NEON USER GUIDE TO CONTINUOUS DISCHARGE (DP4.00130.001)

PREPARED BY	ORGANIZATION	DATE
Kaelin M Cawley	AQU	06/09/2025
Nick Harrison	AQU	06/09/2025
Zachary Nickerson	AQU	06/09/2025
Bobby Hensley	AQU	06/09/2025

CHANGE RECORD

REVISION	DATE	DESCRIPTION OF CHANGE
A	01/11/2018	Initial Release
B	03/01/2021	Major updates to workflow including added input tables, added data relationships, migration of gauge-pressure relationship data from being transitioned in IS to being transitioned in OS, updates to gauge and TROLL offset calculation, updates to flow series calculation, updates to systematic uncertainty propagation, and added information about including lake inflow and outflow locations in this data product.
C	01/25/2022	Updated to include details of how TOMB Continuous discharge is produced and fix typos.
D	05/15/2023	General textual updates throughout, added information on the inclusion of data from the Discharge field collection (DP1.20048.001) data product and science review flags from the Elevation of surface water (DP1.20016.001) data product in L4 processing, added information about continuous discharge active periods, updated algorithm implementation steps within the docker containers to match current processing workflow, added detail on data entry validation, data revision, provisional data status, regression coefficients, uncertainty, and quality flagging.
D.1	08/24/2023	Added table in 'Special Considerations' section detailing the variables sufficient to plot continuous discharge, stage, associated discharge and stage uncertainties, and final quality flags associated with the discharge data.
E	04/02/2024	Updated to detail new L4 discharge input data table structure, added information about inactive provisional data, updated NEON logo.
F	05/28/2025	Removed 'Algorithm Theoretical Basis' and associated sections (now published in NEON.DOC.005403), added information regarding newly-published tables for periods of corrected data, updated referenes to L0 raw pressure data inputs to be L1 water column height data inputs.

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 L4 data product, which is generated from L4 and L1 OS data, L1 IS data, and associated metadata. In the NEON data products framework, the raw data collected in the field (i.e. staff gauge measurements from a single collection event or pressure transducer readouts at 1 min interval) are considered the lowest level (L0). Raw data that have been quality checked and simple metrics that emerge from the raw data are considered L1 data products. Level 4 data products rely on inputs of any level data, often from multiple input products, and may involve calculations that use data collected over a range of spatial or temporal scales.

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 L4 data creation.

1.2 Scope

This document describes the steps needed to generate the L4 Continuous discharge (DP4.00130.001) data product and associated metadata from input data and calculations. This document also provides details relevant to the publication of the data product via the NEON data portal, with additional detail available in AD[03], provided in the download package for this data product. For information on the data products that are used as inputs for this data product, see the following: for L1 water column height data, AD[06]; for gauge height, AD[07] and AD[08]; for discharge field collection, AD[09] and AD[10]; for wadeable stream morphology, AD[11]; for bathymetric and morphological maps, AD[12]; for stage-discharge rating curve, AD[11]; for USGS streamflow data, AD[17]. Documents are available for download with the respective L1 or L4 data package or from the USGS website. Please note that raw or lower level source data products (denoted by 'DP0') may not always have the same numbers (e.g., '20048') as the corresponding L1 or L4 data product.

This document does not describe, in detail, the theory of measurement, custom algorithms, or data correction methods. That information can be found in the NEON Algorithm Theoretical Basis Document (ATBD): Stage-discharge rating curves and continuous discharge (AD[05]).

2 RELATED DOCUMENTS AND ACRONYMS

2.1 Associated Documents

AD[01]	NEON.DOC.000001	NEON Observatory Design (NOD) Requirements
AD[02]	NEON.DOC.002652	NEON Level 1, Level 2 and Level 3 Data Products Catalog
AD[03]	DP4.00130.001_variables.csv	NEON Data Variables for Continuous Discharge (DP4.00130.001)
AD[04]	NEON.DOC.001152	NEON Aquatic Sampling Strategy
AD[05]	NEON.DOC.005403	NEON Algorithm Theoretical Basis Document (ATBD): Stage-discharge rating curves and continuous discharge
AD[06]	NEON.DOC.001198	NEON Algorithm Theoretical Basis Document (ATBD): Surface Water Elevation
AD[07]	NEON_gaugeHeight_userGuide.pdf	NEON User Guide to Gauge Height (DP1.20267.001)
AD[08]	DP0.20267.001.validation.csv	NEON Data Validations for Gauge Height (DP1.20267.001)
AD[09]	NEON_fieldDischarge_userGuide.pdf	NEON User Guide to Discharge Field Collection (DP1.20048.001)
AD[10]	DP0.20048.001.validation.csv	NEON Data Validations for Discharge Field Collection (DP1.20048.001)
AD[11]	NEON_ratingCurve_userGuide.pdf	NEON User Guide to Stage-Discharge Rating Curves (DP1.00133.001)
AD[12]	NEON.DOC.000008	NEON Acronym List
AD[13]	NEON.DOC.000243	NEON Glossary of Terms
AD[14]	NEON.DOC.000927	NEON Calibration and Sensor Uncertainty Values
AD[15]	NEON.DOC.001113	Algorithm Theoretical Basis Document: Quality Flags and Quality Metrics for TIS Data Products
AD[16]	NEON.DOC.011081	Algorithm Theoretical Basis Document: QA/QC Plausibility Testing
AD[17]	https://doi.org/10.3133/wsp2175	Measurement and Computation of Streamflow Testing

2.2 Acronyms

Acronym	Definition
AOS	Aquatic Observational System
API	Application Programming Interface
BaM	Bayesian Modeling
BaRatinAGE	Bayesian Rating Curve Advanced Graphical Environment
CI	Confidence Interval
DP	Data Product
DPID	Data Product Identification Code
GUI	Graphical User Interface
hr	hour
IS	Instrumental Systems
kg	Kilogram
kPa	Kilopascal
L	Liter
L0	Level Zero (Unprocessed) Data
L1	Level One (Processed) Data
L4	Level Four (Derived and Processed) Data
m	Meter
MCMC	Markov Chain Monte Carlo
min	Minute
OS	Observational Systems
Pa	Pascal
Q	Discharge
QAQC	Quality Assurance Quality Checking
s	Second
S1	Aquatic Sensor Set One
S2	Aquatic Sensor Set Two
SRF	Science Review Flag
USACE	United States Army Corps of Engineers
USGS	United States Geological Survey
wk	Week

3 DATA PRODUCT DESCRIPTION

The Continuous discharge (DP4.00130.001) data product provides calculated stage, discharge and associated uncertainty values every minute at stream, river, and lake inflow/outflow locations. One minute resolution continuous discharge and stage data are derived from once per minute pressure readings, empirical gauge heights, pre-developed gauge height-water column height regressions, and stage-discharge rating curves. TOMB data is published at the same frequency as input USGS data, usually at the 30 minute or 1 hour resolution.

3.1 Spatial Sampling Design

Continuous discharge is developed at the site level (except for Toolik Lake, where two stations have continuous discharge produced at the inflow and outflow to the lake) using data collected near either sensor set #1 or sensor set #2, whichever is closer to the staff gauge in wadeable streams, at the near shore sensor set (i.e. sensor set #1) at large rivers, or at sensor sets located at the inflow and outflow locations of Toolik Lake. TOMB continuous discharge data is produced using USGS data and a relationship between empirical readings collected by NEON and the contemporaneous USGS flow readings. The geospatial information related to the input data is published as part of the data product package, including: **siteID** - the 4 character NEON site code, **stationHorizontalID** - the 3 digit code for the sensor set (e.g., S1 = 101/131, S2 = 102/132), and **namedLocation** (the configured location of the pressure transducer L0 input data in csd_continuousDischarge, the 4 character NEON site code in csd_continuousDischargeUSGS and sdrc_gaugePressureRelationship, and the staff gauge named location [SITE.AOS.gauge] in csd_gaugeWaterColumnRegression).

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

3.2 Temporal Sampling Design

L0 surface water pressure data is collected at a 1 min resolution. The 1 min resolution is retained in the L4 continuous discharge data product in csd_continuousDischarge.

Data in the csd_continuousDischargeUSGS table are published at approximately 1 hour resolution for TOMB only.

The gaugings published in sdrc_gaugePressureRelationship retain the resolution of the L1 data, which is measured whenever field technicians work at an aquatic site. This is at least bi-weekly for stream sites and monthly for lake and river sites.

Data for a water year in the csd_gaugeWaterColumnRegression and csd_dischargeRegressionUSGS tables are published during the last month of each water year (September), regardless of the **regressionEndDate** or **usgsDischargeRegEndDate**.

Table 1: List of geolocation inputs for Continuous Discharge.

Data Source	field	contents
NEON geolocation database	active period startDate	SITE.AOS.continuous.discharge
NEON geolocation database	active period endDate	SITE.AOS.continuous.discharge
NEON geolocation database	location startDate	SITE.AOS.gauge SITE.AOS.discharge CFGLOC
NEON geolocation database	location endDate	SITE.AOS.gauge SITE.AOS.discharge CFGLOC
NEON geolocation database	location elevation	SITE.AOS.gauge SITE.AOS.discharge CFGLOC
NEON geolocation database	location zOffset	SITE.AOS.gauge SITE.AOS.discharge CFGLOC

3.3 Variables Reported

All data and geolocation variables used as inputs for continuous discharge are listed in Table 1 and Table 2. All variables reported in the published data are also provided separately in AD[03]. 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 spatial resolution at which data are reported is a site, except at Toolik Lake (TOOK) where data are published from both the lake inflow and the lake outflow.

3.5 Temporal Resolution and Extent

The finest temporal resolution at which gauge-pressure relationship data are reported is the date and time of a set of pressure measurements and corresponding empirical staff gauge reading for the sdrc_gaugePressureRelationship table. The finest temporal resolution at which continuous stage and discharge data are reported is 1 min for the csd_continuousDischarge table and approximately 1 hour for the csd_continuousDischargeUSGS table.

Tables associated with data correction are not published at a regular temporal resolution. Records in csd_dataGapToFillMethodMapping will be published with an end date corresponding to the end of a distinct date range identified for data correction. Records in csd_constantBiasShift will be published with an end date corresponding to the end of a shifted period of record. Records in csd_gapFillingRegression will be published with an end date corresponding to the end of a period of record used to develop a regression used for data correction. See AD[05] for more information on data correction methods and temporal

Table 2: For each derived data table in this data product, all NEON input tables and data sources are listed with the data product in which they are published on the NEON Data Portal

Continuous Discharge Table	Portal DP	Portal DPID	Input Table/Data Source
sdrc_gaugePressureRelationship	Continuous Discharge	DP4.00130.001	csd_gaugeWaterColumnRegression
sdrc_gaugePressureRelationship	Gauge Height	DP1.20267.001	gag_fieldData
sdrc_gaugePressureRelationship	Discharge Field Collection	DP1.20048.001	dsc_fieldData
sdrc_gaugePressureRelationship	Elevation of Surface Water	DP1.20016.001	EOS_1_min
csd_continuousDischarge	Stage-Discharge Rating Curves	DP4.00133.001	sdrc_curveIdentification
csd_continuousDischarge	Continuous Discharge	DP4.00130.001	csd_gaugeWaterColumnRegression
csd_continuousDischarge	Elevation of Surface Water	DP1.20016.001	EOS_1_min
csd_continuousDischarge	Stage-Discharge Rating Curves	DP4.00133.001	sdrc_controlInfo
csd_continuousDischarge	Stage-Discharge Rating Curves	DP4.00133.001	sdrc_priorParameters
csd_continuousDischarge	Continuous Discharge	DP4.00130.001	sdrc_gaugePressureRelationship
csd_continuousDischarge	Stage-Discharge Rating Curves	DP4.00133.001	sdrc_stageDischargeCurveInfo
csd_continuousDischarge	Stage-Discharge Rating Curves	DP4.00133.001	sdrc_sampledParameters
csd_continuousDischarge	Stage-Discharge Rating Curves	DP4.00133.001	sdrc_gaugeDischargeMeas
csd_continuousDischargeUSGS	Continuous Discharge	DP4.00130.001	sdrc_dischargeRegressionUSGS

extent.

3.6 Associated Data Streams

The data from this L4 data product are derived from a L1 data product: Elevation of surface water (DP1.20016.001). Unique combinations of csd_continuousDischarge:**siteID**x**stationHorizontalID**x**endDate** can be linked to unique combinations of EOS_1_min:**siteID**x**horizontalPosition**x**endDateTime**.

The data from this L4 data product are derived from a L1 data product: Gauge height (DP1.20267.001). These data products can be linked by gag_fieldData:**eventID** and sdrc_gaugePressureRelationship:**gaugeEventID**.

The data from this L4 data product are derived from a L1 data product: Discharge field collection (DP1.20048.001). Unique combinations of sdrc_gaugePressureRelationship:**siteID**x**gaugeCollectDate** can be linked to unique combinations of dsc_fieldData:**siteID**x**collectDate**.

The data from this L4 data product are derived from a L4 data product: Stage-discharge rating curves (DP4.00133.001). Data in csd_continuousDischarge can be linked to sdrc_stageDischargeCurveInfo, sdrc_gaugeDischargeMeas, sdrc_posteriorParameters, sdrc_sampledParameters, sdrc_resultsResiduals, and sdrc_curveIdentification by **curveID**. Data in csd_continuousDischarge can be linked to hydrologic controls (priors) data by first linking to sdrc_curveIdentification by **curveID** then linking sdrc_curveIdentification:**siteID**x**controlSurveyEndDateTime** to sdrc_controlInfo:**siteID**x**endDate** and sdrc_priorParameters:**siteID**x**endDate**

3.7 Product Instances

The NEON Observatory contains 24 wadeable streams, 3 large rivers, and 1 lake site containing an inflow and an outflow where discharge is measured.

At each site or location, this data product yields a maximum of 104 gauge and mean pressure readings per year (~2 per wk) in the sdrc_gaugePressureRelationship table, 525,600 records per year (~1 per min) in the csd_continuousDischarge table, and 8,760 records per year (~1 per hr) in the csd_continuousDischargeUSGS.

3.8 Data Relationships for All Sites Except TOMB

The algorithm used for this L4 data product produces as many records in sdrc_gaugePressureRelationship as there are unique gauging events (site *x* day resolution; i.e., SITE.YYYYMMDD formatted **eventID** in gag_fieldData) for a site across DP1.20267.001 (gag_fieldData table) and DP1.20048.001 (dsc_fieldData table) for a given water year (October 1st to September 30th). Beginning with RELEASE-2023 data and forward, when gauging events are available in both data products on the same day, the gauging from DP1.20048.001 is given precedence due to its direct association with the rating curve (see AD[11]). The number of total unique gauging events may differ from the number of records in sdrc_gaugePressureRelationship for a given site and water year because a gauging will not be processed into sdrc_gaugePressureRelationship if there are no water column height measurements within 20 min of the gauging's collect date. Additionally, if all the water column height measurements within 20 min of a gauging have received a SRF in DP1.20016.001, the gauging will also not be processed into

sdrc_gaugePressureRelationship. The sdrc_gaugePressureRelationship table reports the relationship between measured gauge heights and calculated stage values derived from previously-developed gauge height-water column height linear regressions. The data in this table is used to calculate **systematicUncertainty** in csd_continuousDischarge. Processing of sdrc_gaugePressureRelationship data will occur annually for the previous water year following the end of the water year when the DP4.00133.001 data product is processed.

Data in both sdrc_gaugePressureRelationship and csd_continuousDischarge use staff gauge offset values to correct data following changes in the physical location of infrastructure. Offsets are calculated using information derived from the NEON geolocation database (Table 1). The processes for calculating offsets in both staff gauge and pressure transducer infrastructure is the same:

1. A total reference elevation is calculated for a location as the sum of its **elevation** above sea level (m) and **zOffset** (m) (vertical correction needed in order for the location to be relatable to previous locations).
2. The total reference elevation for the initial location is subtracted from each subsequent location to obtain the offset value for each subsequent location.
3. Every staff gauge record between the start date and end date of a specific location will have the appropriate offset applied.

During active periods, one record is created in csd_continuousDischarge per min regardless of whether the pressure transducer was producing data. An active period is defined as a period of time where a sensor is installed and data are expected. In the case of Continuous discharge, active periods can be adjusted due to seasonal conditions (e.g. at OKSR at TOOK where pressure transducers are installed year round but do not log data during the winter due to power limitations). However, for some situations (e.g. a high flow event disrupts the pressure transducer causing to it stop collecting measurements for > 1 calendar month), an active period may be adjusted to avoid producing data files with no usable data. For timestamps when the pressure transducer is not collecting measurements within an active period, records in csd_continuousDischarge are produced containing the timestamp, applicable flags, and metadata, but no stage or discharge series data. Some NEON sites are seasonal due to climate or logistical constraints. Records will not be processed and published at seasonal sites when sensors are intentionally deactivated or removed from the site.

The regression coefficients used to calculate sdrc_gaugePressureRelationship: **calculatedStage** and csd_continuousDischarge:**estimatedStage** are published in csd_gaugeWaterColumnRegression. This table is available as part of the expanded download packages for this data product. The coefficient of determination (R^2) is provided along with slope and y-intercept values to allow users to assess the correlation of the model.

sdrc_gaugePressureRelationship.csv - > One record expected per unique gauging event for a site or location in the past water year that also has at least 1 min of water column height data available within 20 min of the staff gauge reading without science review flags in DP1.20016.001.

csd_continuousDischarge.csv - > One record expected per min during active periods regardless of available water column height data. Inactive periods will contain gaps in data.

csd_gaugeWaterColumnRegression - > One record expected per unique gauge height - water column

height linear regression. Each record in this table will contain a unique identifier as the **regressionID** variable.

regressionID can be used to link sdrc_gaugePressureRelationship, csd_continuousDischarge, and csd_gaugeWaterColumnRegression data.

3.9 Data Relationships for TOMB

Continuous discharge at one NEON site is not developed using the Bayesian model. At D08 TOMB (located on the Tombigbee River in southwest Alabama) a downstream lock and and dam system operated by the USACE highly influences the stage-discharge relationship at the NEON site. Publicly available discharge data collected by the USGS at the dam are thus utilized to publish continuous discharge. The csd_continuousDischargeUSGS table represents the data published in this manner. The csd_dischargeRegressionUSGS table contains input data for csd_continuousDischargeUSGS and describes the fit between empirical NEON discharge measurements and published USGS discharge data.

csd_continuousDischargeUSGS - > Approximately one record per hour that contains USGS discharge and associated uncertainty based on the fit of USGS data with corresponding discharge measurements collected at the adjacent NEON site.

csd_dischargeRegressionUSGS - > One record expected per unique NEON discharge - USGS discharge linear regression developed at a site. Every record in this table will contain a unique identifier as the **usgsRegressionRegID** variable.

usgsRegressionRegID can be used to link csd_continuousDischargeUSGS and bat_dischargeRegressionUSGS data.

3.10 Data Relationships for Corrected Data

Beginning with RELEASE-2025, data corrections were applied to csd_continuousDischarge and csd_continuousDischargeUSGS for distinct time periods. See AD[05] for details on the data correction process, methods, and timeline. Up to three tables can be published during time periods when data were corrected, with the following relationships:

csd_dataGapToFillMethodMapping - > One record expected per data gap. Data gap is defined as a period of missing record or a period of record flagged for correction (AD[05]). Unique records in this table are identified by combinations of **namedLocation**x**endDate**x**gapFilledDataStream** and can be linked to csd_continuousDischarge and csd_continuousDischargeUSGS by matching **siteID** and date ranges.

csd_constantBiasShift - > One record expected per period of record shifted. Unique records in this table are identified by combinations of **namedLocation**x**endDate**x**gapFilledDataStream** and can be linked to csd_continuousDischarge and csd_continuousDischargeUSGS by matching **siteID** and date ranges.

csd_gapFillingRegression - > One record expected per unique regression developed for correcting data at a site. Every record in this table will contain a unique identifier as the **gapFillRegressionID** variable. **gapFillRegressionID** can be used to link csd_gapFillingRegression and csd_dataGapToFillMethodMapping data when csd_dataGapToFillMethodMapping:**gapFillMethod** is ‘transducer’, ‘conductivity’, or ‘usgs’. See AD[05] for details on each data correction via regression method.

3.11 Special Considerations

On Github, NEON maintains the public NEON-stream-discharge repository (https://github.com/NEONScience/NEON-stream-discharge) that contains scripts, packages, and an R shiny application that allow users to access and interpret NEON hydrology data published in this and other data products.

In the NEON-stream-discharge repository, the /L4Discharge/ directory contains the R package *stageQCurve*. The *stageQCurve* package serves two purposes: 1) to mirror the internal processing codes that produce L4 data tables for DP4.00133.001 and DP4.00130.001, and 2) to run BaM in 'predictive' mode with DP4.00133.001 data downloaded from the NEON Data Portal that produces posterior rating curves and associated uncertainties across a range of stages.

Also, R scripts in the /hydrologicControls/ directory to generate and visualize the hydrologic control data used to model the prior rating curve in BaM. The directory is structured by the 4 letter NEON site code. Each site sub directory contains a script for each survey of the discharge cross-section at that site (linked to unique combinations of **siteID** and **controlSurveyEndDateTime** in sdrc_curveIdentification, see AD[11] for more information).

Also available in the NEON-stream-discharge repository is a data visualization application developed using R Shiny. The openFlow app, available to run locally in the /shiny-openFlow/ directory, combines multiple data products to provide users with a tool to holistically view the hydrology of NEON stream and river sites. In openFlow, users can view data from this data product (DP4.00130.001), Stage-discharge rating curves (DP4.00133.001), Precipitation (DP1.00006.001) and Land-water interface images (DP1.20002.001). For more information about installing and running openFlow and its dependencies in R, view the Github README in the /shiny-openFlow/ directory. Once in the app, users can click the "About the App" tab for a detailed README that explains the features, use and functionality of openFlow.

The openFlow application is also available as a stand-alone user interface publicly-available on the NEON website. Visit https://openflow.neonscience.org/ to interact with multiple NEON ecohydrology data products!

Table 3: For most research purposes, the following variables listed are sufficient to plot continuous discharge, stage, associated discharge and stage uncertainties, and final quality flags associated with the discharge data. More details are available in the variables file that accompanies each download package.

Continuous Discharge Table	Variable	Description
csd_continuousDischarge	maxpostDischarge	Discharge time series (L/s)
csd_continuousDischarge	withRemnUncQUpper2Std	Upper bound of the 95% CI for maxpostDischarge, expressing remnant error (L/s)
csd_continuousDischarge	withRemnUncQLower2Std	Lower bound of the 95% CI for maxpostDischarge, expressing remnant error (L/s)
csd_continuousDischarge	equivalentStage	Stage time series (m)
csd_continuousDischarge	stageUnc	Uncertainty associated with equivalentStage, expressing the sum of systematic and non-systematic uncertainty (m)
csd_continuousDischargeUSGS	usgsDischarge	For the D08 TOMB site only, discharge time series from the nearby USGS site (L/s)
csd_continuousDischargeUSGS	withRegressionUncQUpper2Std	For the D08 TOMB site only, the upper bound of the 95% CI for usgsDischarge, expressing error associated with the fit of measured NEON discharge with USGS discharge (L/s)
csd_continuousDischargeUSGS	withRegressionUncQLower2Std	For the D08 TOMB site only, the lower bound of the 95% CI for usgsDischarge, expressing error associated with the fit of measured NEON discharge with USGS discharge (L/s)
csd_continuousDischarge csd_continuousDischargeUSGS	dischargeFinalQF	Final quality flag indicating whether the discharge time series has passed or failed overall quality assessment (1 = fail, 0 = pass).

4 DATA QUALITY

4.1 Data Entry Constraint and Validation

Many quality control measures are implemented at the point of data entry (i.e., the L1 data that are used as an input for this data product) within a mobile data entry application or web user interface. See AD[08] and AD[10] for more details.

In this L4 data product, L1 input data are subject to additional constraint and validation in the form of manual review of the stage-discharge relationship and gauge-water column height regressions by NEON scientists. When a rating curve is developed, gaugings published in DP1.20048.001 may be excluded if they are determined to be outliers due to human error (i.e., transcription error), contain quality flags in the L1 data product (See AD[09] for information on L1 quality flags), or report temporary hydrologic conditions at the discharge cross-section (information found in dsc_fieldData:**dscTempHydroCond**). When gauge-water column height regressions are developed, NEON scientists review the quality of both the L1 gauge height and L1 water column height data. If the L1 gauge height is determined to be in error, the **initialStageHeight** and **endStageHeight** values will be deleted from gag_fieldData in DP1.20267.001 (see AD[07] and AD[08]). If water column height data are determined to be invalid, a science review flag will be added to the affected time range in DP1.20016.001, which will programatically exclude associated gaugings in the processing codes (see Section 3.8 of this document).

4.2 Provisional Data and Data Revisions

All data are Provisional until a tagged version is released. Annually, NEON releases a static version of all or almost all data products, annotated with digital object identifiers (DOIs). 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 has the most up-to-date information and contains a history of changes and known errors. This data product was first included in RELEASE-2022, the second NEON data release.

Publication Timeline and Annual Reprocessing: Throughout the year, Provisional csd_continuousDischarge and csd_continuousDischargeUSGS data are automatically published using input data (Table 2) from the previous water year, following the publication schedule described in Section 3.2 of this document. Note that these Provisional data are not fully QAQC'ed. At the end of each water year, NEON scientists develop, review, and publish all the input data (e.g. stage discharge rating curves, stage models, etc.) for each site or location for that water year. Once the rating curves for the complete water year have been developed, csd_continuousDischarge and csd_continuousDischargeUSGS data are reprocessed and re-published for inclusion in the next data Release.

Due to this data processing schedule, the difference between Provisional and Released data is larger in Continuous discharge (DP4.00130.001) than in most NEON data products. Provisional data in csd_continuousDischarge can be inaccurate if calibration factors, geolocation data, or the physical conditions of the site (e.g. discharge cross-section morphology) have changed relative to input data from the previous water year. Most Provisional data issues of this nature will not be addressed until annual reprocessing occurs at the end of a water year.

Provisional data may be made inactive, and therefore not published, if the morphological and hydrologic

regimes become vastly different than that used to develop the models used to estimate Provisional stage and discharge (e.g., new surveys of the staff gauge and/or discharge cross section required due to major flow event, rating curve shift, or issues with the instrument infrastructure). If users observe a site with inactive Provisional data (i.e., recent, unpublished months), please review the issue log where inactive provisional date ranges will be reported.

In csd_continuousDischargeUSGS, the **usgsValueQualCode** field reports USGS Instantaneous and Daily Data-Value Qualification Codes included in the data downloaded from the *dataRetrieval* R package (P = Provisional, A = Approved; De Cicco et al., 2022). Data users should be aware of the USGS Provisional Data Statement (https://waterdata.usgs.gov/provisional-data-statement/) when working with Provisional data in csd_continuousDischargeUSGS.

Users are encouraged to check the curveID and regressionID associated with continuous discharge and stage data, respectively, in order to determine which rating curve or stage model coefficients were applied to calculate the value. Users should also check the issue log of this data product for any major Provisional data quality issues that cannot be resolved until annual reprocessing at the end of a water year. If a user's research requires the use of Provisional data (i.e., for near real-time hydrologic forecasting), users can contact NEON scientists (https://www.neonscience.org/about/contact-us) for more information on the validity and potential use of Provisional data in this data product.

4.3 Uncertainty

One of the benefits of using BaM and MCMC sampling is that there are a large number of realizations from the posterior distribution, which can be used to quantify uncertainty associated with the maximum likelihood posterior parameters (BaRatin statistical model documentation and Le Coz et al., 2014). NEON publishes both the parametric and remnant (structural) error based off of 500 realizations from the posterior distribution.

Note that the uncertainty published in the basic download package of this data product is expanded uncertainty, i.e. multiplied by a factor of 1.96 to cover two standard deviations, or, the 95% confidence interval. When using the BaRatin GUI tool, the uncertainty should be represented the same way as NEON publishes it. Note that for the BaM executable, uncertainty is represented as one standard deviation. These values are also published in this data product as part of the expanded download package. The 1 standard deviation uncertainty from the expanded download package should be used when writing out NEON data and configurations.

For TOMB data, the standard error of the regression between NEON empirical discharge readings and USGS discharge data are used to calculate the 95% confidence intervals (2 standard deviations) associated with the discharge prediction.

4.4 Quality Flagging

For the quality flags in the csd_continuousDischarge table, see the descriptions in AD[13] and AD[14] for more details on the automated quality flagging associated with instrument data.

A general overview of the Science Review Flag is also described in AD[13]. In this data product, **dischargeFinalQFSciRvw** contains science review flags in csd_continuousDischarge and

csd_continuousDischargeUSGS. Most science review flags in this data product will be assigned at the end of a water year during annual review and reprocessing for a site. During annual review, NEON scientists will assess periods of erratic or erroneous sensor data and apply science review flagging as needed. Situations in which continuous discharge may be flagged include, but are not limited to, ice periods (flow over ice, ice pinching sensor housing, etc.), sensor instability within the housing, sensor biofouling, or an exposed sensor during baseflow or dry periods. Peak flows in csd_continuousDischarge may also receive a science review flag if **maxpostDischarge** far exceeds the highest empirical gauging in the associated rating curve (**curveID**) or the predicted flow appears implausible given historic data.

Because the Elevation of surface water (DP1.20016.001) data product and this data product share the same data source (L0 pressure transducer), science review flags from DP1.20016.001 are published as csd_continuousDischarge:**sWatElevFinalQFSciRvw** in the expanded download package.

In csd_continuousDischargeUSGS, the **usgsValueQualCode** field reports USGS Instantaneous and Daily Data-Value Qualification Codes included in the data downloaded from the *dataRetrieval* R package (De Cicco et al., 2022). In addition to reporting the Provisional status of the data, this field also reports any codes that relate to data quality (e.g., *e* for discharge values that are edited or estimated; more information on USGS data quality codes can be found at https://help.waterdata.usgs.gov/codes-and-parameters/instantaneous-value-qualification-code-uv_rmk_cd).

5 REFERENCES

De Cicco, L.A., R.M. Hirsch, D. Lorenz, W.D. Watkins, M. Johnson (2022) dataRetrieval: R packages for discovering and retrieving water data available from Federal hydrologic web services, v.2.7.12. DOI: 10.5066/P9X4L3GE

Le Coz, J., B. Renard, L. Bonnifait, F. Branger, R. Le Boursicaud (2014) Combining hydraulic knowledge and uncertain gaugings in the estimation of hydrometric rating curves: A Bayesian approach. *Journal of Hydrology*, 509:573-587. DOI: 10.1016/j.hydrol.2013.11.2016