NEON USER GUIDE TO CONTINUOUS DISCHARGE (DP4.00130.001)

PREPARED BY	ORGANIZATION	DATE
Kaelin M Cawley	AQU	01/25/2022
Nick Harrison	AQU	01/25/2022
Zachary Nickerson	AQU	01/25/2022
Bobby Hensley	AQU	01/25/2022

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.

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 4 data product, which is generated from Level 4 OS data, Level 0 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 (Level 0). Raw data that have been quality checked and simple metrics that emerge from the raw data are considered Level 1 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 the file, NEON Data Variables for Continuous Discharge (DP4.00130.001) (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 dischare (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, NEON Algorithm Theoretical Basis Document (ATBD): Surface Water Elevation (AD[05]); for gauge height, NEON User Guide to Gauge Height (DP1.20267.001) (AD[06]) and NEON Data Validations for Gauge Height (DP1.20048) (AD[07]); for wadeable stream morphology, NEON User Guide to Wadeable Stream Morphology (DP4.00131.001) (AD[08]); for bathymetric and morphological maps, NEON User Guide to Bathymetric and Morphological Maps (DP4.00132.001) (AD[09]); for stage-discharge rating curve, NEON User Guide to Stage-Discharge Rating Curves (DP1.00133.001) (AD[10]); for USGS streamflow data, Measurement and Computation of Streamflow Testing (AD[16]). 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 (DP1.20048)
AD[08]	NEON_streamMorphology_userGuide.pdf	NEON User Guide to Wadeable Stream Morphology (DP4.00131.001)
AD[09]	NEON_bathymetry_userGuide.pdf	NEON User Guide to Bathymetric and Morphological Maps (DP4.00132.001)
AD[10]	NEON_ratingCurve_userGuide.pdf	NEON User Guide to Stage-Discharge Rating Curves (DP1.00133.001)
AD[11]	NEON.DOC.000008	NEON Acronym List
AD[12]	NEON.DOC.000243	NEON Glossary of Terms
AD[13]	NEON.DOC.000927	NEON Calibration and Sensor Uncertainty Values
AD[14]	NEON.DOC.001113	Algorithm Theoretical Basis Document: Quality Flags and Quality Metrics for TIS Data Products
AD[15]	NEON.DOC.011081	Algorithm Theoretical Basis Document: QA/QC Plausibility Testing
AD[16]	https://doi.org/10.3133/wsp2175	Measurement and Computation of Streamflow Testing

2.2 Acronyms

Acronym	Definition
Q	Discharge
L0	Level Zero (Unprocessed) Data
L1	Level One (Processed) Data
L4	Level Four (Derived and Processed) Data
QAQC	Quality Assurance Quality Checking
S1	Aquatic Sensor Set One
S2	Aquatic Sensor Set Two
API	Application Programming Interface
AOS	Aquatic Observational System
min	Minute
wk	Week
m	Meter
kPa	Kilopascal
Pa	Pascal
kg	Kilogram
s	Second
BaM	Bayesian Modeling
BaRatinAGE	Bayesian Rating Curve Advanced Graphical Environment
GUI	Graphical User Interface
MCMC	Markov Chain Monte Carlo
USGS	United States Geological Survey

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 30-minute or hourly data.

3.1 Spatial Sampling Design

Continuous discharge is developed at the site level (except for Toolik Lake, where two stations will have continuous discharge produced at the inflow and outflow to the lake) using data collected at either sensor set #1 or sensor set #2, whichever is closer to the staff gauge in wadeable streams, at the near shore sensor set (AKA 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 taken 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.

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

3.2 Temporal Sampling Design

Level 0 pressure transducer is collected at a 1 min resolution. The 1 min resolution is retained in the L4 continuous discharge data product in csd_continuousDischarge.

The gauge heights that are published in the gauge-pressure relationship data 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 in the csd_continuousDischargeUSGS table are published at approximately 1 hour resolution for TOMB only.

3.3 Variables Reported

All data and geolocation variables used as inputs for continuous discharge are listed in Table 1, Table 2, and Table 3. For more information on calibration and validation assessment of sensors repeatability,

Table 1: List of input data for Continuous Discharge. Note: The sdrc_gaugePressureRelationship table is processed as part of the Stream-discharge rating curves (DP4.00133.001) data product, but is published with the Continuous discharge (DP4.00130.001) data product; therefore, the table is described in this document and not described in NEON User Guide to Stage-Discharge Rating Curves (DP1.00133.001) AD[10].

Data Source	DPID	table
L1 Gauge Height	DP1.20267	gag_fieldData
L4 Stream Morphology Map	DP4.00131.001	geo_controlInfo
L4 Stream Morphology Map	DP4.00131.001	geo_priorParameters
L4 Stream Morphology Map	DP4.00131.001	geo_curveIdentification
L4 Stream Morphology Map	DP4.00131.001	geo_gaugeWaterColumnRegression
L4 Bathymetric and Morphological Maps	DP4.00132.001	bat_controlInfo
L4 Bathymetric and Morphological Maps	DP4.00132.001	bat_priorParameters
L4 Bathymetric and Morphological Maps	DP4.00132.001	bat_curveIdentification
L4 Bathymetric and Morphological Maps	DP4.00132.001	bat_gaugeWaterColumnRegression
L4 Bathymetric and Morphological Maps	DP4.00132.001	bat_dischargeRegressionUSGS
L4 Stage-Discharge Rating Curves	DP4.00133.001	sdrc_stageDischargeCurveInfo
L4 Stage-Discharge Rating Curves	DP4.00133.001	sdrc_sampledParameters
L4 Stage-Discharge Rating Curves	DP4.00133.001	sdrc_gaugeDischargeMeas
L4 Stage-Discharge Rating Curves	DP4.00133.001	sdrc_gaugePressureRelationship

see AD[13]. All variables reported in the published data (L4 data) are also provided separately in the file, NEON Data Variables for Continuous Discharge (DP4.00130.001) (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 is published from two stations.

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.

Table 2: 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 3: List of geolocation inputs for Continuous Discharge.

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

3.6 Associated Data Streams

The data from this L4 data product shares 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.

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**.

For wadeable streams, the data from this L4 data product are derived from a L4 data product: Stream Morphology Map (DP4.00131.001). For rivers and Toolik Lake locations, the data from this L4 data product are derived from a L4 data product: Bathymetric and Morphological Maps (DP4.00132.001). Data in this L4 data product are linked to the Stream Morphology Map and Bathymetric and Morphological Maps data products by the same fields in both data products:

- csd_continuousDischarge can be linked to geo/bat_curveIdentification by **curveID**.
- csd_continuousDischarge can be linked to hydraulic controls (priors) data by first linking to geo/bat_curveIdentification by **curveID** then linking geo/bat_curveIdentification:**controlSurveyEndDateTime** to geo/bat_controlInfo:**endDate** and geo/bat_priorParameters:**endDate**.

Note: These tables are available for download as part of expanded download packages in the Stage-Discharge Rating Curves (DP4.00133.001) data product.

- csd_continuousDischarge and sdrc_gaugePressureRelationship can be linked to geo/bat_gaugeWaterColumnRegression by **regressionID**.
- csd_continuousDischargeUSGS can be linked to bat_dischargeRegressionUSGS by **usgsDischargeRegID**.

Note: These tables are available for download as part of expanded download packages in the Continuous Discharge (DP4.00130.001) data product.

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 by **curveID**.

3.7 Product Instances

The NEON Observatory contains 24 wadeable streams, 3 large rivers, and 1 lake site containing measurable inflow and outflow.

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 hour) 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 for a site in the gag_fieldData table of the Gauge Height (DP1.20267.001) data product for a given water year (October 1st to September 30th). The values may differ between the two tables because a record from gag_fieldData will not be processed into the sdrc_gaugePressureRelationship table is there are no L0 pressure transducer measurements within 20 min of the gag_fieldData record's **collectDate**. 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 as an 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 L4 Stage-discharge rating curves (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 after infrastructure changes location. Offsets are calculated using information derived from the NEON geolocation database (Table 3). 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 (vertical correction needed in order for the location to be relatable to previous locations; m), (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.

One record is created in csd_continuousDischarge per min regardless of whether the pressure transducer was producing data. For timestamps when the pressure transducer is not streaming, records in csd_continuousDischarge are produced containing the timestamp, applicable flags, and metadata, but no stream or flow series data. Some NEON sites are seasonal due to climate or logistical constraints. Records will not be processed and published at seasonal sites when infrastructure was intentionally deactivated or removed from the site.

The regression equations used to calculate stage in both sdrc_gaugePressureRelationship and csd_continuousDischarge tables are published in the geo_gaugeWaterColumnRegression table for stream sites and bat_gaugeWaterColumnRegression for river sites and lake locations. These tables are available as part of the expanded download packages for this data product. Information on the regression table used to develop continuous discharge can be found in AD[08] for stream sites and AD[09] for river sites and lake locations.

sdrc_gaugePressureRelationship.csv - > One record expected per staff gauge reading collected 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.

csd_continuousDischarge.csv - > One record expected per min regardless of available L0 pressure data except at seasonal sites, which will contain gaps in data.

geo_gaugeWaterColumnRegression - > One record expected per unique gauge height - water column height linear regression developed at a stream site. Every 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 site or lake inflow and outflow locations. Every 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 BaM. At D08 TOMB (located on the Tombigbee River in southwest Alabama) a downstream lock and and dam system operated by the US Army Corp of Engineers 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 - > Once 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 Publication Timeline and Annual Reprocessing

When a new water year starts (October 1st), csd_continuousDischarge and csd_continuousDischargeUSGS data will be published provisionally using input data (Table 1) from the most recent previous water year following the publication schedule described in Section 3.2 of this document. At the end of a water year, NEON scientists will develop, review, and publish input data (see Table 1 for list of inputs) for each site or location. Upon publication of the previous water year's input data, csd_continuousDischarge and csd_continuousDischargeUSGS data will be reprocessed and re-published.

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 height of the water column, 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).

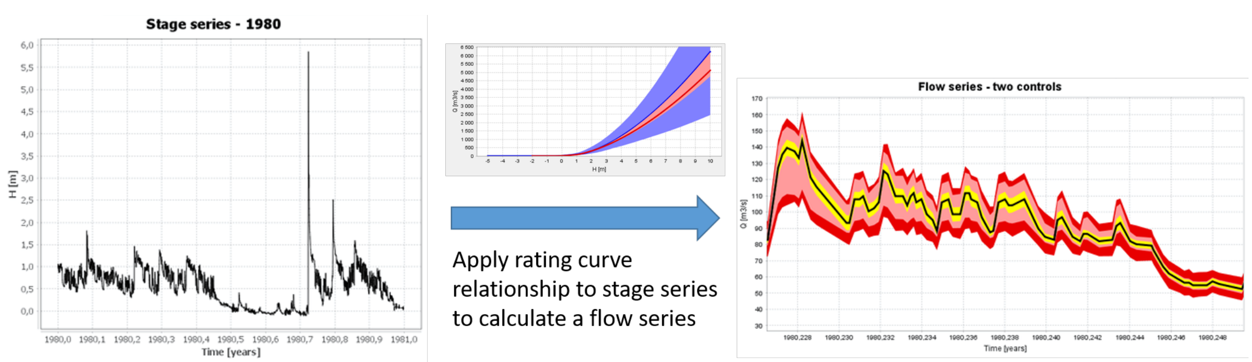

Figure 1: Example stage series and flow series output. Pictures taken from "Stage series" and "Flow series" documentation for BaRatinAGE (Le Coz et al., 2013; Le Coz et al., 2014).

4.2 Theory of Algorithm for All Sites Except for TOMB

Pressure transducers are installed at all NEON sites and are used to develop a continuous discharge record at wadeable streams and large rivers. For more information on the NEON pressure transducers, see AD[05]. Calibrated pressure can be used to estimate stage by first converting to water column height and applying an offset from the site survey (Equation 1) and then by applying the appropriate water column height-gauge height linear regression to the offset estimated water column height (Equation 2).

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

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$\cdot$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 using the stage-discharge rating curve (DP4.20133.001). The NEON Stage-discharge rating curves and Continuous discharge stage series 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 available freely with an individual license by sending an email to: baratin.dev@lists.irstea.fr.

The rating curve relies on a “prior” rating curve that is developed for 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).

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

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 a Docker container containing R code to estimate stage and flow series data at 1 min resolution (similar infrastructure to Metzger, 2017 without the use of HDF5 file formats).

Within the Docker container:

1. L0 pressure data for the site and day of continuous discharge is queried from the NEON database.
2. Calibration factors for the L0 pressure data are queried from the NEON database.
3. Geolocation information for the staff gauge location (SITE.AOS.gauge) and level TROLL (CFGLOC) is queried from the NEON database.
4. Calibration factors and geolocation data are used to convert raw pressure transducer to estimated stage.
5. L4 gauge-pressure relationship data are queried from the NEON OS database.
6. Calibration uncertainty and gauge-pressure relationship uncertainty are combined to estimate uncertainty in calculated stage.
7. L4 stage-discharge relationship and hydraulic controls data are queried from the NEON OS database to use as inputs to BaM.
8. Configuration files for BaM are created.
9. BaM executable is run.
10. BaM outputs are written to the L4 data tables in the NEON database for publication to users.

4.4 Theory of Algorithm for TOMB

About 14 km downstream of the NEON TOMB discharge cross-section is a lock and dam structure operated by the Army Corps of Engineers and USGS. This structure aims to maintain a stable stage level to support navigation of the channel across a broad range of discharge. Thus, the stage-discharge relationship at the NEON site is heavily impacted by the opening and closing of gates that are part of this downstream structure and is not valid for inputs to the BaM model. However, a relationship between the USGS discharge data and NEON empirical measurements were 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.

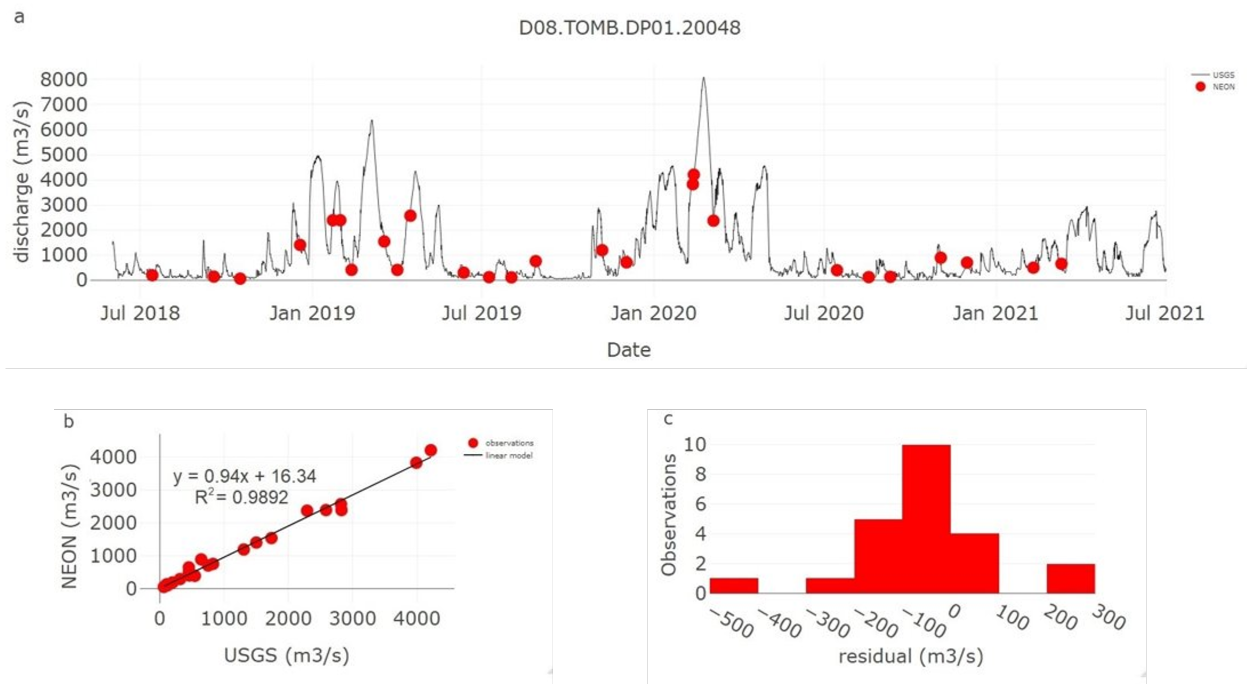

Figure 2: Top panel (a) shows the USGS flow timeseries with red dots for NEON empirical measurements, bottom panel (b) shows a linear regression and equation between USGS data at the time of NEON empirical measurements, and bottom panel (c) shows the distribution of residuals from the regression relationship.

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. These data are then regressed against empirical discharge measurements (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 a Docker container containing R code to estimate flow at the NEON discharge cross-section based on the flow at a downstream USGS gauge station (similar infrastructure to Metzger, 2017 without the use of HDF5 file formats).

Within the Docker container:

1. USGS data is 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 is 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 is used as an input for this data product) within a mobile data entry application or web user interface. See NEON User Guide to Gauge Height (DP1.20267.001) (AD[06]) and NEON Data Validations for Gauge Height (DP1.20048) (AD[07]) for more details.

5.2 Data Revision

All data are provisional until a numbered version is released; the first release of a static version of NEON data, annotated with a globally unique identifier, is planned to take place in 2020. During the provisional period, QAQC 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 Change Log section of the data product landing page has the most up-to-date information (as opposed to the readme in the download, which is current only to the date of data publication) and contains a history of major known errors and revisions.

This data product was first included in RELEASE-2022.

5.3 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 NEON data downloads is expanded uncertainty, i.e. multiplied by a factor of 1.96 to cover two standard deviations. When using the BaRatin GUI tool the uncertainty should be represented the same was as NEON publishes it. For the BaM executable, though, uncertainty is represented as one standard deviation. So, the NEON data should be divided by a factor of 1.96 before writing out data and configurations.

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

5.4 Quality Flagging

This data product contains a **dataQF** field in each data record that is a quality flag for known errors applying to the record in the csd_pressureGaugeRelationship table. Please see below for an explanation of **dataQF** codes specific to this product.

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.

fieldName	value	definition
dataQF	legacyData	Data recorded using a paper-based workflow that did not implement the full suite of quality control features associated with the interactive digital workflow

6 REFERENCES

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.