NEON USER GUIDE TO CONTINUOUS DISCHARGE (DP4.00130.001)

PREPARED BY	ORGANIZATION	DATE
Kaelin M Cawley	AQU	08/29/2023
Nick Harrison	AQU	08/29/2023
Zachary Nickerson	AQU	08/29/2023
Bobby Hensley	AQU	08/29/2023

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.

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 and L0 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.

This document describes the process for performing custom calculations to convert NEON (all sites except for TOMB) or USGS data (only the TOMB site) into Continuous discharge (DP4.00130.001). For information of the raw data that are ingested and processed for the source data product see the following: for L0 pressure transducer data, AD[05]; for gauge height, AD[06] and AD[07]; for discharge field collection, AD[08] and AD[09]; for wadeable stream morphology, AD[10]; for bathymetric and morphological maps, AD[11]; for stage-discharge rating curve, AD[12]; for USGS streamflow data, AD[18]. 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.

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.001198vB	NEON Algorithm Theoretical Basis Document (ATBD): Surface Water Elevation
AD[06]	NEON_gaugeHeight_userGuide.pdf	NEON User Guide to Gauge Height (DP1.20267.001)
AD[07]	DP1.20267.001_validation.csv	NEON Data Validations for Gauge Height (DP1.20267.001)
AD[08]	NEON_fieldDischarge_userGuide.pdf	NEON User Guide to Discharge Field Collection (DP1.20048.001)
AD[09]	DP1.20048.001_validation.csv	NEON Data Validations for Discharge Field Collection (DP1.20048.001)
AD[10]	NEON_streamMorphology_userGuide.pdf	NEON User Guide to Wadeable Stream Morphology (DP4.00131.001)
AD[11]	NEON_bathymetry_userGuide.pdf	NEON User Guide to Bathymetric and Morphological Maps (DP4.00132.001)
AD[12]	NEON_ratingCurve_userGuide.pdf	NEON User Guide to Stage-Discharge Rating Curves (DP1.00133.001)
AD[13]	NEON.DOC.000008	NEON Acronym List
AD[14]	NEON.DOC.000243	NEON Glossary of Terms
AD[15]	NEON.DOC.000927	NEON Calibration and Sensor Uncertainty Values
AD[16]	NEON.DOC.001113	Algorithm Theoretical Basis Document: Quality Flags and Quality Metrics for TIS Data Products
AD[17]	NEON.DOC.011081	Algorithm Theoretical Basis Document: QA/QC Plausibility Testing
AD[18]	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
SRFs	Science Review Flags
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 geo/bat_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 geo_gaugeWaterColumnRegression, bat_gaugeWaterColumnRegression, and bat_dischargeRegressionUSGS tables are published during the last month of each water year (September), regardless of the **regressionEndDate** or **usgsDischargeRegEndDate**.

Table 1: List of calibration inputs for Continuous Discharge.

Internal Notation	Description
CVAL_B0	Calibration coefficient for LT500 pressure sensor
CVAL_B1	Calibration coefficient for LT500 pressure sensor
CVAL_B2	Calibration coefficient for LT500 pressure sensor
U_CVALA2	LT500 pressure sensor repeatability

Table 2: 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 3, Table 1, and Table 2. For more information on calibration and validation assessment of sensors repeatability, see AD[15]. All variables reported in the published data (L4 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.

Table 3: 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 (Note: L0 data is not available on the NEON Data Portal).

Continuous Discharge Table	**Portal DP**	**Portal DPID**	**Input Table/Data Source**
sdrc_gaugePressureRelationship	Continuous Discharge	DP4.00130.001	bat_gaugeWaterColumnRegression geo_gaugeWaterColumnRegression
sdrc_gaugePressureRelationship	Gauge Height	DP1.20267.001	gag_fieldData
sdrc_gaugePressureRelationship	Discharge Field Collection	DP1.20048.001	dsc_fieldData dsc_fieldDataADCP
sdrc_gaugePressureRelationship	Elevation of Surface Water	DP1.20016.001	SRF
sdrc_gaugePressureRelationship	Elevation of Surface Water, L0	DP0.20016.001	L0 Pressure
csd_continuousDischarge	Stage-Discharge Rating Curves	DP4.00133.001	bat_curveIdentification geo_curveIdentification
csd_continuousDischarge	Continuous Discharge	DP4.00130.001	bat_gaugeWaterColumnRegression geo_gaugeWaterColumnRegression
csd_continuousDischarge	Elevation of Surface Water	DP1.20016.001	SRF
csd_continuousDischarge	Elevation of Surface Water, L0	DP0.20016.001	L0 Pressure
csd_continuousDischarge	Stage-Discharge Rating Curves	DP4.00133.001	bat_controlInfo geo_controlInfo
csd_continuousDischarge	Stage-Discharge Rating Curves	DP4.00133.001	bat_priorParameters geo_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	bat_dischargeRegressionUSGS

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.

3.6 Associated Data Streams

The data from this L4 data product share the same L0 pressure input data as the L1 Elevation of surface water (DP1.20016.001) data product. While the same L0 inputs are used for the two data products, the units and temporal resolution of the output data are different. Science review flags (SRFs) from DP1.20016.001 (**sWatElevFinalQFSciRvw** published in EOS_30_min and EOS_5_min) are also used during data processing and are published in csd_continuousDischarge in the expanded download package. For more information on SRFs, see AD[16].

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 either dsc_fieldData:**siteID**x**collectDate** or dsc_fieldDataADCP:**siteID**x**endDate**.

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 geo/bat_curveIdentification by **curveID**. Data in csd_continuousDischarge can be linked to hydrologic controls (priors) data by first linking to geo/bat_curveIdentification by **curveID** then linking geo/bat_curveIdentification:**siteID**x**controlSurveyEndDateTime** to geo/bat_controlInfo:**siteID**x**endDate** and geo/bat_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 and quantified.

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, and dsc_fieldDataADCP tables) 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[12]). 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 L0 pressure transducer measurements within 20 min of the gauging's collect date. Additionally, if all the L0 pressure transducer 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 and pressure transducer 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 2). 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 measurement (stage measurement for staff gauges and calibrated water column height for pressure transducers) 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 the geo_gaugeWaterColumnRegression table for stream sites and bat_gaugeWaterColumnRegression for river sites and lake inflow and outflow locations. These tables are 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 L0 pressure 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 L0 pressure data. Inactive periods will contain gaps in data.

geo_gaugeWaterColumnRegression - > One record expected per unique gauge height - water column height linear regression developed at a stream site. Each record in this table will contain a unique identifier as the **regressionID** variable.

bat_gaugeWaterColumnRegression - > One record expected per unique gauge height - water column height linear regression developed at a river or lake inflow and outflow site. 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 geo/bat_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 bat_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.

bat_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 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.

In 2023, NEON will begin publishing code 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 geo/bat_curveIdentification, see AD[12] for more information). Until all surveys are uploaded to the repository, controls scripts for any survey can be made available upon request to NEON staff.

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.

For ease of access, NEON also hosts a constrained version of openFlow on a public Shiny server. Visit https://neonproject.shinyapps.io/openFlow_alpha/ to access the publicly-hosted version of openFlow. *Note: this version of openFlow is constrained to a 90 day date range and single user access.*

Table 4: 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 ALGORITHM THEORETICAL BASIS

4.1 Theory of measurement

Continuous discharge is the volume of water flowing through a stream over time and is a function of the channel cross-sectional area and water velocity. Practically, a stage-discharge rating curve can be developed to enable the conversion of stage, a relative measure of water column height, to stream discharge (Figure 1).

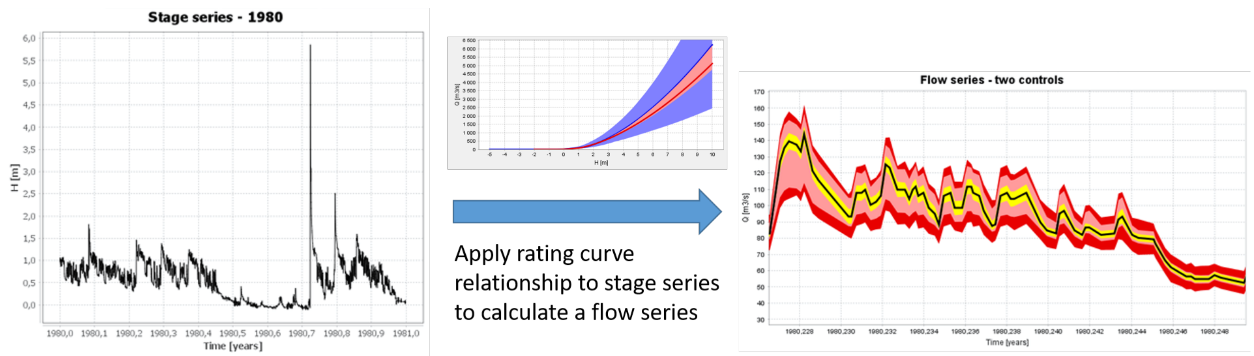

Figure 1: Example stage series and flow series output. Pictures taken from ”Stage series” and ”Flow series” documentation for BaRatinAGE (Le Coz et al., 2013; Le Coz et al., 2014).

4.2 Theory of Algorithm for All Sites Except for TOMB

Pressure transducer sensors are installed at all NEON aquatic sites and are used to develop a continuous discharge record at stream, river and lake inflow and outflow sites. For more information on the pressure transducers NEON deploys, see AD[05]. Calibration coefficients are applied to raw surface water pressure data to derive calibrated pressure. Calibrated pressure can be used to estimate stage by (1) converting to water column height; (2) applying an offset from the site survey (Equation 1); and (3) by applying the appropriate water column height-gauge height regression coefficients to the offset estimated water column height (Equation 2).

$$h_{wc} = \frac{P_{sw}}{\rho \cdot g} \cdot 1000 + h_{stage} \quad (1)$$

where,

h_{wc} = offset estimated water column height, m

P_{sw} = calibrated surface water pressure, kPa

ρ = Density of water, 999 kg/m^3

g = Acceleration due to gravity, 9.81 m/s^2

1000 = conversion from kPa to Pa (1 Pa is equivalent to 1 kg·m^{-1}s^{-1})

h_{stage} = offset between pressure transducer and staff gauge reading, m

$$h = m_{reg} \cdot h_{wc} + b_{reg} \tag{2}$$

where,

h = estimated stage, m

m_{reg} = slope of the water column height-gauge height linear regression, unitless

b_{reg} = intercept of the water column height-gauge height linear regression, m

The estimated stage timeseries is converted to discharge by applying coefficients derived from the stage-discharge rating curve (DP4.20133.001) to the stage timeseries data. NEON stage-discharge rating curves and continuous discharge stage timeseries are developed using a Bayesian modeling technique developed by the Bayesian Rating Curve Advanced Graphical Environment (BaRatinAGE) development team (Le Coz et al., 2013; Le Coz et al., 2014). The executable and/or a GUI is freely available with an individual license by sending an email to: baratin.dev@lists.irstea.fr.

The stage-discharge rating curve relies on a "prior" rating curve that is developed from the hydraulic controls. The physical dimensions of the channel, the number of hydraulic controls selected, and the physical dimensions of the hydraulic controls are derived from cross-section survey data. Exponential equations for each control are then calculated (Equation 3). A "posterior" rating curve is then fit using the "prior" rating curve and the gauging records using Bayesian estimation of the rating curve and a Markov Chain Monte Carlo (MCMC) sampling (Le Coz, 2014). Equation 3 and text below are taken from "Rating curve equation" documentation for BaRatinAGE (Le Coz et al., 2013).

$$Q(h) = \sum_{r=1}^{N_{segment}} \left(1_{[K_{r-1};K_r]}(h) \times \sum_{j=1}^{N_{control}} M(r,j) \times a_j (h - b_j)^{c_j} \right) \tag{3}$$

In the above equation, $M(r, j)$ is the matrix of controls, and the notation $1_I(h)$ denotes a function equal to 1 if h is included in the interval I, and zero otherwise. This equation shows that the stage discharge relation is a combination of power functions, and the matrix of controls is used to specify how this combination operates (succession or addition of controls). For more information on the development of NEON stage-discharge rating curves, see AD[10].

4.3 Algorithm Implementation for All Sites Except for TOMB

The NEON IS transition system runs Docker containers containing R code to generate the sdrc_gaugePressureRelationship and csd_continuousDischarge tables (similar infrastructure to Metzger, 2017 without the use of HDF5 file formats).

Within the Docker container used to generate sdrc_gaugePressureRelationship, run on a water year of data:

1. L1 gauging data for the site and water year are queried from the NEON database.

2. L1 surface water elevation science review flags are queried from the NEON database.
3. Gauging records that fall within periods of L1 surface water elevation science review flags are removed from dataset.
4. Geolocation information for the staff gauge location (SITE.AOS.gauge/SITE.AOS.discharge) are queried from the NEON database.
5. Gauging records are offset based on geolocation information.
6. L0 pressure data for the site for each day of gauging events are queried from the NEON database.
7. L0 pressure data are averaged +/- 10 min around gauging collect date and appended to the gauging data table. Gaugings with no associated mean L0 pressure data are removed from the dataset.
8. Calibration factors for the L0 pressure data are queried from the NEON database.
9. Geolocation information for the pressure transducer location (CFGLOC) is queried from the NEON database.
10. Calibration factors and geolocation data are used to convert raw pressure transducer to calculate and offset the height of the water column.
11. L4 gauge height - water column height regression coefficients are queried from the NEON database.
12. Regression coefficients are used to convert water column height to estimated stage.
13. L4 data tables are written to the NEON database for publication to users.

Within the Docker container used to generate csd_continuousDischarge, run on a single day of data:

1. The date processed is checked against L4 continuous discharge active periods queried from the NEON database.
2. L0 pressure data for the site and day are queried from the NEON database.
3. L0 data is regularized to fill gaps with null timestamps at a 1 min temporal resolution.
4. Calibration factors for the L0 pressure data are queried from the NEON database.
5. Geolocation information for the pressure transducer location (CFGLOC) is queried from the NEON database.
6. Calibration factors and geolocation data are used to convert L0 pressure data to offset water column height.
7. L4 gauge height - water column height regression coefficients are queried from the NEON database.
8. Regression coefficients are used to convert water column height to estimated stage.
9. L4 gauge-pressure relationship data are queried from the NEON OS database.
10. Calibration uncertainty and gauge-pressure relationship uncertainty are combined to estimate uncertainty in calculated stage.
11. L4 stage-discharge relationship, hydraulic controls, and sampled MCMC parameter data are queried from the NEON OS database to use as inputs to BaM.
12. Configuration files for BaM are created.
13. BaM executable is run and L4 data tables are output.
14. L4 continuous discharge threshold data are queried from the NEON database.
15. Thresholds are applied to L4 data tables outputs and automatic quality flags are calculated.
16. L1 surface water elevation science review flags are queried from the NEON database and appended to L4 data tables.
17. L4 data tables are written to the NEON database for publication to users.

4.4 Theory of Algorithm for TOMB

A lock and dam structure is operated by the USACE and USGS agencies approximately 14 km downstream of the NEON discharge cross-section at D08 TOMB. This structure aims to maintain a stable stage level to support navigation of the channel across a broad range of discharge. The stage-discharge relationship at the NEON site is heavily impacted by the opening and closing of floodgates, which invalidates the inputs to the BaM model. However, a relationship between the USGS discharge data and NEON empirical measurements are well-correlated (Figure 2). With support from the aquatic ecology community, a different data processing pipeline was developed for TOMB using USGS data as the primary input.

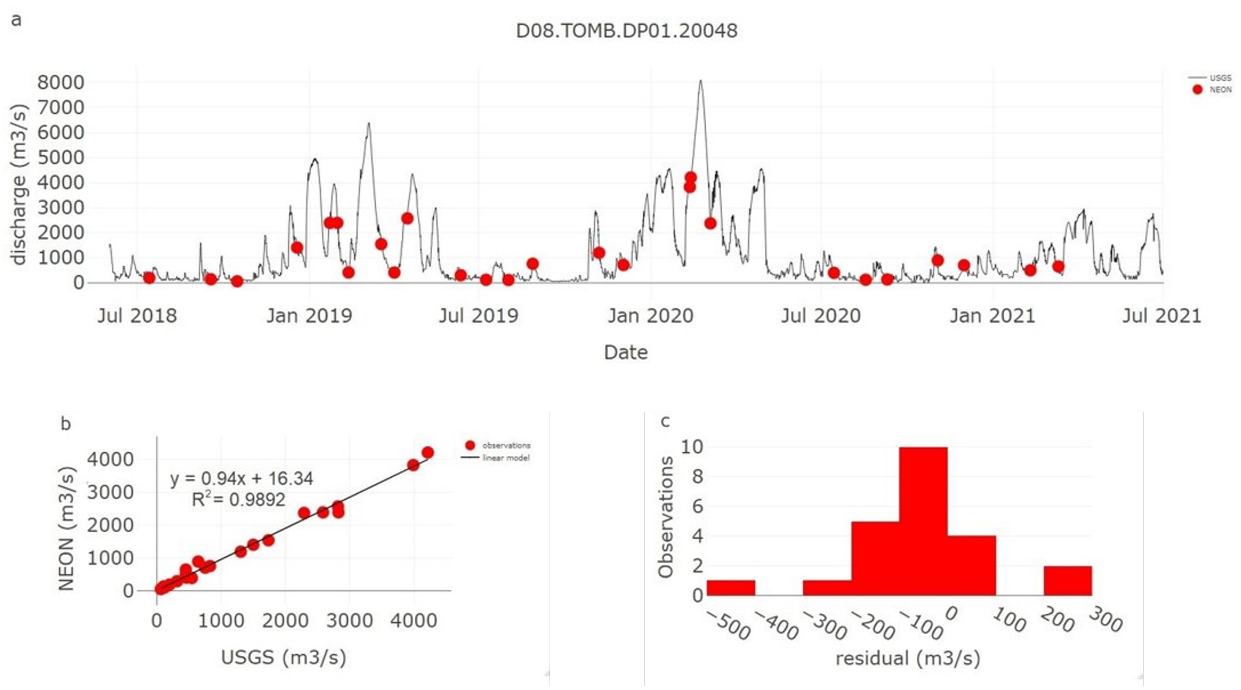

Figure 2: Top panel (a) shows the USGS flow timeseries with red dots for NEON empirical measurements, bottom panel (b) shows a linear regression and equation between USGS data at the time of NEON empirical measurements, and bottom panel (c) shows the distribution of residuals from the regression relationship.

At D08 TOMB continuous discharge is derived by estimates made by the USGS at site 02469761 (Coffeeville Dam). At the end of each water year, discharge data are downloaded from the USGS API using the USGS *dataRetrieval* R package (De Cicco et al., 2022). These data are then regressed against empirical discharge measurements (dsc_fieldDataADCP table in DP1.20048.001) collected by NEON staff at the NEON site. Data contained within the regression may span multiple water years depending on the stability of the relationship. Because discharge values span multiple orders of magnitude, they are first log-transformed. The standard error of this regression is then used to calculate the 95% confidence intervals associated with the discharge prediction.

4.5 Algorithm Implementation for TOMB

The NEON IS transition system runs Docker containers containing R code to generate the csd_continuousDischargeUSGS table (similar infrastructure to Metzger, 2017 without the use of HDF5 file formats).

Within the Docker container used to generate csd_continuousDischargeUSGS, run on a single day of data:

1. USGS data are pulled for a time period (1-day) and location (USGS 02469761 TOMBIGBEE R AT COFFEEVILLE L&D NR COFFEEVILLE, AL.).
2. The most recent relationship between USGS data and NEON empirical measurements (published in the bat_dischargeRegressionUSGS table) is used to populate uncertainty bounds around the USGS measurements.
3. Data are formatted and units are converted.
4. Container outputs are written to the L4 data tables in the NEON database for publication to users.

5 DATA QUALITY

5.1 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[07] and AD[09] 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[08] for information on L1 quality flags), or report temporary hydrologic conditions at the discharge cross-section (information found in gag_fieldData:**dscTempHydroCond**, published in the expanded download package of DP1.20048.001). When gauge-water column height regressions are developed, NEON scientists review the quality of both the L1 gauge height and L0 pressure 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[06] and AD[07]). If L0 pressure 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).

5.2 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 3) 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.

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.

5.3 Regression Coefficients

In the expanded download package of this data product, NEON publishes the geo/bat_gaugeWaterColumnRegression table. This table allows users to view the gauge - water column height regression equation (m_{reg} and b_{reg} in Equation 2) and coefficient of determination (R^2) for each **regressionID** published. The regression coefficients can be used to assess the accuracy and precision of the csd_continuousDischarge:**estimatedStage** variable used as an input to BaM to model continuous discharge. See Section 3.8 of this document to learn how to link the data tables in this data product to the appropriate regression coefficients.

5.4 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.

5.5 Quality Flagging

For the quality flags in the csd_continuousDischarge table, see the descriptions in AD[14] and AD[15] 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[14]. 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).

6 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

Le Coz, J., C. Chaleon, L. Bonnifait, R. Le Boursicaud, B. Renard, F. Branger, J. Diribarne, M. Valente (2013) Bayesian analysis of rating curves and their uncertainties: the BaRatin method. *Houille Blanche-Revue Internationale De L Eau*, 6:31:41. DOI: 10.1051/lhb/2013048.

Metzger, S., D. Durden, C. Sturtevant, H. Luo, N. Pingintha-Durden, T. Sachs, A. Serafimovich, J. Hartmann, J. Li, K. Xu, A. Desai (2017) eddy4R: A community-extensible processing, analysis and modeling framework for eddy-covariance data based on R, Git, Docker and HDF5. *Geoscientific Model Development Discussions*, 1-26. DOI: 10.5194/gmd-2016-318.