQF** field in *dep_secchi_db* and populate with zeros
    b. For each row in AD[15]:*dep_secchi_in*:
        i. If value ≠ (**secchi1Depth** + **secchi2Depth**)/2
            A. Insert 1 into **secchiMeanDepthQF** field of *dep_secchi_db*
2. Verify **euphoticDepth**
    a. Generate **euphoticDepthQF** field in *dep_secchi_db* and populate with zeros
    b. For each row in AD[15]:*dep_secchi_in*:
        i. If value ≠ **secchiMeanDepth** * 2.5
            A. Insert 1 into **euphoticDepthQF** field of *dep_secchi_db*

5.3 Run the following processing steps for the *dep_depthprofile_in* sheet in the Secchi depth and depth profile data ingest workbook (AD[15])

1. Verify **thermalStratification**
    a. Generate **thermalStratificationQF** field in *dep_depthprofile_db* and populate with zeros
    b. For each row in AD[15]:*dep_depthprofile_in*:
        i. If **thermalStratification** = 'stratified', value in **epilimnionDepth** field = real number
2. Verify that conditional fields are complete
    a. Generate **completeEpilimnionDepth** in *dep_depthprofile_db* and populate with zeros
    b. For each row in **epilimnionDepth** field of *dep_depthprofile_db*:
        i. If value is NULL:
            A. Value in **thermalStratification** field = non-stratified

5.4 Run the following processing steps during the creation of *dep_afm_pub* from AD[14] AQU general field metadata ingest workbook

1. Populate new fields from AD[29]: *afm_perdate_db* in *dep_afm_pub*
    a. Populate all fieldnames *dep_afm_pub* from *afm_perdate_db*
    b. Link *afm_perdate_db* to *dep_afm_pub* using **siteID** and **date**

5.5 All databases: Run the following processing steps for all databases produced through the algorithms above

1. Generate quality flag summary
    a. Run **Generate: Quality Flag Summary**, in AD[06], where:
        i. List of database tables = *dep_secchi_db, dep_depthprofile_db*

6 UNCERTAINTY

There are multiple sources of uncertainty associated with the Secchi depth and depth profile data: error associated with depth measurements, observer error when reading the Secchi disk, or instrument error when measuring temperature, DO, and conductivity. These factors increase uncertainty in dependent data products and the five sampling protocols that use these data products (AD[08-12]).

6.1 Reported Uncertainty

Although no quantitative algorithms are available to incorporate many of these sources of uncertainty into the associated data products, NEON can produce summary uncertainty reports for observational data products. The summary reports will include: bout-level data entry error rates; automated QA/QC error rates for each data product per bout; and the precision of the measurement tools. The Secchi disk rope measures depth to the nearest 0.5 m. The handheld probes used to measure water temperature, conductivity, and DO measures temperature to with a resolution of 0.1 ^{o}C and accuracy of ± 1 degree, conductivity with a resolution of 0.1 uS/cm and accuracy of ± 0.5%, and DO with a resolution of 0.01 mg/L and accuracy of 1%.

7 VALIDATION AND VERIFICATION

7.1 Algorithm Validation

7.2 Data Product Validation

7.3 Data Product Verification

8 SCIENTIFIC AND EDUCATIONAL APPLICATIONS

NEON Secchi depth and depth profile data are collected similarly to several other monitoring programs across the U.S. Citizen science projects, as well as state monitoring programs (e.g., LakeWatch) also commonly track Secchi depth information. These data not only inform NEON sample procedures and depths, but also may be compared to suites of other data collected at lake sites across the continental U.S. and GLEON studies around the world.

9 FUTURE MODIFICATIONS AND PLANS

1. When personal digital assistant (PDA) devices become available for data entry in the field, a number of steps in this ATBD will require modification.
2. Guidelines for uncertainty calculations and reporting for all data products are currently lacking and need to be addressed where possible.
3. Guidelines for releasing revised data products.

10 BIBLIOGRAPHY

Mischke, U., S. Thackeray, M. Dunbar, C. McDonald, L. Carvalho, C. deHoyos, M. Jarvinen, C. Laplace-Treyture, G. Morabito, B. Skjelbred, A. Solheim, B. Brierly, and B. Dudley. 2012. Deliverable D3.1-4: Guidance document on sampling, analysis and counting standards for phytoplankton in lakes. WISER Deliverable D3.1-4, February 2012. http://www.wiser.eu/results/deliverables/

Wehr, J. D. and R. G. Sheath. 2003. Freshwater Algae of North America. Academic Press, San Diego, California. 918 pages.

Wetzel, R. G. 2001. Limnology: Lake and River Ecosystems, Third Edition. Academic Press, San Diego, Claifornia. 1006 pages.

11 CHANGELOG