NEON ALGORITHM THEORETICAL BASIS DOCUMENT: TOS HERBACEOUS BIOMASS – QA/QC OF RAW FIELD AND LAB DATA

PREPARED BY	ORGANIZATION	DATE
Courtney L. Meier	FSU	2014-09-16

APPROVALS	ORGANIZATION	APPROVAL DATE

RELEASED BY	ORGANIZATION	RELEASE DATE

See Configuration Management System for approval history.

Change Record

REVISION	DATE	ECO #	DESCRIPTION OF CHANGE

TABLE OF CONTENTS

1 DESCRIPTION ... 3
- 1.1 Purpose ... 3
- 1.2 Scope ... 3

2 RELATED DOCUMENTS AND ACRONYMS ... 4
- 2.1 Applicable Documents ... 4
- 2.2 Reference Documents ... 4
- 2.3 Acronyms ... 4

3 DATA PRODUCT DESCRIPTION ... 5
- 3.1 Variables Reported ... 5
- 3.2 Temporal Resolution and Extent ... 6
- 3.3 Spatial Resolution and Extent ... 7
- 3.4 Associated Data Streams ... 9
- 3.5 Product Instances ... 9

4 SCIENTIFIC CONTEXT ... 9
- 4.1 Theory of Measurement/Observation ... 9
- 4.2 Theory of Algorithm ... 10
  - 4.2.1 Summary of Algorithm for the Herbaceous biomass and productivity per bout data ... 10
  - 4.2.2 Summary of Algorithm for the Herbaceous biomass and productivity mass data ... 10
- 4.3 Special Considerations ... 10

5 ALGORITHM IMPLEMENTATION ... 11
- 5.1 Processing steps for data in hbp_perbout_in ... 11
- 5.2 Processing steps for data in hbp_massdata_in ... 13

6 UNCERTAINTY ... 15

7 VALIDATION AND VERIFICATION ... 15
- 7.1 Algorithm Validation ... 15
- 7.2 Data Product Validation ... 15
- 7.3 Data Product Verification ... 16

8 SCIENTIFIC AND EDUCATIONAL APPLICATIONS ... 16

9 FUTURE MODIFICATIONS AND PLANS ... 16

10 BIBLIOGRAPHY 18
11 CHANGELOG 19
12 APPENDIX: DATA ENTRY VALIDATION VIA PDA AND WEBUI 20
12.1 All Data: Run the following steps for all fields in in the Raw Data Ingest Workbook for TOS Herbaceous Plant Biomass (AD[10]) 20
12.2 Validation steps for hbp_perbout_in 20
12.3 Validation steps for hbp_massdata_in 21

LIST OF TABLES AND FIGURES

Table 1. List of subproducts and spatial and temporal data produced in this ATBD for the data product herbaceous clip harvest (NEON.DOM.SITE.DP1.10023).. 6

Figure 1. A 20m x 20m (400m^2) NEON plot showing potential clip-harvest "cells" (dashed blue lines). 8

Figure 2. A 40m x 40m (1600 m2) NEON Tower plot showing clip-harvest cells within a 400 m2 subplot. 8

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 and lab, for example, the dry weights of herbaceous functional groups, 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 dry weight herbaceous biomass, 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.

Herbaceous biomass will be harvested at all NEON sites, and there will be a minimum of one sampling bout per growing season. Clipped biomass will be sorted to functional group prior to drying and weighing in the laboratory, such that the contribution of functional groups to total herbaceous biomass may be assessed. At sites managed for grazing, as well as those sites that have more than one phenological peak in herbaceous biomass, there will be more than one clip-harvest sampling bout per growing season, and samples in these additional bouts will not be sorted to functional group. This document details the processes necessary to QA/QC Level 0 herbaceous clip harvest data. 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 the quality controlled and calibrated L1 herbaceous clip harvest data product, and associated metadata, from specified input data. This document 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 (NEON Data Publication Workbook for TOS Herbaceous Plant Biomass: QA/QC of Raw Field and Lab Data (AD[11])).

The algorithms described herein are designed to ingest data entered by NEON Field Operations technicians, and perform automated quality assurance and control procedures on the data collected in the field and lab that are associated with the TOS Field and Lab Protocol for Herbaceous Plant Biomass (AD[09]). The raw data that are processed in this document are detailed in the NEON Raw Data Ingest Workbook for TOS Herbaceous Plant Biomass (AD[10]). In addition to quality checked dry weight data per herbaceous functional group, this document details production of a suite of quality-checked spatial, temporal, and sampling metadata that are linked to the clip-harvest sampling event.

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 Algorithm Theoretical Basis Document: QA/QC Data Validation and Plausibility Testing of TOS and AOS Field and Lab Data
AD[07]	NEON.DOC.004309	NEON Field Site Information
AD[08]	NEON.DOC.000914	TOS Science Design for Plant Biomass, Productivity, and Leaf Area Index
AD[09]	NEON.DOC.014037	TOS Field and Lab Protocol for Herbaceous Biomass
AD[10]	NEON.DOC.001920	NEON Raw Data Ingest Workbook for TOS Herbaceous Biomass
AD[11]	NEON.DOC.001931	NEON Data Publication Workbook for TOS Herbaceous Biomass: QA/QC of Raw Field and Lab Data
AD[12]	NEON.DOC.002104	NEON Algorithm and Theoretical Basis Document: TOS Herbaceous Biomass – Aboveground Productivity Calculations
AD[13]	NEON.DOC. 002261	TOS Spatial Data
AD[14]	NEON.DOC.00XXXX	TOS Herbaceous Biomass Clip-Strip Coordinates Lookup

2.2 Reference Documents

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

2.3 Acronyms

AGB	Aboveground biomass
ANPP	Annual net primary productivity

3 DATA PRODUCT DESCRIPTION

The herbaceous clip harvest data products include total aboveground dry mass herbaceous production per plot per sampling bout. Additionally, for one sampling bout per growing season, the aboveground dry mass production associated with key herbaceous functional groups will be provided. These functional groups are defined as:

- Bryophytes (for which annual production is discernable)
- C3 cool-season graminoids
- C4 warm-season graminoids
- Leguminous forbs
- Non-leguminous forbs
- Woody-stemmed shrubs with diameter at decimeter height < 1 cm

All data products include spatial and temporal sampling details, as well as required metadata. Salient details associated with the temporal and spatial herbaceous biomass sampling design and protocol are presented in Sections 3.2 and 3.3. For additional information on the sampling design and the associated protocol, see TOS Science Design for Plant Biomass, Productivity, and Leaf Area Index (AD[08]), and the TOS Field and Lab Protocol for Herbaceous Plant Biomass (AD[09]). Following data collection in the field and the lab, several standard QA/QC procedures are implemented, as described in Section 5 of this document, in order to produce quality-checked Level 1 herbaceous clip harvest data products.

3.1 Variables Reported

This Algorithm Theoretical Basis Document (ATBD) describes the steps needed to generate the Level 1 data product: herbaceous clip harvest (NEON.DOM.SITE.DPL.10023). This data product is comprised of both subproducts and associated metadata, the latter of which are intended to provide a meaningful context to the end user for the herbaceous clip harvest mass subproducts. QA/QC checks described in this document are performed on both data subproducts and metadata. **Table 1** lists salient subproducts and metadata included within the herbaceous clip harvest data product, and an exhaustive list is provided in the NEON Data Publication Workbook for TOS Herbaceous Plant Biomass: QA/QC of Raw Field and Lab Data (AD[11]). Names of specific data and metadata fields listed in AD[11] 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 conform to the standards set forth in the NEON Coordinate Systems Specification (AD[05]).

Table 1. List of subproducts and spatial and temporal data produced in this ATBD for the data product herbaceous clip harvest (NEON.DOM.SITE.DP1.10023). The list is not exhaustive, and additional subproducts and supporting metadata are also available.

Number	Field Name	Description
	plotID	Unique plot identifier
	plotType	NEON plot type in which sampling occurred: Tower, Distributed or Gradient
	clipID	Unique identifier for the clip-harvest location within the plot
	eventID	An identifier for the set of information associated with the event, which includes information about the place and time of the event
	exclosure	Identifies whether clip harvest took place in an area protected by a grazing exclosure
	herbGroup	Categories indicating herbaceous plant functional group
	dryMass	Oven-dried mass of sample or subsample
	sampleID	Identifier for sample

3.2 Temporal Resolution and Extent

NEON field and lab technicians will generate the herbaceous clip harvest data product at least once per growing season for each site. However, the temporal sampling resolution depends on plot type: While Distributed plots located throughout the site are only ever harvested once per growing season, Tower plots may be sampled more than once per growing season if certain criteria are met. The default sampling frequency for Tower plots is once per year, but the following factors may result in additional sampling bouts per growing season, and hence additional instances of the data product, and are assessed on a per site basis:

- Sites with distinct early-season (C3 plants) and late-season biomass peaks (C4 plants) may be clip-harvested twice per growing season.
- Sites actively managed for grazing are clip-harvested every 4 weeks.
- Sites with continuous growth and decomposition (e.g., tropical and sub-tropical grasslands) are harvested more frequently than sites with periods of senescence.

The finest temporal resolution at which instances of the herbaceous clip harvest data product are tracked is the date of harvest for a given plot or subplot. Date values are assigned a boutNumber in the field, and the boutNumber, year of sampling, and siteID in which the clip harvest occurred are used to create a unique eventID associated with each herbaceous dry weight mass value. Harvested herbaceous biomass is sorted to functional group for a minimum of one sampling bout per growing season.

3.3 Spatial Resolution and Extent

The Herbaceous Clip Harvest protocol is implemented in all Tower plots, and in a maximum of twenty randomly selected Distributed plots per site. In Distributed plots, the maximum potential sampling effort (n=20 plots) corresponds to sampling in ½ to ⅔ of the established Distributed plots per site. The Distributed plots at which the herbaceous clip harvest protocol is performed are chosen randomly in order to create an unbiased estimate of herbaceous biomass at the site scale; some of the samples may be comprised of non-randomly located Gradient plots if Distributed plots fail to span the full dynamic range of total plant biomass at a given site (see AD[09] for details). Whether a plot was selected randomly or along a gradient is denoted in the plotType field of TOS Spatial Data (AD[13]).

For both plot types, the clip-harvest is performed in 0.1 m x 2 m clip strips that exist within 0.5 m x 3 m gridded clip "cells" (**Figure 1**). For typical sampling bouts, one gridded clip-area "cell" is randomly chosen per 400 m^2 plot or subplot. However, at sites managed for grazing, an additional grid cell is placed within a grazing exclosure, resulting in two clip-strips harvested per sampling bout. Grid cells that overlap with 25m^2 nested subplots (used for seedling and sapling counts) may be used for clip-harvests, but those that overlap with 1m^2 and 10m^2 nested subplots (used for diversity sampling) are never clipped. This ensures that generation of herbaceous plant productivity data products minimally affects the quality of plant biodiversity data products. The overall spatial hierarchy of the herbaceous clip harvest data product is subtly different for Tower plots compared to Distributed and Gradient plots. For Tower plots that are 40 m x 40 m and larger, the gridded clip harvest locations exist within subplotIDs within each plotID (**Figure 2**):

Tower plots (≥ 1600m^2): clipID → subplotID → plotID → habitat (NLCD veg class) → siteID → domainID

All other plots: clipID → plotID → habitat (NLCD veg class) → siteID → domainID

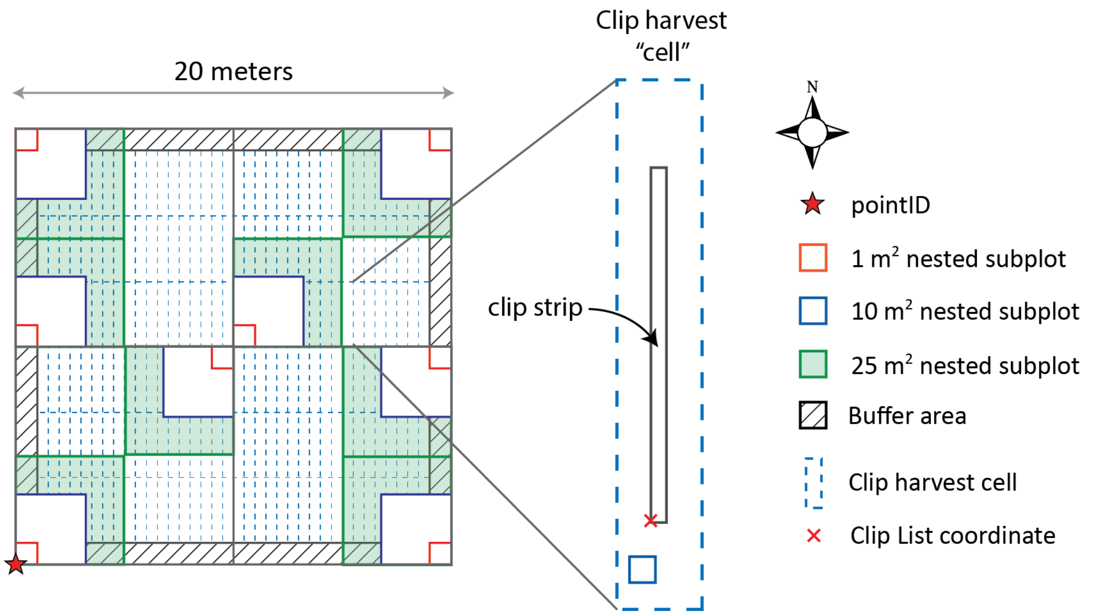

Figure 1. A 20m x 20m (400m^2) NEON plot showing potential clip-harvest "cells" (dashed blue lines). Cells overlapping 10m^2 nested subplots are not sampled for herbaceous biomass. When grazing managment is absent, one randomly selected grid cell is chosen for harvest per sampling bout. At sites managed for grazing, two grid cells are sampled per bout. The pointID (red star) is used to calculate the location in UTM of a given Clip List coordinate (red 'x').

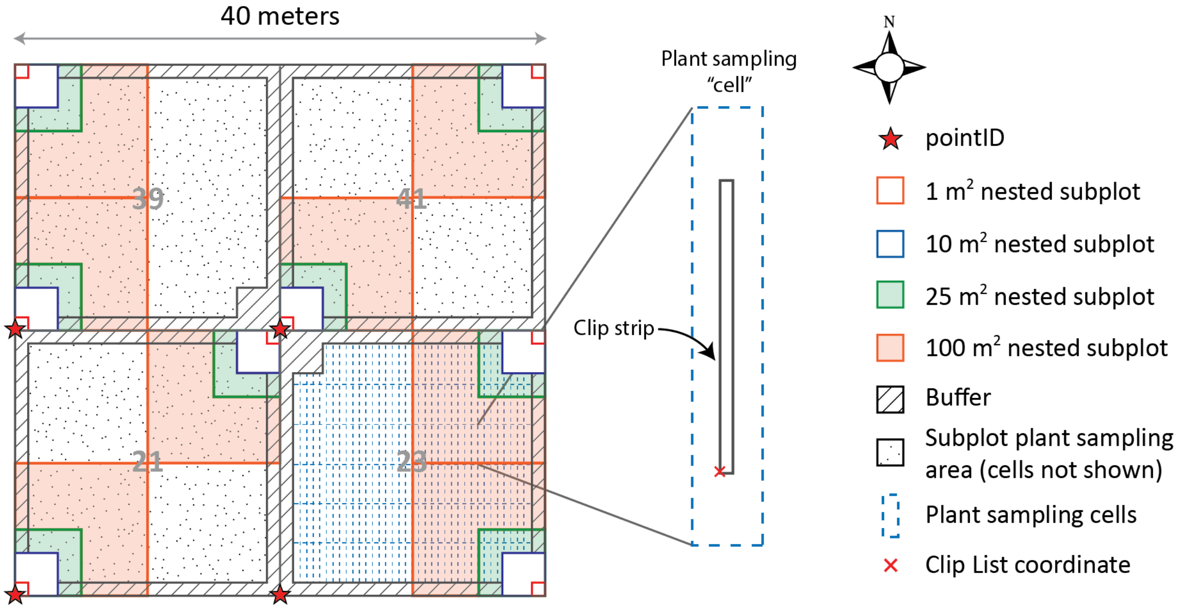

Figure 2. A 40m x 40m (1600 m2) NEON Tower plot showing clip-harvest cells within a 400 m2 subplot. Cells overlapping 10m^2 nested subplots are not sampled for herbaceous biomass. Clip harvest samples are generated from two randomly selected subplots per sampling bout, but for clarity, cells are not shown in 3 of 4 subplots. At sites managed for grazing, two grid cells per selected subplot are sampled per bout. For each subplotID (grey numbers), the pointID at the SW corner (red stars) is used to calculate the location in UTM of a given Clip List coordinate (red 'x').

3.4 Associated Data Streams

The QA/QC herbaceous clip harvest data product is used as an input to the herbaceous production data product, the latter of which is detailed in NEON ATBD: TOS Herbaceous Biomass – Aboveground Productivity Calculations (AD[12]). The plant material generated from herbaceous clip harvest may subsequently be analyzed for isotopes and chemistry. The results of these analyses can be found in NEON's plant foliar stable isotopes (NEON.DOM.SITE.DP1.10053) and plant foliar chemical properties (NEON.DOM.SITE.DP1.10051) data products.

3.5 Product Instances

The herbaceous clip harvest data product will be produced once per unique eventID per growing season from plots at all core and relocatable sites that contain qualifying herbaceous vegetation. Distributed plots qualify to produce the herbaceous clip harvest data product if the NLCD vegetation type is not forest (i.e., not equal to Deciduous Forest, Evergreen Forest, or Mixed Forest), and if aerial percent herbaceous cover is ≥ 50% (as assessed by field technicians). Tower plots qualify to produce the herbaceous clip harvest data product if herbaceous vegetation is ≥ 10% total cover, including understory cover. Plots that are not sampled due to insufficient qualifying cover are denoted with a 'targetTaxaPresent' = 'N'. For each plot with qualifying vegetation, dry mass values are reported for up to 6 herbaceous functional groups for each clipped "cell" per sampling event. Dry mass samples are then split, if sufficient sample is available, into two subsamples: one for chemical analysis, and another for archive.

4 SCIENTIFIC CONTEXT

4.1 Theory of Measurement/Observation

Herbaceous aboveground annual net primary productivity (ANPP) is defined as the amount of aboveground biomass (AGB) produced by herbaceous plants in a given year (or growing season when growing season spans multiple calendar years). NEON has selected long, narrow clip-harvest strips (0.1 m x 2 m) due to the fact that this shape results in the most consistent estimates of AGB and ANPP, compared to square clip areas, when plants may be distributed in a spatially heterogeneous manner. Conversion of above-ground biomass estimates to ANPP is not covered in this ATBD; but relevant equations can be found in Sala and Austin (2000), Sala et al. (1988), and Biondini et al. (1991).

Clipped herbaceous biomass will be sorted to functional group at least once per growing season because functional traits influence how plants respond to various change drivers (e.g. Collatz et al. 1998). The functional groups NEON will employ are listed in Section 3. When grazing by herbivores is significant, it is also necessary to account for consumption of AGB and compensatory regrowth by plants (McNaughton et al. 1996). At heavily grazed sites, NEON has adopted the standard method of sampling AGB inside and outside of a number of portable grazing exclosures that are moved at regular intervals throughout the growing/grazing season to facilitate ANPP estimation in these systems (Knapp et al. 2007).

4.2 Theory of Algorithm

This document describes algorithms needed for assessing the integrity of the Level 0 data stream generated by the field sampling described in AD[09]. The approaches described herein are simple yet necessary components of quality control and assurance, and include: validation tests for all data and metadata provided by Field Operations technicians, verifying that all required data are recorded for each sampling event, transformation of date and time variables into standard formats, generation of unique eventIDs, and association of the data with required spatial sampling information.

4.2.1 Summary of Algorithm for the Herbaceous biomass and productivity per bout data

1. Generate eventID of the form HBP.siteID.YYYY.boutNumber
2. Collate metadata about the plot from plot-level spatial data lookup table
3. Provide plot-level coordinate uncertainty associated with each clip-harvest location
4. Check for duplicate records
5. Check for missing records

4.2.2 Summary of Algorithm for the Herbaceous biomass and productivity mass data

1. Assign times to date values using matching records from the parent sampling event in hbp_perbout
2. Extract eventID from the parent sampling event in hpb_perbout
3. Collate metadata about the plot from plot-level spatial data lookup table
4. Calculate the duration of storage prior to oven-drying
5. Calculate the duration of oven-drying
6. Check for duplicate records
7. Check for missing records

4.3 Special Considerations

None.

5 ALGORITHM IMPLEMENTATION

Throughout the algorithm implementation section of this ATBD, 'nodata', 'null', and/or 'NA' indicates a blank cell. All variables reported from the field or laboratory technician (L0) are listed in the data ingest workbook (AD[10]) notated here as "hbp_*tablename*_in". Unless otherwise specified in the algorithm below, all variables that appear in tables "hbp_*tablename*_pub" have been passed directly from the L0 variables with the same name, as listed in the data publication workbook (AD[11]). Algorithm implementation should proceed in the order of the subsections provided here, e.g., the processing of "perbout" field collected data (subsection 5.1) should occur prior to that of laboratory-collected "massdata" (subsection 5.2), and so on.

The logic steps outlined below produce the Level 1 quality-checked herbaceous clip harvest data product from Level 0 raw field and lab data provided by Field Operations in AD[10].

5.1 Processing steps for data in hbp_perbout_in

1. For each complete group of related records submitted with the same **boutNumber,** generate a unique eventID using values from **siteID**, **date**, and **boutNumber** fields, and assign to corresponding **eventID** field of hbp_perbout_pub
    a. For the complete group of related records, extract values in the **date** field and sort. Examine the year ('YYYY') portion of the sorted **date** values, and assign the smallest value to the temporary variable 'year'
    b. To define the **eventID**, concatenate: 'HBP' + '**siteID'** + 'year' (as defined above) + **boutNumber,** separated by '.'
        i. *Example*: *HBP.SJER.2014.1*
2. Generate plot-level spatial information and uncertainty using **plotID** and the TOS plot-level spatial data lookup table (AD[13]). Get values for the following fields from AD[13], based on matching the **plotID** and where the subtype = 'basePlot' in AD[13], and insert these values into corresponding fields in hbp_perbout_pub:
    a. **plotType**
    b. **plotSize**
    c. **nlcdClass**
    d. **slopeAspect**
    e. **slopeGradient**
3. Generate point-level spatial information and uncertainty using **plotID**, **subplotID** and the TOS point-level spatial data lookup table (AD[13]), and adjusting the point location using offsets associated with each unique value of **clipCellNumber**.
    a. For each record, find the point-level spatial information in UTM format of the SW corner of the subplot using **plotID**, **subplotID** and the TOS point-level spatial data

lookup table. This location information includes: northing, easting, utmZone, geodeticDatum, coordinate uncertainty, elevation and elevation uncertainty. If point-level spatial information is not found for the given value of **subplotID**, set the point-level spatial fields of hbp_perbout_pub to null values.

b. If point-level spatial information is found, calculate the correct UTM easting and northing coordinates associated with the location of the record. Use **subplotID** and **clipCellNumber** to find the easting and northing offsets in the clipCellNumber_lookup.csv (AD[14]). Add the selected offsets from AD[14] to the values for easting and northing associated with **pointID** that were found in step a.

c. Convert the adjusted UTM coordinates from step b, i.e., the coordinates associated with the actual location of sampling for a given record, to WGS84 coordinates with decimal latitude and longitude units, and fill in the following fields of hbp_perbout_pub:

  - **geodeticDatum**
  - **utmZone (string including zone number and “N” or “S”;** the latter indicate the hemisphere in which sampling occurred**)**
  - **easting**
  - **northing**
  - **decimalLatitude**
  - **decimatLongitude**
  - **elevation**
  - **coordinateUncertainty**: Add 0.6 meters to the value for coordinateUncertainty looked up from AD[13] in step (a).
  - **elevationUncertainty**: Add 10 meters to the value for elevationUncertainty looked up from AD[13] in step (a)

4. Check for duplicate records:

   a. Create quality flag field in hbp_perbout_pub, and populate with zeroes:

      i. QF Name = **duplicateClipQF**

   b. List of fieldNames = [**clipID**, **eventID**]

   c. Check for duplicate records based on exact matches of fields listed in fieldNames.

   d. If duplicate records are identified:

      i. If values in all remaining fields, *except* **uid** and **enteredBy** fields, are identical:

         A. Pass one of the records into hbp_perbout_pub

         B. Enter a ‘1’ into the corresponding **duplicateClipQF** field

   ii. Else:

      A. Pass duplicate records into hbp_perbout_pub

      B. Enter a '2' into the corresponding **duplicateClipQF** field

5. Determine whether unique values of **clipID** occur within more than one **eventID** within t-12 months of current **date**.

   a. Create quality flag in hbp_perbout_pub, and populate with zeros:

      i. QF Name = **multipleClipQF**

   b. For all records occurring within a 12 month period:

      i. If any **clipID** value occurs > 1 time, insert 1 into the **multipleClipQF** field of each record with that **clipID** value in hbp_perbout_pub

6. For each complete group of related records submitted with the same **boutNumber,** perform a quality check on the **exclosure** data to ensure that the values of **exclosure** conform to what is expected from the protocol: If a given site is managed for grazing, for some plots we expect there will be two values of **clipID** for each unique combination of **plotID** and **subplotID**, and we also expect that **exclosure** values for these two **clipIDs** will be different – i.e., one with exclosure='Y' and the other with exclosure='N'.

   a. Generate **exclosureQF** and populate with zeroes.
   b. If there are two records with: 1) the same **plotID** and **subplotID;** 2) different **clipID;** and 3) the same value of **exclosure**, assign **exclosureQF** = 1 to both records, and change the value of **exclosure** to 'uncertain' for both records.

5.2 Processing steps for data in hbp_massdata_in

1. Across a complete group of related records with the same **boutNumber**, for each record find the required 'dateTime' information for the **date** field by searching the hbp_perbout_pub table and matching on **clipID.** Assign 'dateTime' data hbp_perbout_pub to the **date** field in hbp_massdata_pub

   a. If no match exists in hbp_perbout_pub, remove record from hbp_massdata_in

2. Populate the **eventID, exclosure,** and **exclosureQF** fields for each record in hbp_massdata_pub by searching hbp_perbout_pub and matching on the value of **clipID**.

   - If more than one record from hbp_perbout_pub matches **clipID** (i.e., there is a duplicate in hbp_perbout_pub), choose one of the records to populate the required fields.

3. Populate **plotType** in hbp_massdata_pub, based on the TOS Plot Spatial Data lookup table (AD[13])
    a. For each record in hbp_massdata_in, find the corresponding record in AD[13] based on matching values in the **plotID** field of the two tables, and where AD[13]:subType = 'basePlot'
        i. Assign AD[13]:**plotType** to corresponding field in hbp_massdata_pub
4. Calculate **storageHours**, the amount of time samples were kept in cold storage before being placed in the drying oven:
    a. storageHours_temp = **ovenInDate** – **date** (must include date and time data)
    b. If any of the following are true, then storageHours_temp = 'NULL':
        i. **ovenInDate** = 'NULL', or
        ii. **date** = 'NULL', or
        iii. **ovenInDate** – **date** is negative
    c. Round storageHours_temp to the nearest hour
    d. Assign storageHours_temp to **storageHours** field in hbp_massdata_pub
5. Calculate **dryingHours,** the amount of time samples were oven-dried:
    a. dryingHours_temp = **ovenOutDate** – **ovenInDate** (must include date and time data)
    b. If either **ovenOutDate** or **ovenInDate** are null: dryingHours_temp=NULL
    c. Else: Round dryingHours_temp to the nearest hour
    d. Assign dryingHours_temp to **dryingHours** field in hbp_massdata_pub
6. Check for duplicate records:
    a. Create quality flag field in hbp_massdata_pub, and populate with zeroes:
        i. QF Name = **duplicateMassQF**
    b. List of fieldNames = [**sampleID**, **qaDryMass**]
    c. Check for duplicate records based on exact matches of fields listed in fieldNames.
    d. If duplicate records are identified:
        i. If values in all remaining fields, *except* **uid** and **enteredBy** fields, are identical:
            A. Pass one of the records into hbp_massdata_pub
            B. Enter a '1' into the corresponding **duplicateMassQF** field
        ii. Else:
            A. Pass duplicate records into hbp_massdata_pub

B. Enter a '2' into the corresponding **duplicateMassQF** field of both records

7. Check for missing records in hbp_massdata_pub:

   a. *Identification:* The protocol instructs that for each clip that occurs in the field (identified uniquely by combinations of **eventID** and **clipID** in the perbout table), if the technicians actually collect a sample (targetTaxaPresent='Y'), this collection should result in primary weights (**qaDryMass**='N') of either: 1) A single mass record in the massdata table with herbGroup='ALL' (i.e., the sample is not sorted), or 2) 6 mass records (with values of herbGroup = {BRY, CSG, WSG, LFB, FRB, WST}. In evaluating missing records, mass records with **qaDryMass=**'Y' can be ignored since the protocol does not prescribe exactly which samples are reweighed for quality control purposes.

   b. *How to apply flags:*
      - Create new records for each missing record.
      - Populate **missingRecordsMassQF** flag set to 1, and populate any known ancillary information (e.g. **subplotID, plotType, clipID, eventID, exclosure**).
      - Add text 'This record is generated by the system' to each newly created row.
      - Set **missingRecordsMassQF** to -1 for all records with **qaDryMass**='Y', since this test is irrelevant to the repeat weights.
      - For existing records (not those created for missing records), set the **missingRecordsMassQF** to 0.

6 UNCERTAINTY

Analysis of Meaurement Uncertainty is included as part of the "Future Modifications and Plans" in Section 9.

7 VALIDATION AND VERIFICATION

7.1 Algorithm Validation

Inputs not available for initial version of this document.

7.2 Data Product Validation

Inputs not available for initial version of this document.

7.3 Data Product Verification

Inputs not available for initial version of this document.

8 SCIENTIFIC AND EDUCATIONAL APPLICATIONS

Production of herbaceous biomass can be a significant component of aboveground ANPP, particularly in grassland sites and sites that are managed for grazing. As such, the herbaceous clip harvest data product is an important part of NEON's effort to characterize ANPP, and provides insight into how biomass production associated with this important plant group changes through time. Additionally, estimates of herbaceous aboveground ANPP made in Tower plots within the NEON Tower footprint allow an assessment of how this vegetation group contributes to net ecosystem productivity. Finally, because herbaceous biomass is sorted to functional group at least once per growing season, the herbaceous clip harvest data product enables detection of functional changes in herbaceous plant community composition through time.

9 FUTURE MODIFICATIONS AND PLANS

- Currently, records in hbp_massdata_in that have no match in hbp_perbout_pub based on **clipID** and **date** are discarded. In future, it is strongly desired to constrain data entry in hbp_massdata_in such that **clipID** and **date** entries are checked in real time against those values of hbp_perbout_pub that already exist in the NEON database.
- Create a **missingRecordsPerBoutQF** once a stable list of plot-level spatial information is compiled, using the approach below:
    1. Check for missing records in hbp_perbout_pub:
        a. Create quality flag field in hbp_perbout_pub, and populate with zeroes:
            i. QF Name = **missingRecordsPerBoutQF**
        b. For each **eventID**, determine minimum '# expected records':
            i. If **plotType** = 'Distributed': Insert -1 into **missingRecordsPerBoutQF**
                1. Note: This test cannot be performed for these plots, as sampling effort is variable per site
            ii. Else if **plotType** = 'Tower' AND **plotSize** = 400 AND the unique values for **exclosure** = ['N']:
                A. '# expected records' = 30

        iii. Else if **plotType** = 'Tower' AND **plotSize** = 400 AND the unique values for **exclosure** = ['N','Y']:
            A. '# expected records' = 60
        iv. Else if **plotType** = 'Tower' AND **plotSize** = 1600 AND the unique values for **exclosure** = ['N']:
            i. '# expected records' = 40
        v. Else if **plotType** = 'Tower' AND **plotSize** = 1600 AND the unique values for **exclosure** = ['N', 'Y']:
            1. '# expected records' = 80
    c. Calculate # missing records:
        i. missing_temp = '# expected records' – # observed records per **eventID**
    d. If missing_temp > 0: Create new rows in hbp_perbout_pub, one for each missing record.
        i. Assign **uid** and **eventID** to the new rows
        ii. Assign **domainID** and **siteID** location information to the new rows
        iii. Based on the known list of Tower plots for the **siteID** in which the bout was performed, match the list of plotIDs for which records exist for the bout against the list of all known Tower plots that should have been sampled at that site for the bout, and determine which plots are missing. Assign the missing plotIDs to the **plotID** field of the new rows.
        iv. Assign **date** for the N inserted rows based on the median value of **date** for all records with a matching **eventID** AND with **missingRecordsPerBoutQF** = 0
        v. Assign **missingRecordsPerBoutQF** = 1 for the new rows
        vi. Add text 'This record is generated by the sytem' to the **remarks** field of the inserted rows.

- Addition of QA/QC checks for a new data ingest table that will be designed to accommodate herbaceous clip harvest data from sites with continuous growth and decomposition of herbaceous biomass.
- Updates and modifications to address problems with QA/QC tests that arise once validation and verification datasets are available.
- Calculation of "Measurement Uncertainty" (see below)

Measurement Uncertainty

1. Create a "hbp_uncertaintyreport" table to hold uncertainty results calculated in the steps below.
   a. List of fields = [eventID, percentHerbaceousMassQF, meanQAMassDiff]
2. On a per **eventID** basis, calculate the percent of rows in hbp_massdata_db for which QA herbaceous mass values differ from the original mass value by more than 10%.
   a. For the *ith* unique **eventID** in hbp_massdata_db:
      i. rowNum = total number of rows with qaHerbaceousMass != 0 or *null*
      ii. countMassQF = total number of rows with valueQAHerbaceousMassQF = 1
      iii. percentHerbaceousMassQF = (countMassQF/rowNum) * 100
      iv. Insert percentHerbaceousMassQF into the cell of the **percentHerbaceousMassQF** field of hbp_uncertaintyreport associated with the *ith* value of **eventID.**
3. On a per **eventID** basis, calculate the mean difference between herbaceousMass and qaHerbaceousMass.
   a. For the *ith* unique **eventID** in hbp_massdata_db:
      i. massDiff = NA
      ii. qaRows = subset of rows in *ith* **eventID** with qaHerbaceousMass != 0 or *null*
      For *j* in 1 to number of rows in qaRows:
      1. massDiff[j] = abs(qaRows$herbaceousMass[j] – qaRows$qaHerbaceousMass[j])
      iii. Insert mean(massDiff) into the cell of the **meanQAMassDiff** field of hbp_uncertaintyreport associated with the *ith* value of **eventID**

10 BIBLIOGRAPHY

Biondini, M. E., W. K. Lauenroth, and O. E. Sala. 1991. Correcting estimates of net primary production: Are we overestimating plant-production in rangelands? Journal of Range Management **44**:194-198.

Collatz, G. J., J. A. Berry, and J. S. Clark. 1998. Effects of climate and atmospheric CO2 partial pressure on the global distribution of C4 grasses: present, past, and future. Oecologia **114**:441-454.

Knapp, A. K., J. M. Briggs, D. L. Childers, and O. E. Sala. 2007. Estimating aboveground net primary production in grassland- and herbaceous-dominated ecosystems. Pages 27-48 *in* T. Fahey and A. K. Knapp, editors. Principles and standards for measuring primary production. Oxford University Press, New York.

McNaughton, S. J., D. G. Milchunas, and D. A. Frank. 1996. How can net primary productivity be measured in grazing ecosystems? Ecology **77**:974-977.

Sala, O. E., and A. T. Austin. 2000. Methods of estimating aboveground net primary productivity. Pages 31-43 *in* O. E. Sala, R. B. Jackson, H. A. Mooney, and R. W. Howarth, editors. Methods in Ecosystem Science. Springer-Verlag, New York, NY.

Sala, O. E., M. E. Biondini, and W. K. Lauenroth. 1988. Bias in estimates of primary production: An analytical solution. Ecological Modelling **44**:43-55.

11 CHANGELOG

12 APPENDIX: DATA ENTRY VALIDATION VIA PDA AND WEBUI

12.1 All Data: Run the following steps for all fields in in the Raw Data Ingest Workbook for TOS Herbaceous Plant Biomass (AD[10])

1. Constrain entered values to the correct **dataType**
2. Constrain entered values to conditions specified in **entryValidationRules** field of AD[10]
3. Generate a unique ID for each record (**uid**)
4. Follow guidelines for fields in which no data have been entered, as specified in **noDataOutcomePDA** and **noDataOutcomeUI:**
    a. If **noDataOutcome** = 'fail'
        i. Do not let user finalize record until a value is provided
        ii. Display warning message text, "Please enter a value for [fieldName] to continue," unless alternative text is provided in **warningText**
    b. Else if **noDataOutcome** = 'warn'
        i. Warn user that a value is missing prior to finalizing the record, but allow selection of 'OK' to continue without a value.
        ii. warning message text, "Please confirm that there is no value for [fieldName] to continue," unless alternative text is provided in **warningText**
    c. Else **noDataOutcome** = 'pass'
        i. Allow user to finalize record with no value in this field
5. Follow guidelines for default values in AD[10], as specified in **defaultValuePDA**, **defaultValueUI**, and **case** fields
6. Anonymize all technicianIDs (i.e., fields ending in 'By'; e.g., recordedBy, identifiedBy)

12.2 Validation steps for hbp_perbout_in

1. Provide filtered drop-down menus for location fields:
    a. **domainID** and **siteID**, using the TOS plot-level spatial data lookup table (AD[13])
    b. **plotID**, using the TOS plot-level spatial data lookup table (AD[13]), where AD[13]:**subType** = 'basePlot'
    c. **subplotID**, using the TOS plot-level spatial data lookup table (AD[13]):
        i. If corresponding value of AD[13]:**plotSize** = 400: **subplotID** = 31
        ii. Else: **subplotID** = 21|23|39|41
2. Auto-populate **date** according to provided default value guidelines
    a. Note for UI: for the **date** field, technician enters 'date' (YYYY-MM-DD) and 'time' (HH:mm) into separate fields, and values are concatenated to form correctly formatted value for **date** field.
3. Auto-populate **recordedBy**, **measuredBy**, **boutNumber**, **subplotID**, and **samplingProtocol** with values entered previously, if available
4. Provide drop-down menu for: **exclosure** and **targetTaxaPresent**

5. Constrain format of **boutNumber** field:
    a. If value entered by technician is > 0 and < 10: Append leading zero to value in **boutNumber** field, and accept record.
    b. Else if value entered by technician is ≥ 10 and < 100: Accept value in **boutNumber** field entered by technician.
    c. Else: Display warning text, "Value is out of acceptable range," and do not allow technician to proceed until an acceptable **boutNumber** value is entered.
6. **recordedBy** and **enteredBy** – Maximo user list for FOPs or 'Other'
    a. PDA solution: just type in (i.e., typeahead field) rather than select other
    b. PDA will send values from **recordedBy** as entered on the PDA to both **recordedBy** and **enteredBy** fields in PDR
    c. UI solution: Employ Typeahead field
        i. **recordedBy** and **enteredBy** fields may have different values
7. Constrain **clipCellNumber** field:
    a. If **subplotID** = 21|31: **clipCellNumber** must be in range(1,250)
    b. Else if **subplotID** = 23: **clipCellNumber** must be in range(251,500)
    c. Else if **subplotID** = 39: **clipCellNumber** must be in range(501,750)
    d. Else if **subplotID** = 41: **clipCellNumber** must be in range(751,1000)
    e. If value entered by technician is ≤ 999: Append leading zero to value in **clipCellNumber**, and accept record.
    f. The **clipCellNumber** field is not passed to hbp_massdata_pub, but is used to create **clipID** below.
8. Auto-populate **clipID** field:
    a. **clipID** = concatenate: **plotID** + '_' + **clipCellNumber**
    b. *Example: CPER_001_0147*

12.3 Validation steps for hbp_massdata_in

For each sampling bout (i.e., unique value of **eventID,** defined in Section 5 above), validation steps in sections 12.1 and 12.2 must be completed for hbp_perbout_in data prior to carrying out the steps below on data in hbp_massdata_in.

1. Provide filtered drop-down menus for location fields:
    a. **domainID** and **siteID**, using the TOS plot-level spatial data lookup table (AD[13])
    b. **plotID**, using the TOS plot-level spatial data lookup table (AD[13]), where AD[13]:**subType** = 'basePlot'
    c. **subplotID**, using the TOS plot-level spatial data lookup table (AD[13]):
        i. If corresponding value of AD[13]:**plotSize** = 400: **subplotID** = 31
        ii. Else: **subplotID** = 21|23|39|41

2. Auto-populate the **date** and **weighDate** fields according to provided default value guidelines.
    - The **date** and **weighDate** fields will have no 'time' portion when entered into the webUI.
3. Auto-populate **ovenInDate** and **ovenOutDate** according to provided default value guidelines.
   Notes for UI:
    a. For the **ovenInDate** and **ovenDateOut** fields, technicians enter 'date' (YYYY-MM-DD) and 'time' (HH:mm) into separate fields, and values are concatenated to form correctly formatted dateTime values for **ovenInDate** and **ovenDateOut** fields.
    b. If the value of the 'time' portion of the input for **ovenInDate** or **ovenOutDate** is 'null', assign to 12:00 (noon) local time.

4. Provide drop-down menu for: **herbGroup**, **qaDryMass**

5. Constrain **clipCellNumber** field:
    a. If **subplotID** = 21|31: **clipCellNumber** must be in range(1,250)
    b. Else if **subplotID** = 23: **clipCellNumber** must be in range(251,500)
    c. Else if **subplotID** = 39: **clipCellNumber** must be in range(501,750)
    d. Else if **subplotID** = 41: **clipCellNumber** must be in range(751,1000)
    e. If value entered by technician is ≤ 999: Append leading zero to value in **clipCellNumber**, and accept record.
    f. The **clipCellNumber** field is not passed to hbp_massdata_pub, but is used to create **clipID** below.
6. Auto-populate **clipID** field:
    a. **clipID** = concatenate: **plotID** + '_' + **clipCellNumber**
    b. *Example: CPER_001_0147*
7. **recordedBy** and **enteredBy** – Maximo user list for FOPs or 'Other'
    a. PDA solution: just type in (i.e., typeahead field) rather than select other
    b. PDA will send values from **recordedBy** as entered on the PDA to both **recordedBy** and **enteredBy** fields in PDR
    c. UI solution: Employ Typeahead field
        i. **recordedBy** and **enteredBy** fields may have different values
8. Auto-generate a value for the **sampleID** field:
    a. Assign a temporary herbCode value for the record using the corresponding **herbGroup** value.
        i. 'Bryophyte' - BRY, 'Cool Season Grasses' - CSG, 'Warm Season Grasses' - WSG, 'Leguminous Forbs' - LFB, 'Non-leguminous Forbs' - FRB, 'Woody-stemmed Plants' - WST, 'ALL' - ALL
    b. clipID = value from corresponding **clipID** field with '_' characters stripped out

   c. date = ‘YYYYMMDD’ from corresponding **date** field with other characters stripped out
   d. **sampleID** = concatenate: ‘NEON.hbp.’ + ‘clipID’ + ‘.’ + ‘date’ + ‘.’ + ‘herbCode’
      i. *Example: NEON.hbp.CPER0010147.20140625.BRY*

9. Additional webUI features for constraining data entry:
   a. *Dealing with **dryMass** and **qaDryMass** data entry:* In the database, ‘QA’ records are indicated via the **qaDryMass** field and have a value of **qaDryMass** = ‘Y’. Standard non-QA records are indicated with **qaDryMass** = ‘N’. Masses in the database are only ever reported in the **dryMass** field. However, for a given value of **herbCode**, the webUI allows technicians to enter mass values in grams into both a **dryMass** and a **qaDryMass** field. In the event that mass data are entered into both the **dryMass** and **qaDryMass** fields, the webUI then creates separate records from these mass entries with the same value of **herbCode** for each record, and assigns the appropriate **qaDrymass** = ‘Y’ or ‘N’ in the database. In this way, the webUI better matches the manner in which techs record data in the paper data sheets, and allows for more efficient data entry. For a given value of **herbCode**, there additional cases that require consideration:
      - Technicians enter a mass value only in the **dryMass** field in the webUI. This happens routinely because only 10% of samples are selected for QA. Because there is no mass entered into the webUI **qaDryMass** field, there is no QA record generated, and only one record with **qaDryMass** = ‘N’ is generated.
      - Technicians enter a mass value only in the **qaDryMass** field in the webUI. Because no mass value is entered in the webUI **dryMass** field, there is no record generated with **qaDryMass** = ‘N’, and only one record with **qaDryMass** = ‘Y’ is generated.
      - Technicians mistakenly do not enter a value into either the **dryMass** or the **qaDryMass** fields of the webUI (this is not what the protocol instructs, but may occur). Because no mass values are entered in either field, the webUI does not generate any records in the database for the given value of herbCode.
      - When a value is entered into the **qaDryMass** field, the webUI provides new fields for **recordedBy**, **measuredBy**, and **weighDate** that are specific to the QA mass entered, because the values entered for these fields for the ‘standard’ mass may not be the same. Values from other fields, such as **ovenInDate** and **ovenOutDate** are identical for the QA and ‘standard’ records, and the webUI copies values from these fields to the QA record. A new **remarks** field is not provided for the QA record, so all remarks for a given sample must be entered in the **remarks** field associated with the ‘standard’ record.