NEON USER GUIDE TO MOSQUITOES SAMPLED FROM CO2 TRAPS (DP1.10043.001) AND MOSQUITO-BORNE PATHOGEN STATUS (DP1.10041.001)

PREPARED BY	ORGANIZATION
Katherine LeVan	SCI
Sara Paull	SCI

CHANGE RECORD

REVISION	DATE	DESCRIPTION OF CHANGE
A	07/19/2017	Initial Release
B	09/30/2019	Adds missed bout reporting; Adds new field called sampling impractical
C	10/14/2020	Included general statement about usage of neonUtilities R package and statement about possible location changes. Updated taxonomy information. Clarified that from 2018-present, one bout of sampling occurs over one night and the following day (24 hours). Added information about transition from on to off season and changes to the list of candidate vector species for pathogen testing
D	04/25/2022	Updated section 5.3 Data Revision with latest information regarding data release. Updated information regarding the geoNEON package.

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 presence of a mosquito sample 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 two L1 data products: Mosquitoes sampled from CO2 traps and Mosquito-borne pathogen status 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 files NEON Data Variables for Mosquitoes sampled from CO2 traps (DP1.10043.001) (AD[05]) and NEON Data Variables for Mosquito-borne pathogen status (DP1.10041.001) (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 the data collected in the field pertaining to TOS Protocol and Procedure: Mosquito Sampling (AD[09]). The raw data that are processed in this document are detailed in the file, NEON Raw Data Validation for Mosquitoes sampled from CO2 traps (DP0.10043.001) (AD[04]), provided in the download package for this data product. Please note that raw data products (denoted by 'DP0') may not always have the same numbers (e.g., '10043') 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 Data Products Catalog
AD[04]	Available with data download	Validation csv
AD[05]	Available with data download	Variables csv
AD[06]	Available with data download	Variables csv
AD[07]	NEON.DOC.000910	TOS Science Design for Mosquito Abundance, Diversity and Phenology
AD[08]	NEON.DOC.000911	TOS Science Design for Vectors and Pathogens
AD[09]	NEON.DOC.014049	TOS Protocol and Procedure: Mosquito Sampling
AD[10]	NEON.DOC.000008	NEON Acronym List
AD[11]	NEON.DOC.000243	NEON Glossary of Terms
AD[12]	NEON.DOC.004285	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
AD[15]	Available with data download	Categorical Codes csv

2.2 Acronyms

Acronym	Definition
CDC	Centers for Disease Control and Prevention

3 DATA PRODUCT DESCRIPTION

Mosquitoes are sampled in the field using CDC CO_2 light traps. Following collection, mosquito samples are sent to an external facility where they are sorted to remove bycatch and taxonomically identified (to species and sex, whenever possible). In the case of large field samples, a subsample of up to 200 individual mosquitoes is taxonomically identified but both total weights of the field collected sample and the subsample are provided to inform estimates of total abundance. Identifications for a subset of difficult taxa are verified by DNA barcoding. For additional details on the sampling design and associated protocol, see the TOS Science Design for Mosquito Abundance, Diversity and Phenology (AD[07]) and TOS Protocol and Procedure: Mosquito Sampling (AD[09]).

Mosquito-borne pathogen sampling involves the testing of all or a subset of collected mosquitoes for infection by viral pathogens by one or more external facilities. Only female mosquitoes identified to the species-level and captured in sufficient quantity over a season from likely vector species are eligible for pathogen testing. A set of up to 1000 individual mosquitoes per species per site per year are targeted for pathogen testing of arboviruses within the families Bunyaviridae, Alphaviridae, and Flavivirdae. From 2013-2019, top-priority mosquito species sent for pathogen testing included: *Aedes aegypti*, *Aedes albopictus* (tested if >100 collected per site-year), and *Culex tarsalis*, *Culex pipiens*, and *Aedes triseriatus* (tested if >200 collected per site-year). From 2013-2019, other individuals, particularly those identified to the species-level within the genera of *Aedes* and *Culex*, were tested if >200 were collected per site-year. Beginning in 2020, a smaller set of vectors that reach the minimum abundance threshold of 200 per site-year will be selected for pathogen testing from the following species list: *Culex tarsalis*, *Culex restuans*, *Culex pipiens*, *Culex quinquefasciatus*, *Culex erraticus*, *Culex nigripalpus*, *Culex salinarius*, *Aedes japonicus*, *Aedes triseriatus*, *Culiseta melanura*, *Culiseta morsitans*, *Psorophora columbiae*, *Coquillitidia perturbans*, *Aedes dorsalis*, *Aedes trivittatus*, *Aedes canadensis mathesoni*. Candidate species for testing were selected in consultation with our technical working group (TWG), and include species that are well-represented as potential vectors in the literature, or that have yielded at least one positive virus-specific test in the NEON data. Species that are vectors of urban-adapted, primarily human pathogens will no longer be pathogen-tested (e.g., *Aedes aegypti*, *Aedes albopictus*) due to the non-urban locations of most NEON sites. We are continually evaluating the sampling design for best practices in collaboration with the community in the form of our technical working groups.

Following identification, mosquitoes are combined by species, sex, and bout (e.g., all female mosquitoes of species A collected at site C during sampling bout D). Groups of individuals combined for testing are assigned the same testingID. This large pool is then subdivided into vials (testingVialID), which contains a defined number of mosquitoes (generally a poolSize of 10-50 individuals). Each testingVialID is tested one or more times using a variety of methods which may include RT-PCR, Vero cell culture, and melt curve assays. These methods vary in target specificity, from general (e.g., Vero cell culture) to specific viral species (e.g., RT-PCR). Most pools of mosquitoes are negative because pathogens are rare; when pools are determined to be positive for any virus, the identit(ies) of the virus(es) are determined to the species-level, if possible. Test results yield data on the presence of important mosquito pathogens (e.g., West Nile virus, Eastern and Western equine encephalitis viruses, California encephalitis virus etc.) in a subset of species that are known vectors of disease. See the TOS Science Design for Vectors and Pathogens (AD[08]) for additional background on mosquito-borne pathogen sampling.

3.1 Spatial Sampling Design

Mosquito sampling is executed at all terrestrial NEON sites and follows a spatially-balanced stratified random design (AD[02]). Mosquitoes are sampled at 10 mosquito points per site. Points are randomly positioned within each National Land Cover Database (NLCD) class with representation within each NLCD class set as proportional to its representation at the site; NLCD classes with less than 5% representation are excluded from sampling. For ease of deployment, mosquito plots are always located between 5-45 meters distance from a road accessible to sampling by NEON technicians. Mosquito points must be separated by a minimum of 310m and must be 10m from the edge of other NEON sampling locations.

As much as possible, sampling occurs in the same locations over the lifetime of the Observatory. However, over time some sampling locations may become impossible to sample, due to disturbance or other local changes. When this occurs, the location and its location ID are retired. A location may also shift to slightly different coordinates. Refer to the locations endpoint of the NEON API for details about locations that have been moved or retired: https://data.neonscience.org/data-api/endpoints/locations/

3.2 Temporal Sampling Design

When adult mosquitoes are active, sampling bouts will occur every two weeks at core sites and every four weeks at relocatable sites. The finest temporal resolution at which mosquito data (for the purposes of species richness, abundance and phenology) will be tracked is trapping night or trapping day. The finest level of temporal resolution at which mosquito-borne pathogen status will be tracked is at the level of a sampling bout. A sampling bout is currently (2018-present) comprised of two separate samples per trap (20 samples per site) consisting of one trapping night and the following day for up to ten plots. From 2013-2017 a bout consisted of two trapping nights and the intervening day for up to ten plots (30 samples per site). The setDate (indicating when the trap was set) and collectDate (indicating when the trap was collected) will be recorded for each sample collected during a bout. Bouts are grouped using the **eventID** designation (a descriptor that includes the year of sampling, the site ID, and the calendar week in which a sampling bout occurred). Infrequently, a bout may be scheduled over 2 ISOweeks such that a collection bout will span 2 **eventID**s.

The total number of bouts per year varies among sites based on seasonality of each site (e.g., stopping during winter at temperate sites). During the time of year when mosquitoes are flying, sampling bouts occur every 2 weeks at the core site and every 4 weeks at each relocatable site, alternating between the core and a relocatable such that one site in the domain is sampled each week. After the mosquito season has ended (e.g., upon the onset of winter), weekly sampling at three plots at the core site will monitor for mosquito presence and help determine when the next mosquito sampling season should begin. From 2013-2019 the transition from on-season sampling to off-season monitoring required three zero-catch bouts at the core site, but the number of traps varied from 3-10 during those bouts, and the time interval between sampling varied from 1 to 2 weeks. Beginning in 2020, transition to off-season monitoring requires 3 consecutive on-season bouts (10 traps) conducted 2 weeks apart in which no mosquitoes are captured. When temperatures are too low for sampling to occur this can be treated as a zero-capture bout. Additional details about sampling bout frequency can be found in the TOS Protocol and Procedure: Mosquito Sampling (AD[09]).

3.3 Sampling Design Changes

There have been several design changes that have been implemented over the course of data collection. Such changes arise due to continual evaluation of the sampling design for best practices in collaboration with technical working groups. They also occur when optimization of the design is necessary to ensure that allocation of sampling effort is poised to maximize returns to the scientific community.

Beginning in the 2018 field season, the length of a sampling bout was changed from 2 nights plus the intervening day to 1 night and the following day. Analyses indicated that dropping from 2 nights to one would not negatively impact the diversity data.

Beginning in the 2019 field season, the field **samplingImpractical** was added to the data to allow for the generation of a record when a plot or site could not be sampled for a particular night or day of sampling. If field sampling was not possible **samplingImpractical** is populated with a value other than 'OK' (e.g., 'location flooded') and no data are recorded other than an eventID. From sampling season 2019 and onwards, there will always be 20 sampling records per bout of CDC CO_2 trapping. If sampling could not be conducted for all or part of the bout, the samplingImpractical field will communicate such missing records and the reason therefore.

Beginning in the 2020 field season, transition to off-season monitoring requires 3 consecutive on-season bouts (10 traps) conducted 2 weeks apart in which no mosquitoes are captured. If sampling must be canceled due to low temperatures it counts as a zero-capture bout. Prior to 2020, three consecutive zero-capture bouts were always required, but the time interval between bouts and number of traps set varied.

Beginning in the 2020 field season, the list of vector species sent for pathogen testing was revised and shortened due to low infection prevalence. The revised list of candidate vectors for pathogen testing was developed following analysis of NEON mosquito pathogen data, literature review and consultation with the technical working group. The following vectors will be sent for pathogen testing when they reach the minimum abundance threshold of 200 per site-year: *Culex tarsalis*, *Culex restuans*, *Culex pipiens*, *Culex quinquefasciatus*, *Culex erraticus*, *Culex nigripalpus*, *Culex salinarius*, *Aedes japonicus*, *Aedes triseriatus*, *Culiseta melanura*, *Culiseta morsitans*, *Psorophora columbiae*, *Coquillitidia perturbans*, *Aedes dorsalis*, *Aedes trivittatus*, *Aedes canadensis mathesoni*. Species that are vectors of urban-adapted, primarily human pathogens will no longer be pathogen-tested (e.g., *Aedes aegypti*, *Aedes albopictus*) due to the non-urban locations of most NEON sites.

3.4 Variables Reported

All variables reported from the field or laboratory technician (L0 data) are listed in the file, NEON Raw Data Validation for Mosquitoes sampled from CO2 traps (DP0.10043.001) (AD[04]). All variables reported in the published data (L1 data) are also provided separately in the files, NEON Data Variables for Mosquitoes sampled from CO2 traps (DP1.10043.001) (AD[05]) and NEON Data Variables for Mosquito-borne pathogen status (DP1.10041.001) (AD[06]).

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), 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 Temporal Resolution and Extent

The finest resolution at which temporal data are reported is the **trapHours**, the range between **setDate** and **collectDate**.

collectDate (date an individual trap was collected) → **trapHours**

3.6 Spatial Resolution and Extent

The finest resolution at which spatial data are reported is a single trap (**??**).

plotID (unique ID given to the individual trap) → **siteID** (ID of NEON site) → **domainID** (ID of a NEON domain).

The basic spatial data included in the data downloaded include the latitude, longitude, and elevation of the plot marker where trapping occurred + associated uncertainty due to GPS error (trapping data) or the latitude, longitude and elevation of the NEON tower at each site (pathogenresults), since mosquitoes are typically pooled across plots within a site. Shapefiles of all NEON Terrestrial Observation System sampling locations can be found here: http://www.neonscience.org/science-design/field-sites/maps-spatial-data.

To derive a more precise estimate of the location of each trap, there are two options:

- Use the getLocTOS function from the geoNEON package, available here: https://github.com/NEONScience/NEON-geolocation
- Or follow these steps to perform the same calculation:

trapping:

1. Precise geocoordinates for the plot marker and associated coordinate uncertainly are provided in the downloaded data
2. Technicians are permitted to move up to 10m from the marked location to find a suitable place to install traps. Thus realized coordinate uncertainty on trap placement = **coordinateUncertainty** + 10m

pathogenresults:

1. Precise geocoordinates for the plot marker and associated coordinate uncertainly are provided in the corresponding trapping data records
2. Weighted averaging of trap locations for mosquitoes contributing to a given pathogenpooling **testingID**

3.7 Associated Data Streams

testingID is the linking variable that ties specific samples and associated metadata between the Mosquitoes sampled from CO2 traps data product (DP1.10043.001) and Mosquito-borne pathogen status data product (DP1.10041.001).

3.8 Product Instances

There are a maximum of 26 field season collection bouts per year, with mosquitoes collected from no more than 10 plots per bout. Each plot will yield no more than 3 samples per bout of collection (collection years 2013 - 2017) or 2 samples per bout of collection (collection years after 2018). Thus, no single site should ever exceed 780 trapping data product instances in a given calendar year. The number of records for identification and pathogenresults varies with the diversity of the site and pathogen testing workflows.

3.9 Data Relationships

For collections that occurred before 2018, the protocol dictates that 3 samples are recovered (if present) from each trap per bout (roughly 30 expected records per eventID in mos_trapping). In 2018 and subsequent collection years, sampling was reduced from 40 hours of collection per plot (2 nights and intervening day; up to 3 samples per plot per bout) to 24 hours per plot (one day and one night; up to 2 samples per plot per bout). A record from mos_trapping may have zero or more child records in mos_sorting, depending on whether the trap contained mosquitoes (**targetTaxaPresent** = 'Y') downloaded. A record from from mos_sorting may have one or more child records in mos_identification depending on the number of species detected within the sample, if any. A record from from mos_identification may have zero or one child records in mos_archivepooling (if any material is archived) and zero or more child records in mos_pathogenpooling (if any material is pathogen tested). A record from from mos_pathogenpooling may have one or more child records in mos_pathogenresults depending on the number of tests run on a given testingVialID. Duplicates and/or missing data may exist where protocol and/or data entry aberrations have occurred; users should check data carefully for anomalies before joining tables.

mos_trapping.csv - > One record expected per sampleID for all time; 20 records per eventID (2018 and subsequent collection years) or 30 records per eventID (collection years 2013 - 2017)

mos_sorting.csv - > One record expected per sampleID for all time, generates a single subsampleID if mosquitoes and/or bycatch is present in the trap

mos_expertTaxonomistIDProcessed.csv - > One record expected per subsampleID per scientificName per sex per identificationQualifier combination. The value in **individualCount** per subsampleID per scientificName per sex per identificationQualifier represents the total number of individuals of that type found within the subsample. Up to 10 individuals of a given scientificName/sex/identificationQualifier combination may be removed from a given subsampleID and pinned; the individualID of each pinned individual is given in a pipe-delimited list in the column labelled **individualIDList**. If the mosquitoes of a given scientificName/sex/identificationQualifier combination are pooled from a subsampleID into a pool of mosquitoes for archiving or a pool of mosquitoes for testing, then the archiveID or testingID that the subsampleID contributed to will be listed. Taxonomic identifications in this table have been desynomized using the NEON mosquito taxonomy table.

mos_expertTaxonomistIDRaw.csv - > This table is identical to "mos_expertTaxonomistIDProcessed", except that no taxonomic desynomizing is performed.

mos_identificationHistory.csv - > Zero, two or more records expected per identifier (records may pertain to subsampleID, individualID, testingID, archiveVialIDList); records are only created when data corrections to taxonomy are made. If errors in identification are detected through QAQC processes *after* data publication, then corrected taxonomy will be provided in the mos_expertTaxonomistIDRaw and mos_expertTaxonomistIDProcessed tables. The mos_identificationHistory table is populated with the corrected name and all prior names used for specimen(s) in the data product. When data are populated in the mos_identificationHistory table, **identificationHistoryID** is used as a linking variable between the mos_identificationHistory table and all other mosquito tables where updates were made.

mos_archivepooling.csv - > One record expected per archiveID, which is a mixture of subsampleIDs (listed in archiveID column of the mos_indentification file). Not all subsampleID's contribute to mixtures; some are tested or pinned.

mos_pathogenpooling.csv - > One record expected per testingVialID, which is a mixture of subsampleIDs (listed in testingID column of the mos_indentification file). Not all subsampleID's contribute to mixtures; some are archived or pinned.

mos_pathogenresults.csv - > One or more records expected per testingVialID, which is a subsample from a given testingID (number of individuals within that subsample listed in the 'poolSize' column of the mos_pathogenpooling file).

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

4 TAXONOMY

NEON manages taxonomic entries by maintaining a master taxonomy list based on the community standard, if one exists. Through the master taxonomy list, synonyms submitted in the data are converted to

the appropriate name in use by the standard. The master taxonomy list also indicates the expected geographic distribution for each species by NEON domain and whether it is known to be introduced or native in that part of the range. Errors are generated if a species is reported at a location outside of its known range. If the record proves to be a reliable report, the master taxonomy table is updated to reflect the distribution change.

The full master taxonomy lists are available on the NEON Data Portal for browsing and download: http://data.neonscience.org/static/taxon.html.

4.1 Mosquito Taxonomy

The master taxonomy for mosquitoes is Darsie and Ward (2005), with some modifications based on the Integrated Taxonomic Information System (ITIS) on-line database (http://www.itis.gov). Taxon ID codes used to identify taxonomic concepts in the NEON master taxonomy list were generated for each taxon by concatenating the first three letters of the genus name together with the first three letters of the specific epithet to make a unique taxon ID for each scientific name. Where such concatenation would produce duplicate taxon ID codes, numbers were appended to the taxon ID until it was unique within the NEON database (e.g.,*Aedes cantator* as AEDCAN1 and *Aedes canadensis* as AEDCAN2). The master taxonomy list includes all mosquito species from the continental United States, supplemented with species that are expected to occur at NEON sites in Alaska, Puerto Rico, and Hawaii. NEON plans to keep the taxonomy updated in accordance with the current literature, starting in 2020 and annually thereafter. Geographic ranges and native statuses used in this data product are primarily from Darsie and Ward (2005), future ranges and nativity statuses will be derived from the ITIS on-line database and the current literature.

4.2 Mosquito Pathogen Taxonomy

The master taxonomy for mosquito pathogens is The Universal Virus Database of the International Committee on taxonomy of Viruses (https://talk.ictvonline.org/taxonomy). Taxon ID codes used to identify taxonomic concepts in the NEON master taxonomy list are 3-5 character alphanumeric codes derived from the virus name. The list includes many of the virus species found in mosquitoes of North America, Puerto Rico, and Hawaii that can infect humans. Because mosquitoes are tested first at the family level, the list also includes family-level and genus-level virus identities. NEON plans to review the list of pathogens that are tested starting in 2020 and on an annual basis in consultation with external labs performing the tests. Geographic ranges and native statuses used in this data product are from Centers for Disease Control and Prevention, Arbovirus Catalogue (https://wwwn.cdc.gov/arbocat/VirusDetails.aspx), and the U.S. Geological Survey (USGS) ArboNET arboviral disease maps (http://diseasemaps.usgs.gov/mapviewer/).

5 DATA QUALITY

5.1 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 Mosquitoes sampled from CO2 traps (DP0.10043.001) (AD[04]), 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. Also included in this file is a field named 'entryValidationRulesParser', which describes syntactically the validation rules for each field that is performed upon ingest of the data into the NEON Cyberinfrastructure, based on a standardized data validation language (Nicl) internal to NEON. Please see AD[13] for more information about the Nicl language.

Data collected prior to 2017 were processed using a paper-based workflow that did not implement the full suite of quality control features associated with the interactive digital workflow.

5.2 Automated Data Processing Steps

Following data entry into a mobile application of 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]).

5.3 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.

5.4 Quality Flagging

The **samplingImpractical** field in each data record (table: mos_trapping) is a quality flag that communicates the reason for missed sampling events. From sampling season 2019 and onwards, there will always be 20 sampling records per bout of CDC CO_2 trapping. If sampling could not be conducted for all or part of the bout, the samplingImpractical field will communicate such missing records and the reason therefore.

Records of land management activities, disturbances, and other incidents of ecological note that may have a potential impact are found in the Site Management and Event Reporting data product (DP1.10111.001)

Table 1: Descriptions of the samplingImpractical codes for quality flagging

Value	Description
OK	Sampling occurred at the intended plot location and time
Location snow covered	Sampling did not occur at the intended plot location and time due to the presence of heavy snow cover in excess of 1.8 m at the plot or impeded access to the plot due to snow cover
Location flooded	Sampling did not occur at the intended plot location and time due to flooding at the plot or impeded access to the plot due to flooding
Temperature low	Temperatures at the site were below the threshold for off season sampling, thus off season sampling was not conducted; note that low temperatures are never a valid reason to cancel a field season collection bout
Logistical	Logistical reasons such as site access or staffing difficulties prevented sampling at the scheduled time for that plot
Management	Management activities such as controlled burn, grazing, managed hunts, etc prevented personnel from accessing the site location; see the Site Management and Event Reporting product (DP1.10111.001) for additional details
Extreme weather	Extreme weather such as tornado, hurricane, wildfire, etc present at the site prevented staff from accessing the plot on the scheduled interval
Other	Other activities prevented sampling on schedule at this plot location; these reasons are described in the remarks

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.

5.5 Analytical Facility Data Quality

A quality field (**weightBelowDetection** in table: mos_sorting) in each data record communicates when generated weights are below detection limits for the laboratory scale used.

The laboratory contract for mosquito identification specifies that a subsample of approximately 200 mosquitoes are identified to species-level from each sample; laboratories populate the mos_sorting table with the weight in grams of the whole sample (**totalWeight**), the weight of the subsample that receives taxonomic identification services (**subsampleWeight**) and the weight of any non-mosquito bycatch remaining from the subsample after identification of all mosquitoes from the subsample is complete (**bycatchWeight**). Thus, **subsampleWeight** will be less than or equal to **totalWeight** and **bycatchWeight** will be less than or equal to **subsampleWeight**. If the lab measures a mosquito sample that is below the minimum detection limit of scale used, they populate a flag in the **weightBelowDetection** field to indicate which weight values are below the scale's detection limit.

Because samples can vary in the amount of icing present, users are discouraged from using weights to assess properties of individual mosquitoes in the sample. Two samples with identical mosquito composition and quantity could differ substantially in weight if one has a lot of ice locked into the sample and another is moisture-free. The purpose of these weights is to provide a quantitative scaling mechanism where the abundances of taxa from subsamples that are completely identified (e.g. **subsampleWeight** equals **totalWeight**) versus subsamples where only a portion of the mosquitoes are identified (e.g. **subsampleWeight** less than **totalWeight**)

Table 2: Descriptions of the weightBelowDetection codes for quality flagging

Value	Description
OK	no weights below scale detection limit
total	total weight below scale detection limit
subsample	subsample weight below scale detection limit
bycatch	bycatch weight below scale detection limit
bycatch and subsample	bycatch and subsample weights below scale detection limit
all	all weights below scale detection limit

Laboratories also double check a subset of their identifications to confirm accuracy and detection data transcription errors. Three percent of all samples are quality checked for taxonomic difference by reprocessing at the external facility; laboratories calculate percent difference in enumeration (PDE) and Percent Taxonomic Disagreement (PTD) (Stribling et al. 2008). Percent difference in enumeration (**PDE** in table mos_sorting) must not exceed 5%; PTD must not exceed 2% at the genus level and 5% at the species level (**genusPTD** and **speciesPTD** in the table mos_sorting). Where values exceed these thresholds, discrepancies are reconciled in the final datasheets. Notes on subsamples where QC was performed and results can be found in the 'remarks' column of the applicable tables.Details on the calculations of these fields can be found in the external lab SOP.

If errors in identification are detected during initial taxonomic analysis, only the corrected identification will be provided in the mos_expertTaxonomistIDRaw and mos_expertTaxonomistIDProcessed tables. **PDE**, **genusPTD**, and **speciesPTD** will be populated as described above. If errors in identifica-

tion are detected through QAQC processes *after* data publication, then corrected taxonomy will be provided in the mos_expertTaxonomistIDRaw and mos_expertTaxonomistIDProcessed tables. The mos_identificationHistory table is populated with the corrected name and all prior names used for specimen(s) in the data product. When data are populated in the mos_identificationHistory table, **identificationHistoryID** is used as a linking variable between the mos_identificationHistory table and all other mosquito tables where updates were made.

6 REFERENCES

Darsie, R. F., and R. A. Ward. 2005. Identification and geographical distribution of the mosquitos of North America, north of Mexico. University Press of Florida, Gainesville

Stribling, J. B., K. L. Pavlik, S. M. Holdsworth, and E. W. Leppo. 2008. Data quality, performance, and uncertainty in taxonomic identification for biological assessments. Journal of the North American Benthological Society. 27: 906-919.