NEON USER GUIDE TO BREEDING LANDBIRD POINT COUNTS (DP1.10003.001)

PREPARED BY	ORGANIZATION
Kate Thibault	FSU
Sarah Elmendorf	DPS

CHANGE RECORD

REVISION	DATE	DESCRIPTION OF CHANGE
A	06/29/2017	Initial Release
B	05/25/2020	Included general statement about usage of neonUtilities R package and statement about possible location changes. Updated taxonomy information.
C	04/11/2022	Added language in section 4 Taxonomy addressing RTE species obfuscation in the data. Updated section 5.3 Data Revision with latest information regarding data release. Updated information regarding the geoNEON package.
C.1	12/07/2023	Added identification history section 4.2 to Taxonomy and updated Data Relationships section (3.7) and Analytical Facility Data Quality section (5.5) to reference identification history. Converted alphanumeric pointID values to all-numeric values.
D	04/09/2024	Minor formatting updates
E	02/18/2025	Updated the url for spatial data in section 3.4. Added information about the new neonUtilities Python 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 dry weights of litter functional groups from a single collection event, are considered the lowest level (Level 0). Raw data that have been quality checked via the steps detailed herein, as well as simple metrics that emerge from the raw data are considered Level 1 data products.

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

1.2 Scope

This document describes the steps needed to generate the L1 data product Breeding landbird point counts - counts, distance from observer, and taxonomic identification of breeding landbirds observed during point counts - and associated metadata from input data. This document also provides details relevant to the publication of the data products via the NEON data portal, with additional detail available in the file, NEON Data Variables for Breeding landbird point counts (DP1.10003.001) (AD[05]), 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: Breeding Landbird Abundance and Diversity (AD[07]). The raw data that are processed in this document are detailed in the file, NEON Raw Data Validation for Breeding landbird point counts (DP0.10006.001) (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., '10033') as the corresponding L1 data product.

2 RELATED DOCUMENTS AND ACRONYMS

2.1 Associated Documents

AD[01]	NEON.DOC.000001	NEON Observatory Design (NOD) Requirements
AD[02]	NEON.DOC.000913	TOS Science Design for Spatial Sampling
AD[03]	NEON.DOC.002652	NEON Data Products Catalog
AD[04]	Available with data download	NEON Raw Data Validation for Breeding landbird point counts (DP0.10006.001)
AD[05]	Available with data download	NEON Data Variables for Breeding landbird point counts (DP1.10003.001)
AD[06]	NEON.DOC.000916	TOS Science Design for Breeding Landbird Abundance and Diversity
AD[07]	NEON.DOC.014041	TOS Protocol and Procedure: Breeding Landbird Abundance and Diversity
AD[08]	NEON.DOC.000913	TOS Science Design for Spatial Sampling
AD[09]	NEON.DOC.000913	TOS Science Design for Plant Diversity
AD[10]	NEON.DOC.000008	NEON Acronym List
AD[11]	NEON.DOC.000243	NEON Glossary of Terms
AD[12]	NEON.DOC.004825	NEON Algorithm Theoretical Basis Document: OS Generic Transitions
AD[13]	Available on NEON data portal	NEON Ingest Conversion Language Function Library
AD[14]	Available on NEON data portal	NEON Ingest Conversion Language Function Library
AD[15]	Available with data download	Categorical Codes csv

3 DATA PRODUCT DESCRIPTION

The breeding landbird point counts product provides records of species identification, as well as metadata which can be used to model detectability, e.g. weather, distances from observers to birds, and detection methods. Breeding landbirds are defined as "smaller birds (usually exclusive of raptors and upland game birds) not usually associated with aquatic habitats" (Ralph et al. 1993). They are typically censused during the first half of the breeding season, when birds are "most active, paired, on territories, and vocal" (Ralph et al. 1993).

Breeding landbirds are sampled using the point count method. Point counting entails one or more observers going to pre-established points and recording all the birds heard and/or seen during a set period of time (Figure 1). The NEON point count method is adapted from the Integrated Monitoring in Bird Conservation Regions (IMBCR): Field protocol for spatially-balanced sampling of landbird populations. (Hanni et al. 2016; http://bit.ly/17ekDNB). Point counts are six minutes long, with each minute tracked by the observer, following a two-minute settling-in period. All birds are recorded to species and sex, whenever possible, and the distance to each individual or flock is measured with a laser rangefinder, except in the case of flyovers.

Breeding landbird point counts data may be used to inform estimates of abundance, density and diversity of bird populations at plot, site, and continental scales. They also can be used to inform models of mosquito-borne disease dynamics.

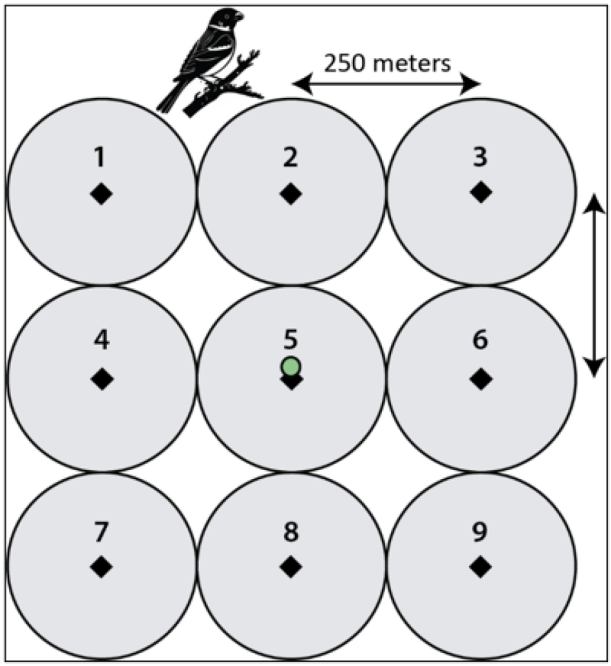

Figure 1: NEON plot (grid of points) for sampling breeding landbirds

3.1 Spatial Sampling Design

Breeding landbird point counts are conducted at terrestrial NEON sites. Depending on the size of the site, sampling for this product occurs either at either points or across grids.

At larger sites, point count sampling occurs across 5-10 9-point grids, with grid centers collocated with distributed plots centers, if possible. The 9-point grids are distributed within the sites proportional to permitted sampling area contained within each National Land Cover Database (NLCD) class, with the restriction that >50% of the points within each grid must fall within the target NLCD class. Grids may be intersected by roads and buildings, but are situated to avoid large bodies of water and to ensure that any point on the grid is at least 10m from any road. Any point on a grid that intersects smaller bodies of water is shifted to dry ground. Bird grids must be separated by a minimum distance of 250m.

At smaller sites (i.e., sites that cannot accommodate a minimum of 5 grids, using the criteria above), point counts occur at the southwest corner (point 21) of 5-25 distributed base plots that are distributed according to a stratified-random design proportional to permitted sampling area contained within each of the National Land Cover Database (NLCD) classes found throughout the entire site. Distributed base plots must be a minimum of 10m from roads and 50m from buildings. Small streams (<1m wide) can bisect plots. If the southwest corner of the plot is in the stream, observations are made from the closest dry location. Distributed base plots selected for bird point count sampling must be separated by a minimum of 250m.

The distance from each observed bird to the location of the observer (pointID) is recorded in the data as part of the protocol for visual detection of birds (observerDistance).

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/

See TOS Science Design for Breeding Landbird Abundance and Diversity (AD[06]), TOS Protocol and Procedure: Breeding Landbird Abundance and Diversity (AD[07]), and TOS Science Design for Plant Diversity (AD[09]) for further details.

3.2 Temporal Sampling Design

Point counts are conducted at each point once or twice per year depending on site size, with grids at larger sites sampled only once per year, during an expert-identified window of breeding activity specific to the local area.

3.3 Variables Reported

All variables reported from the field technician (L0 data) are listed in the file, NEON Raw Data Validation for Breeding landbird point counts (DP0.10006.001) (AD[04]). All variables reported in the published data (L1 data) are also provided separately in the file, NEON Data Variables for Breeding landbird point counts (DP1.10003.001) (AD[05]).

Field names have been standardized with Darwin Core terms (http://rs.tdwg.org/dwc/; accessed 16 February 2014), the Global Biodiversity Information Facility vocabularies (http://rs.gbif.org/vocabulary/gbif/; accessed 16 February 2014), the VegCore data dictionary (https://projects.nceas.ucsb.edu/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 Geoid12A geoid model for its vertical reference surface. Latitudes and longitudes are denoted in decimal notation to six decimal places, with longitudes indicated as negative west of the Greenwich meridian.

Some variables described in this document may be for NEON internal use only and will not appear in downloaded data.

3.4 Spatial Resolution and Extent

The finest resolution at which spatial data are reported is a single point (Figure 1).

pointID → **plotID** (ID of plot/grid within site) → **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 centroid of the plot/grid where sampling occurred + associated uncertainty due to GPS error and plot width. Shapefiles of all NEON Terrestrial Observation System sampling locations can be found on the NEON science webpage at https://www.neonscience.org/data-samples/data/spatial-data-maps.

To derive a more precise estimate of the location of each point, 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:

    1. The namedLocation field in the data is the named location of the plot; more precise geographic data require the named location of the point. **pointCountMinute**s of '88' reflect incidental observations of birds after the count period has ended or while traversing between plots. These observations are certain to have occurred at the sites, somewhere near the plot, but coordinate uncertainty is unknown. For all **pointCountMinute**s other than '88', construct the named location of the point of each record in brd_perpoint and/or brd_countdata by concatenating the fields for namedLocation and pointID as: namedLocation + '.' + pointID, e.g. subplotID '1' of namedLocation 'GRSM_013.birdGrid.brd' has a complete named location of 'GRSM_013.birdGrid.brd.1'.
    2. Use the API (http://data.neonscience.org/data-api; e.g. http://data.neonscience.org/api/v0/locations/GRSM_013.birdGrid.brd.1) to query for elevation("locationElevation"), easting("locationUtmEasting"), northing("locationUtmNorthing"), coordinateUncertainty ("Value for Coordinate uncertainty"), elevationUncertainty ("Value for Elevation uncertainty"), and utmZone ("locationUtmZone") as inputs to the next step.
    3. Increase coordinateUncertainty by an appropriate amount to account for error introduced by navigating within plots. Specifically, pointIDs of '21' and '5' are monumented; the realized coordinate uncertainty off the high resolution GPS units used in plot establishment are recorded

in the coordinate uncertainty value in the spatial data. **pointCountMinute**s of '88' have an indeterminate amount of coordinate uncertainty. Other pointIDs are not monumented and are located using recreational-grade GPS units, with a realized uncertainty of ~15m.

3.5 Temporal Resolution and Extent

The finest temporal resolution at which point count data will be recorded is the minute at which an observation occurred within a six-minute sampling period, with start time and date specified. This resolution allows analyses that involve temporal subsetting of the sampling period, and also allows aggregation to coarser scales. An **eventID** is used to uniquely identify the point and time at which a point count took place, and is comprised of the plotID, pointID, and date of sampling.

The NEON Data Portal provides data in monthly files for query and download efficiency. Queries including any part of a month will return data from the entire month. Code to stack files across months is available here: https://github.com/NEONScience/NEON-utilities

3.6 Product Instances

There are a maximum of 10 9-point bird grids per site (large sites), and a minimum of 10 distributed points per site. Each point is surveyed 1-2x year, yielding a range of 20-135 6-minute point counts per site per year. The number of birds observed during each point count varies with site-specific detection probability and abundance.

3.7 Data Relationships

Each technician takes up to 3 bird quizzes over the course of the season, generating up to 3 records per technicianID per observerInstitutionName per year in brd_personnel.

Each field crew reports the source or sources used to identify birds (identificationReferences) per site per year. There is one record per site per year in brd_references, with multiple sources concatenated and separated by a semi-colon.

The protocol dictates that each pointID within a grid is counted once or twice per year (one or two expected records per pointID per plotID per year in brd_perpoint). Each combination of plotID*pointID*date generates a single eventID, which should uniquely identify each record of brd_perpoint. A record from brd_perpoint may have one or more child records in brd_countdata depending on the number of birds observed; records from these two tables can be joined on eventID.

Records from brd_countdata where pointID is '88' are not expected to have a corresponding record in brd_perpoint; observations with this pointID are incidental observations of rare or unusual birds outside of the formal point count locations and/or times.

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.

brd_personnel.csv - > One record expected per technicianID per observerInstitutionName per quiz per year.

brd_references.csv - > One record expected per siteID per year.

brd_perpoint.csv - > One record expected per date (startDate, date and time) per pointID per plotID per siteID.

brd_countdata.csv - > One record expected per date (startDate, date and time) per observation per plotID per siteID. Observations are recorded separately by taxonID, sexOrAge, clusterSize and observerDistance to facilitate distance sampling analyses.

brd_identificationHistory.csv -> One or more records expected per identificationHistoryID. Records are only created when data corrections to taxonomic identifications are made. If errors in identification are detected through QAQC processes after data publication, then corrected taxonomy will be provided in the brd_countdata table. The brd_identificationHistory table is populated with all prior names used for specimen(s) in the data product. When data are populated in the brd_identificationHistory table, **identificationHistoryID** is used as a linking variable between the brd_identificationHistory table and the other table where updates were made.

Data downloaded from the NEON Data Portal are provided in separate data files for each site and month requested. The neonUtilities package in R and the neonutilities package in Python contain functions to merge these files across sites and months into a single file for each table. The neonUtilities R 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. The neonutilities package in Python is available on the Python Package Index (PyPi; https://pypi.org/project/neonutilities/) and can be installed using pip. For instructions on using the package in either language 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 Bird Taxonomy

The master taxonomy for breeding landbirds is the American Ornithological Society (AOS) Checklist of North and Middle American Birds (7th Edition and Supplements), current through the 64th supplement (Chesser et al. 2023). Taxon ID codes used to identify taxonomic concepts in the NEON master taxonomy list are based on the standard four-character codes provided by the Institute for Bird Populations Four-

letter (English Name) codes and six-letter (Scientific Name) Alpha Codes for 2186 Bird Species (and 140 Non-Species Taxa) in accordance with the 64th AOS Supplement (2023), sorted by English name and prepared by Peter Pyle and David F. DeSante (The Institute for Bird Populations http://www.birdpop.org/pages/birdSpeciesCodes.php). The list includes North American species, supplemented with species that are expected to occur at NEON sites in Puerto Rico and Hawaii. NEON plans to keep the taxonomy updated in accordance with the AOS supplements, starting in 2019 and annually thereafter.

Geographic ranges and native statuses used in this data product are from BirdLife International, Cambridge, UK and NatureServe, Arlington, USA (Version 5.0, downloaded in 2015, http://datazone.birdlife.org/species/requestdis, Bird species distribution maps of the world).

Prior to the 2022 data release, publication of species identifications were obfuscated to a higher taxonomic rank when the taxon was found to be listed as threatened, endangered, or sensitive at the state level where the observation was recorded. The state-level obfuscation routine was removed from the data publication process at all locations excluding sites located in D01 and D20, and data have been reprocessed to remove the obfuscation of state-listed taxa for all years. Federally listed threatened and endangered or sensitive species remain obfuscated at all sites and sensitive species remain redacted at National Park sites.

4.2 Identification History

Beginning in 2023, the identificationHistory table was added to track any changes to taxonomic identifications that have been published in NEON data. Such taxonomic revisions may be necessary when errors are found in QAQC checks, or when evidence from genetic analysis of samples or re-analysis of archived samples indicate a revision is necessary. Requests for taxonomic changes are reviewed by NEON science staff. Proposed changes are evaluated based on evidence in the form of photographs, existing samples, genetic data, consultation with taxonomic experts, or range maps. Upon approval, the existing record in brd_countdata is updated with the new taxonomic information and a unique identifier is added to the identificationHistoryID field. A record with the same **identificationHistoryID** is created in the brd_identificationHistory table where the previous taxonomic information is archived along with the date the change was made.

5 DATA QUALITY

5.1 Data Entry Constraint and Validation

Constraints and data validation 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 Breeding landbird point counts (DP0.10006.001), provided with every download of this data product. Contained within this file is a field named 'entryValidationRulesParser', which describes syntactically the validation rules for each field built into the data entry application. Data entry constraints are described in Nicl syntax in the validation file provided with every data download, and the Nicl language is described in NEON's Ingest Conversion Language (NICL) specifications ([AD[13]).

5.2 Automated Data Processing Steps

Following data entry into a mobile application or web user interface, the steps used to process the data through to publication on the NEON Data Portal are detailed in the NEON Algorithm Theoretical Basis Document: OS Generic Transitions (AD[14]).

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 **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.

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)

5.5 Analytical Facility Data Quality

Breeding landbird point counts are conducted by an external contractor. Quality checks conducted before data are returned to NEON include:

1. A 100% check of all data entered into the contractor database to eliminate transcription errors
2. An “uncommon bird” check wherein contractor staff and partners check through the species detected at each NEON site to look for potentially misidentified species that do not regularly occur in the area.
    - Often, extremely rare or out of place birds are the result of a data entry error, the use of an incorrect bird code, or a misidentification.
    - If the database and datasheet concur, the proofer will attempt to verify the record using local knowledge, eBird records, field guides, and other relevant sources.
    - If the record is still questionable after these steps have been taken, the proofer will attempt to contact the field technician who collected the data and verify the record.
    - If the technician can provide supporting documentation to support the record (photos, recordings), it will remain in the database. If the record cannot be verified, it will be changed to “unknown bird” on both the datasheet and in the database.

If errors in identification are detected through QAQC processes after data publication, then corrected taxonomy is provided in the brd_countdata table and previous taxonomic information is preserved in the

brd_identificationHistory (see Section 4.2 Identification History and Section 3.7 Data Relationships above for more details).

6 REFERENCES

Chesser, R. T., S. M. Billerman, K. J. Burns, C. Cicero, J. L. Dunn, B. E. Hernández-Baños, R. A. Jiménez, A. W. Kratter, N. A. Mason, P. C. Rasmussen, J. V. Remsen, Jr., and K. Winker. 2023. Check-list of North American Birds (online). American Ornithological Society. https://checklist.americanornithology.org/taxa/

Hanni, D. J., C. M. White, N. J. Van Lanen, J. J. Birek, J. M. Berven, and M. F. McLaren. 2016. Integrated Monitoring in Bird Conservation Regions (IMBCR): Field protocol for spatially-balanced sampling of land-bird populations. Unpublished report. Bird Conservancy of the Rockies, Brighton, Colorado, USA.

Ralph, C. J., T. E. Martin, G. R. Geupel, D. F. Desante, and P. Pyle. 1993. Handbook of Field Methods for Monitoring Landbirds. Gen. Tech. Rep. PSW-GTR-144-www. Page 41. Pacific Southwest Research Station, Forest Service, U.S. Department of Agriculture, Albany, CA.