ALGORITHM THEORETICAL BASIS DOCUMENT (ATBD): STAGE-DISCHARGE RATING CURVES & CONTINUOUS DISCHARGE

PREPARED BY	ORGANIZATION	DATE
Bobby Hensley	AQU	05/14/2025
Nick Harrison	AQU	05/14/2025
Zach Nickerson	AQU	05/14/2025

APPROVALS	ORGANIZATION	APPROVAL DATE
Kate Thibault	SCI	05/14/2025

RELEASED BY	ORGANIZATION	RELEASE DATE
Tanisha Waters	CM	05/14/2025

See configuration management system for approval history.

The National Ecological Observatory Network is a project solely funded by the National Science Foundation and managed under cooperative agreement by Battelle. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Change Record

REVISION	DATE	ECO #	DESCRIPTION OF CHANGE
A	05/14/2025	ECO-07137	Initial release

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

1 DESCRIPTION

Contained in this document are details concerning the process for creating the NEON L4 Stage-discharge rating curves and L4 Continuous discharge data products.

1.1 Purpose

This document details the algorithms used for deriving NEON L4 Stage-discharge rating curves and L4 Continuous discharge from L4 and L1 OS data, L1 IS data, geolocation data, and associated metadata. In the NEON data products framework, raw data collected in the field (e.g., surface water pressure collected by a pressure transducer sensor) are considered L0. Data that have calibration data applied, undergone basic conversion of units (e.g., pressure to water column height), and have undergone basic quality checks to identify (but not remove) erroneous data are considered L1. Data products derived from multiple lower-level data products are considered L4. This document 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

L4 Stage-discharge rating curves are derived from the L1 Discharge field collection and Level 4 Stream morphology map data products. The respective Protocol and Procedure documents for Discharge field collection (AD[01], AD[02], AD[03], AD[04]) and Stream morphology map (AD[05]) describe the corresponding instruments and collection procedures for these dependencies. This document details how these dependencies are used to generate the L4 Stage-Discharge rating curves data product.

L4 Continuous discharge is derived from the L4 Stage-discharge rating curves and L1 Elevation of surface water data products. The L1 Elevation of surface water ATBD (AD[06]) describes the corresponding instruments, calibration procedures, uncertainty, and error handling for this data product. This document details how these dependencies are used to generate L4 Continuous discharge.

2 RELATED DOCUMENTS, ACRONYMS AND VARIABLE NOMENCLATURE

2.1 Applicable Documents

AD[01]	NEON.DOC.001085	NEON AOS Protocol and Procedure: DSC – Stream Discharge
AD[02]	NEON.DOC.005277	NEON AOS Protocol and Procedure: GAG – Aquatic Staff Gauge Measurement Readings
AD[03]	NEON.DOC.005388	NEON Standard Operating Procedure: DCS – Configurations, Settings and Collection Methods for Stream Discharge Measurements using the Flowmeter Method
AD[04]	NEON.DOC.005389	NEON Standard Operating Procedure: DCS – Configurations, Settings and Collection Methods for Stream Discharge Measurements using the ADCP Method
AD[05]	NEON.DOC.003162	NEON AOS Protocol and Procedure: Wadeable Stream Morphology
AD[06]	NEON.DOC.001198	NEON Algorithm Theoretical Basis Document (ATBD): Surface Water Temperature, Elevation and Specific Conductance
AD[07]	NEON.DP4.00133.001 _variables.csv	NEON Data Variables for Stage-discharge rating curves
AD[08]	NEON.DP4.00130.001 _variables.csv	NEON Data Variables for Continuous discharge
AD[09]	NEON.DOC.001113	Quality Flags and Quality Metrics for TIS Data Products

2.2 Reference Documents

RD[01]	Le Coz, J., B. Renard, L. Bonnifait, F. Branger and R. 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.jhydrol.2013.11.016
RD[02]	Turnipseed, D.P., and Sauer, V.B., 2010, Discharge measurements at gaging stations: U.S. Geological Survey Techniques and Methods book 3, chap. A8, 87 p.
RD[03]	Sauer, V.B., 2002, Standards for the Analysis and Processing of Surface-Water Data and Information Using Electronic Methods: U.S. Geological Survey Water-Resources Investigations Report 01–4044, 91 p
RD[04]	De Cicco, L.A., Hirsch, R.M., Lorenz, D., Watkins, W.D., Johnson, M., 2025, dataRetrieval: R packages for discovering and retrieving water data available from Federal hydrologic web services, v.2.7.18. doi:10.5066/P9X4L3GE

2.3 Acronyms

Acronym	Explanation
A	Channel cross-sectional area
a	Hydraulic coefficient
ADCP	Acoustic Doppler Current Profiler
ATBD	Algorithm Theoretical Basis Document
b	Hydraulic activation stage
c	Hydraulic exponent

DP	Data Product
h	Stage
IS	Instrument System
K_S	Strickler coefficient
κ	Hydraulic transition stage
L0	Level 0
L1	Level 1
L4	Level 4
OS	Observational System
Q	Discharge
QFSciRvw	Science Review Quality Flag
QA/QC	Quality Assurance Quality Control
R	Channel hydraulic radius
S	Channel slope
USGS	United States Geological Survey
WY	Water year (October 1 – September 30)

2.4 Definitions

Term	**Definition**
Cleaning	The process of identifying and applying quality flags to erroneous or suspect periods of data to be corrected using a gap-filling method.
Discharge	The volume of water flowing through a stream or river per unit time.
Gap-filling	The process of filling a data gap through the use of a specified method. Gaps can result from missing data or erroneous data set to NA in the cleaning process.
Gauging	An empirical measurement of stage or discharge.
Hydraulic controls	Hydraulic controls define the physical channel geometry and determine how changes in water level translate into changes in streamflow.
Pressure transducer	An electronic sensor mounted on the streambed that is used to continuously measure water pressure above the sensor.
Stage	The height of the free water surface relative to a fixed datum, usually a staff gauge.
Staff Gauge	A calibrated ruler installed in a body of water used to measure stage.
Stage-discharge rating curve	The relationship between stage and discharge, developed over time using empirical data.
Survey	Measurements that relate the channel topography along the cross-section where discharge measurements are collected to the elevation of the staff gauge. Survey data characterize the physical dimensions of each of the hydraulic controls used in the prior model during the development of the stage-discharge rating curve.

3 DATA PRODUCT DESCRIPTION

3.1 Variables Reported

The variables reported in the Stage-discharge rating curves L4 data product provided by the algorithms documented in this ATBD are provided in the accompanying file NEON Data Variables for Stage-discharge rating curves (AD[07]).

The variables reported in the Continuous discharge data product provided by the algorithms documented in this ATBD are provided in the accompanying file NEON Data Variables for Continuous discharge (AD[08]).

3.2 Input Dependencies

Table 1 and **Table 2** list other NEON data products used to formulate L4 Stage-discharge rating curves and L4 Continuous discharge tables.

Table 1. Data products used as inputs for the Stage discharge rating curves data product.

Data Product Name	Data Product Number	Input Table Name	Stage-discharge rating curves Table Name
Discharge field collection	DP1.20048.001	dsc_fieldData	sdrc_stageDischargeMeas
Stream morphology map	DP4.00131.001	geo_surveyPoints	sdrc_controlInfo sdrc_controlType sdrc_priorParameters

Table 2. Data products used as inputs for the Continuous discharge data product.

Data Product Name	Data Product Number	Input Table Name	Continuous discharge Table Name
Discharge field collection	DP1.20048.001	dsc_fieldData	csd_gaugeWaterColumnHeightRegression csd_dischargeRegressionUSGS sdrc_gaugePressureRelationship
Gauge height	DP1.20267.001	gag_fieldData	csd_gaugeWaterColumnHeightRegression sdrc_gaugePressureRelationship
Elevation of surface water	DP1.20016.001	EOS_1_min	csd_gaugeWaterColumnHeightRegression sdrc_gaugePressureRelationship csd_continuousDischarge
Stage Discharge Rating Curve	DP4.00133.001	sdrc_controlInfo sdrc_priorParameters sdrc_curveIdentification sdrc_stageDischargeCurveInfo	csd_continuousDischarge

Data Product Name	Data Product Number	Input Table Name	Continuous discharge Table Name
		sdrc_sampledParameters sdrc_gaugeDischargeMeas	

3.3 Temporal Resolution and Extent

Stage-discharge rating curves typically span one water year (1 October – 30 September). However, the relationship between stage and discharge is dynamic and can be altered when changes in channel morphology (e.g., scouring or deposition of sediment along the streambed) occur. Additionally, if the staff gauge is damaged or displaced during a high-flow event, its elevation may not be able to be related pre-and post-disturbance. In these cases, it may be necessary to publish multiple rating curves within a water year or to extend them into prior or subsequent water years.

For Continuous discharge, the gauge-water column height regression models used to convert L1 water column height data into L4 stage typically span one water year, though, as in the case with rating curves, it may be necessary to publish multiple gauge-water column height regression models within a water year or extend them into prior or subsequent water years.

The continuous discharge timeseries is published at 1 minute resolution, except for at the TOMB site, which uses a different processing pipeline (see **Section 4.2.8**) and is published at an approximate 1 hour resolution.

3.4 Spatial Resolution and Extent

Stage-discharge rating curves and Continuous discharge data are published for all 24 NEON wadeable stream sites, 2 of 3 river sites (D03 FLNT – Flint River and D08 BLWA – Black Warrior River), and one lake site (D18 TOOK - Toolik Lake) that contains an inflow and outflow where discharge can be measured. At one river site (D08 TOMB - Tombigbee River), only Continuous discharge data are published (see **Section 4.2.8**). Data are published for one location at each site, with the exception of Toolik Lake where data are published for both the inflow and outflow locations.

4 SCIENTIFIC CONTEXT

Discharge is the volume of water flowing through a stream or river per unit time. It is a function of the channel cross-sectional area and water velocity. A cleaned and gap-filled record of stream discharge is critical for understanding the evolution of channel morphology and benthic habitat and for estimating the flux of nutrients and sediments.

4.1 Theory of Measurement

Because discharge is difficult to measure continuously, it is typically estimated using a stage-discharge rating curve that relates discharge to stage. Rating curves are developed by collecting numerous discrete measurements of discharge (oftentimes spanning multiple water years), while noting the accompanying stage on a nearby staff gauge (AD[01]). Pressure transducer sensors are used to collect a continuous record of water column height (AD[06]). The relationship between water column height and stage is used to generate continuous estimates of stage. Coefficients derived from the stage-discharge rating curve are applied to the continuous stage data to generate a continuous record of discharge.

4.2 Theory of Algorithm

Accurate estimation of continuous discharge is dependent on the quality of the inputs. The continuous record of water column height must be reviewed, and any erroneous, erratic, or shifted data must be either corrected or omitted from the timeseries, and data gaps must be filled to the extent possible. Likewise, empirical discharge and staff gauge measurements must be reviewed for potential outliers and removed from the gauge-water column height regression model or the stage-discharge rating curve. Finally, the continuous discharge record must be reviewed, with any erroneous, erratic, or shifted data corrected and any data gaps filled to the extent possible.

4.2.1 Rating Curve Development

The Stage-discharge rating curves data product is developed using the BaM! Bayesian model (RD[01]). The executable and/or GUI is freely available with individual licenses by sending an email to: baratin.dev@lists.irstea.fr. The advantages of using a Bayesian model approach over simply fitting a regression to the empirical observations are twofold: First, knowledge of the stage-discharge relationship based on the physical channel geometry can be used for the prior distribution, better constraining the rating curve in regions outside the range of empirical measurements. Second, stage-discharge relationships are heteroscedastic in nature, and the Bayesian model both considers uncertainties associated with individual measurements and incorporates uncertainty from a wide range of inputs.

4.2.2 Hydraulic Control Delineation

Hydraulic controls are identified using data collected during topographic surveys of the discharge cross-section (survey data are provided in the Stream morphology map data product, AD[05]). The staff gauge is also mapped during the survey so that channel elevation can be related to stage. Hydraulic controls are identified within the cross-section profile and the activation stage of each control is determined. The primary function of the hydraulic controls is to define the physical geometry of the channel and inform the prior knowledge within the Bayesian model. This is developed using the Manning's equation:

$$Q = AK_S S^{1/2} R^{2/3} \tag{1}$$

where Q is discharge, A is the channel cross-sectional area, K_S is the Strickler coefficient, S is the channel slope and R is the hydraulic radius. This equation can be simplified to:

$$Q = a(h-b)^c \tag{2}$$

where h is stage, a is a coefficient related to the characteristics of the channel (e.x. $K_s \times \sqrt{S} \times width$ for a wide rectangular channel), b is an activation stage, or point of zero flow, and c is an exponent related to the channel geometry (e.g., 5/3 for a wide rectangular channel). The a, b, and c coefficients contained in Equation 2 represent the characteristics of each hydraulic control identified within the channel. New hydraulic controls are created for a site each time there is a new survey of the discharge cross-section, and are recorded in the **sdrc_controlInfo**, **sdrc_controlType** and **sdrc_priorParamters** tables in the Stage-discharge rating curves data product.

The model combines the hydraulic controls piecewise to create the "prior" rating curve:

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

Where the function $1_{[\kappa_{r-1};\kappa_r]}$ defines the transition between stage ranges and is equal to 1 if $\kappa_{r-1} \leq h \leq \kappa_r$ and 0 otherwise. The matrix $M(r,j)$ contains values of either 0 or 1 depending on whether control j is active in stage range r.

$$M(r,j) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \tag{4}$$

NEON aquatic sites typically utilize 3 hydraulic controls (**Figure 1**). In **Equation 4**, control 1 (a riffle) is the only active control in range 1 (**Figure 1**). Control 2 (**Figure 1,** red box) becomes active in range 2 and drowns control 1. In range 3, control 3 (the floodplain, **Figure 1,** green box) becomes active, but control 2 also remains active.

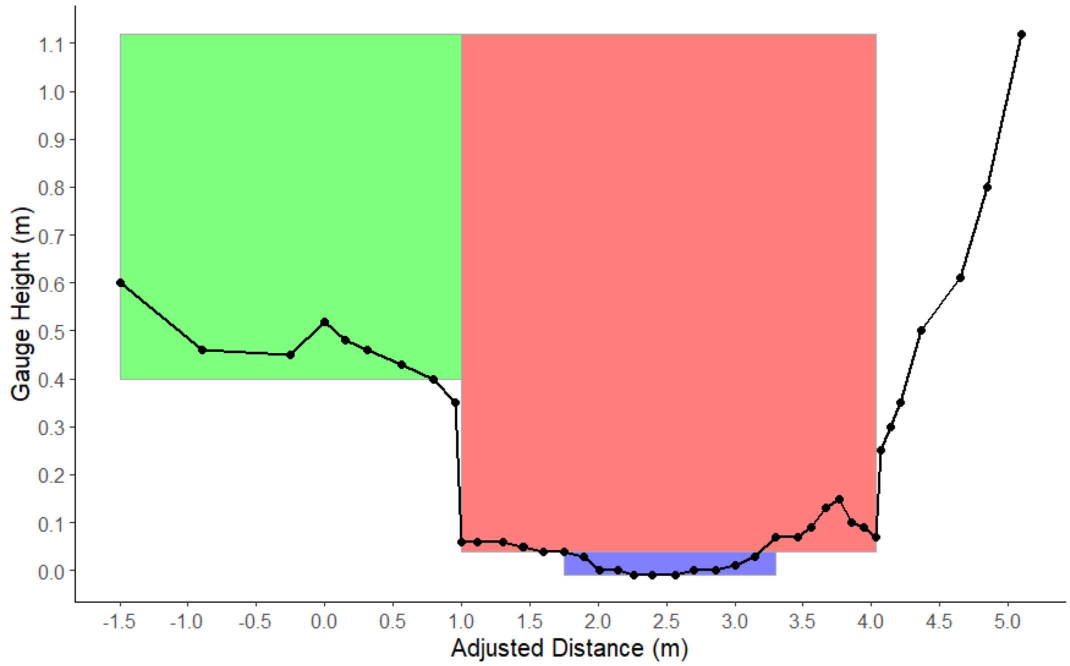

Figure 1. Hydraulic controls overlaying a survey of the channel cross-section. Each dot represents a survey point collected by the total station. The blue box represents the low-flow riffle control, the red box represents the main channel control, and the green box represents a bank control.

4.2.3 Empirical Gaugings

Paired field measurements of stage and discharge, hereafter referred to as "gaugings," are collected for rating curve development. All **finalDischarge** and **streamStage** measurements in the **dsc_fieldData** table of the L1 Discharge field collection data product for a given water year are reviewed and used to create a rating curve for that water year. Potential outliers are identified and not included in the rating curve. Gaugings collected at moderate to high streamflow levels during previous water years may be included in subsequent rating curves to better constrain discharge estimates at these stage levels. All the gaugings included in a particular rating curve (indicated by the **curveID** field) are listed in the **sdrc_gaugeDischargeMeas** table, and the start and end dates of each rating curve are listed in the **sdrc_stageDischargeCurveInfo** table in the Stage-discharge rating curves data product.

4.2.4 Using the Bayesian Model to Develop the Rating Curve

Using the Bayesian model, the posterior probability density function of the rating curve parameters is estimated using the prior rating curve and the gauging records using Markov Chain Monte Carlo (MCMC) simulation, which creates an ensemble of 500 model realizations. The maximum-posterior (mode of the distribution) for each of the model parameters are included in the **sdrc_posteriorParameters** table, and the residuals (differences between observations and model predictions for each gauging) are included in the **sdrc_resultsResiduals** table in the Stage-discharge rating curves data product. From these, a maximum-posterior rating curve (equal to the mode of the posterior distribution) can be determined, along with its parametric and remnant uncertainty intervals (**Figure 2**).

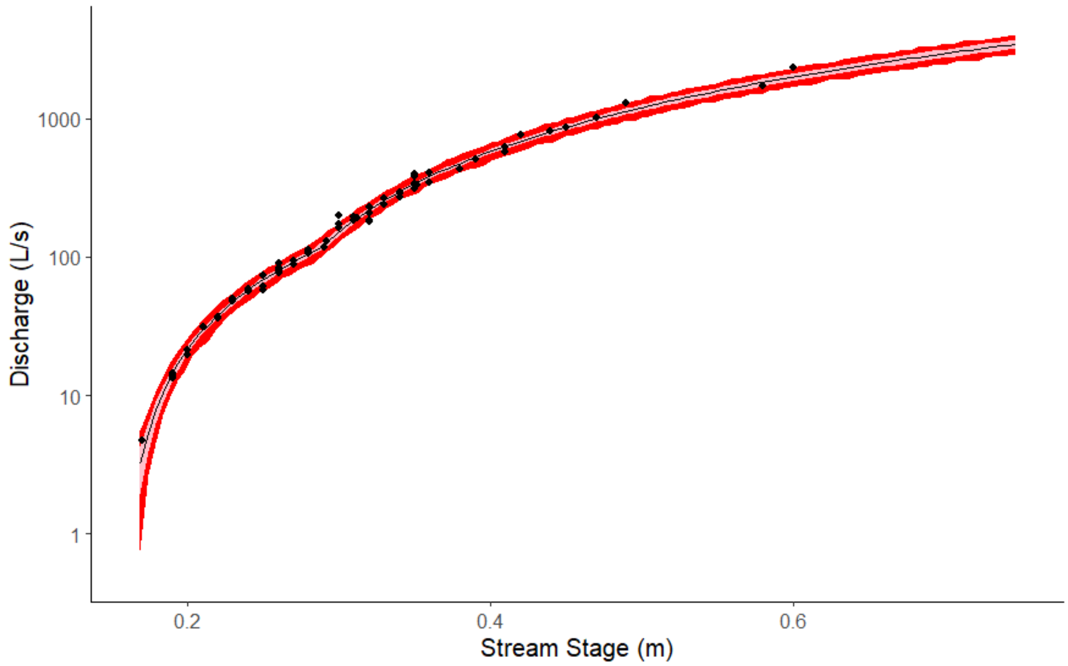

Figure 2. Stage-discharge rating curve showing empirical gaugings (black points), discharge (solid line) and uncertainty (red and pink bands).

4.2.5 Continuous Discharge Development

Continuous discharge is modelled using the same Bayesian model used in the development of the stage-discharge rating curve (RD[01]). This allows for the rating curve and its associated uncertainties to be applied to the continuous stage timeseries, generating a continuous estimate of discharge with uncertainty bands.

4.2.6 Equivalent Stage

To model continuous discharge, there must first be a continuous record of stream stage. A continuous record of Water column height, published as **surfacewaterColumnHeight** at a 1 min resolution in the **EOS_1_min** table of the Elevation of surface water data product, is derived from a pressure transducer sensor (AD[06]) (**Figure 3**). Staff gauge measurements are derived from **dsc_fieldData** in the Discharge field collection data product (AD[01]) and **gag_fieldData** in the Gauge height data product (AD[02]).

A linear model is developed that regresses water column height data against empirical staff gauge measurements (**Figure 3**). When measurements are published in both **dsc_fieldData** and **gag_fieldData** for the same day, the staff gauge measurements from **dsc_fieldData** are chosen for use in the linear regression due to that measurement's direct relationship with rating curve development. Potential outliers are identified and removed from the model as needed. The start date, end date, slope, intercept, and coefficient of determination (R^2) for each regression model are reported in the **csd_gaugeWaterColumnRegression** table in the Continuous discharge data product. The regression equation is then applied to water column height data to calculate **equivalentStage** in the **csd_continuousDischarge** table of the Continuous discharge data product. The **stageUnc** value represents the sum of nonsystematic uncertainty (derived from calibration coefficients) and systematic uncertainty (the mean difference between all pairs of empirical measurements and predicted stage estimates). Because the elevation of the staff gauge and/or the pressure transducer sensor may change over time, there may be instances where multiple stage regressions are applied within a single water year.

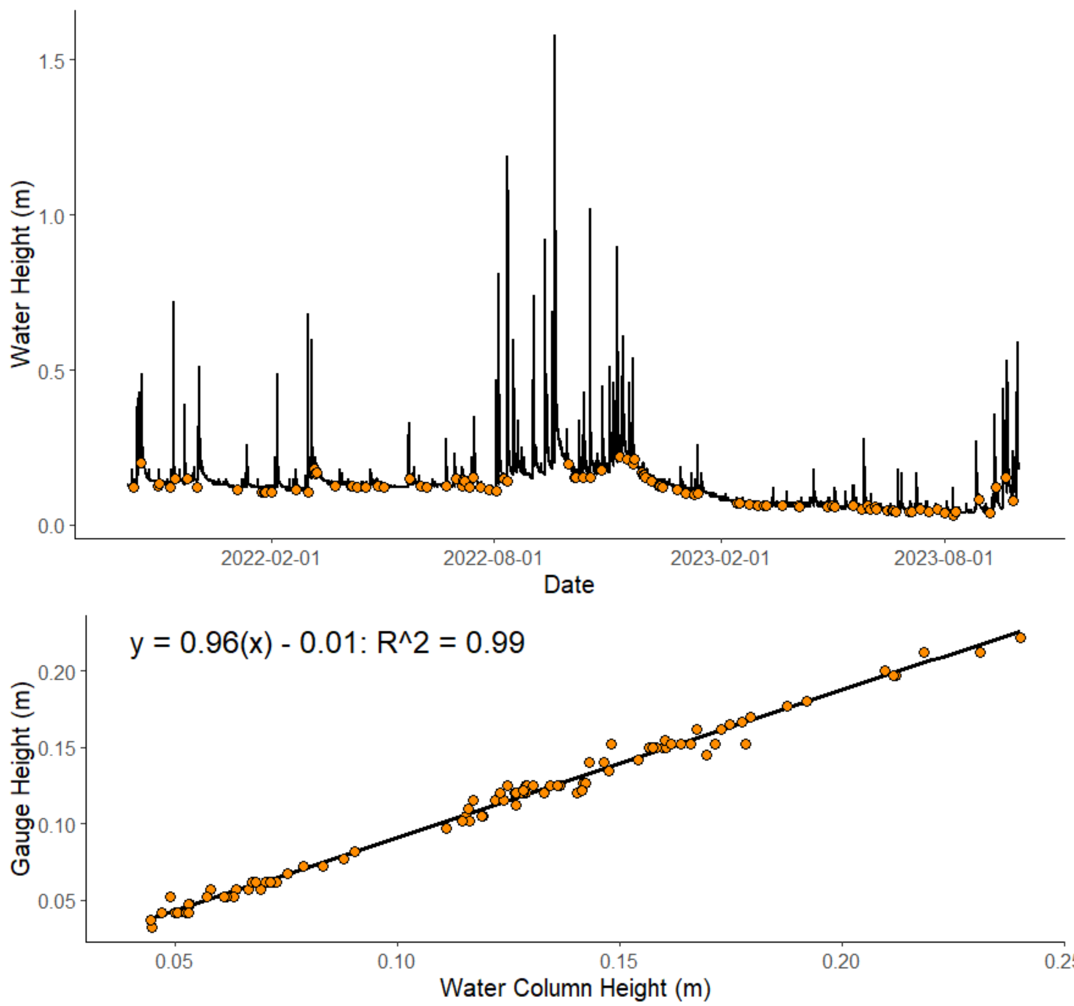

Figure 3. Above: Water column height timeseries (black line) with gauge height measurements (orange dots). Below: Linear model developed from the regression between water column height and gauge height measurements.

4.2.7 Using the Bayesian Model to Develop Continuous Discharge

The stage timeseries is run through the Bayesian model along with data from the prior and posterior rating curve models to estimate continuous discharge. The **maxpostDischarge** values in the **csd_continuousDischarge** table **5.1** are generated by applying the maximum-posterior rating curve coefficients to the **equivalentStage** timeseries. **maxpostDischarge** values may have had corrections (see **Section 5.1**) applied to the **waterColumnHeight** (and resulting **equivalentStage)** used as model inputs, but these records are raw model outputs that have not had any additional discharge corrections applied. Parametric uncertainty (derived from the uncertainty of the model priors) and remnant uncertainty (derived from how the model fits the observations) are also included.

4.2.8 Continuous Discharge Development at the Tombigbee River Site

The NEON site on the Tombigbee River (TOMB) is located just upstream of the Coffeeville Dam in Coffeeville, AL. The dam produces a relatively constant upstream water surface elevation and, as such, a stage-discharge rating curve cannot be developed to accurately estimate discharge at the NEON site. The Stage-discharge rating curves data product is thus not published at TOMB.

The USGS publishes continuous measurements of discharge released through the Coffeeville Dam (site ID 02469761). These data are pulled using the USGS dataRetrieval R package (RD[04]) and published as **usgsDischarge** values in the **csd_continuousDischargeUSGS** table in the Continuous discharge data product. To quantify uncertainty in how the USGS continuous discharge estimates translate to continuous discharge at the NEON site several kilometers away, a regression is fit between the empirical **finalDischarge** values in the **dsc_fieldData** table in the Discharge field collection data product (AD[01]) and the corresponding USGS discharge measurements. The start date, end date, slope, intercept, and coefficient of determination (R^2) for each regression are published in the **csd_dischargeRegressionUSGS** table in the Continuous discharge data product. Standard error values associated with the regression are also included in the **csd_dischargeRegressionUSGS** table and are used to generate the uncertainty ranges included in the **csd_continuousDischargeUSGS** table. The **csd_continuousDischargeUSGS** and **csd_dischargeRegressionUSGS** tables are only generated for TOMB. No cleaning or gap-filling is performed on the TOMB data.

5 ALGORITHEM IMPLEMENTATION

5.1 Data Correction: Reviewing, Cleaning and Gap-Filling the Timeseries

To provide the most accurate and complete record of continuous discharge, it is necessary to review, clean, and gap-fill the timeseries, a process referred to henceforth as “data correction.” Ideally, all data correction occurs at the water column height level before processing in the Bayesian model to compute continuous discharge. There are instances, however, where data correction can only be applied at the continuous discharge level. Examples of such instances include cases of in-channel ice formation (where water column height data should not be corrected but continuous discharge data should) or when the USGS-NEON discharge model is the only gap-filling method available for correction (where water column height data cannot be corrected but continuous discharge data can).

First, the **surfacewaterColumnHeight** timeseries published in the **EOS_1_min** table of the Elevation of surface water data product is reviewed and cleaned (**Figure 4**). During the cleaning process, a value of 1 is applied to the **surfacewaterColumnHeightFinalQFSciRvw** field for all records of **surfacewaterColumnHeight** that require correction. Gap-filling methods (described in **Sections 5.1.1-5.1.4**) are then applied to each of these records, including periods that require shifts (see **Section 5.1.2**), and any other gaps in the timeseries. Once gap-filling is complete, the corrected water column height record is published as **waterColumnHeightCorrected** in the **csd_continuousDischarge** table of the Continuous discharge data product.

Next the Bayesian model is run to produce the uncorrected **maxpostDischarge** timeseries in the **csd_continuousDischarge** table (see **Section 4.2.4**); and this record is reviewed, cleaned, and gap-filled using the same methodology described above for **surfacewaterColumnHeight**. Periods of the **maxpostDischarge** timeseries that require correction are identified with values of 1 in the **dischargeFinalQFSciRvw** field. The corrected discharge record is then published as **continuousDischarge** in the **csd_continuousDischarge** table.

The fields **waterColumnHeightCorrectionApplied** and **dischargeCorrectionApplied** in the **csd_continuousDischarge** table indicate whether individual records have been reviewed and if gap-filling methods have been applied (**Table 3**). Additionally, the **csd_dataGapToFillMethodMapping** table is produced to provide information on the specific gap-filling method (**gapFillMethod** field) applied to each record that was flagged for correction (**Table 4**).

Data correction for the entire water year will typically not occur until after the water year has ended, with expected publication occurring within six months of the end of the water year. For example, data collected from October 1, 2023 to September 30, 2024 would be corrected and available for use by March 1, 2025.In the meantime, provisional, uncorrected, **maxpostDischarge** data (which is generated using the previous year’s rating curve and has not been cleaned or gap-filled) will be published at the end of each month.

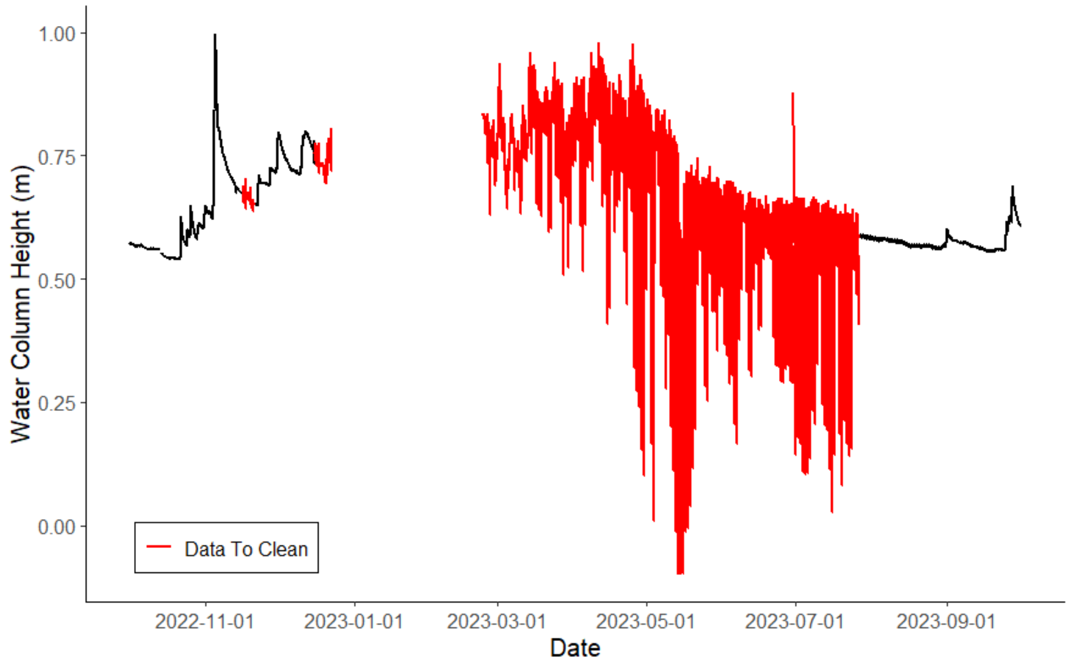

Figure 4. Example of cleaning water column height data. Data shown in red have been flagged and set to NA, then potentially corrected (along with any other gaps in the timeseries) using gap-filling methods.

Table 3. Values for waterColumnHeightCorrectionApplied and dischargeCorrectionApplied fields inform users whether individual records have been reviewed and if a gap-filling method has been applied.

Value	Description
0	Record has been reviewed and no correction was required.
1	Record has been reviewed and a gap-filling method has been applied.
No value	Record has not been reviewed.

Table 4. The csd_dataGapToFillMethodMapping table provides information on the specific gap-filling method applied to each record that was flagged for data correction.

gapFillMethod	Description
interpolation	Water column height data gap-filled using linear interpolation
constant	Water column height or discharge data gap-filled using a constant value
transducer	Water column height data gap-filled using a linear regression model derived from a co-located pressure transducer sensor
conductivity	Water column height data gap-filled using a linear regression model derived from a co-located specific conductance sensor
usgs	Discharge data gap-filled using a linear regression model derived from a NEON-USGS discharge relationship
none	No method available to fill gap, data will be set to *NA*

5.1.1 Gap-filling Methods: Linear Interpolation and Constant Value

Small data gaps (minutes to hours) or larger gaps (days to weeks) may be filled using either linear interpolation (missing values are estimated by assuming a straight line between known data points) or by applying a constant value throughout the gap. Data that have been filled using these gap-filling methods will have the following fields in **csd_continuousDischarge** table set equal to 1:

- waterColumnHeightGapFilledInterpolation
- waterColumnHeightGapFilledConstant
- dischargeGapFilledInterpolation
- dischargeGapFilledConstant

5.1.2 Gap-filling Methods: Constant Bias Shift

Any shifts in the timeseries (often caused by unintentional movement of the pressure transducer sensor) are corrected by adding or subtracting a constant offset. This method enables comparison between the series' original and adjusted values and aligns the record with any gaugings collected during the period of correction (**Figure 5**). Data that have had a constant bias shift applied will have the following fields in **csd_continuousDischarge** table set to 1:

- waterColumnHeightCorrectedShiftPre
- waterColumnHeightCorrectedShiftPost
- dischargeCorrectedShiftPre
- dischargeCorrectedShiftPost

The suffix 'Pre' in the data fieldname indicates the shift was applied before other gap-filling methods were applied (e.g., linear interpolation, constant value, or regression-based methods). The suffix 'Post' in the data fieldname indicates the shift was applied after any pre-correction shift and following any other gap-filling method. Pre-correction shifts are typically applied to adjust existing data, while post-correction shifts are typically applied to adjust data generated with the Regression gap-filling method (see **Section 5.1.3**). The **csd_constantBiasShift** table is published in the Continuous discharge data product which provides shift date ranges, which timeseries the shift applies to, the magnitude and direction of the shift, and whether the shift was applied pre- or post-data corrections.

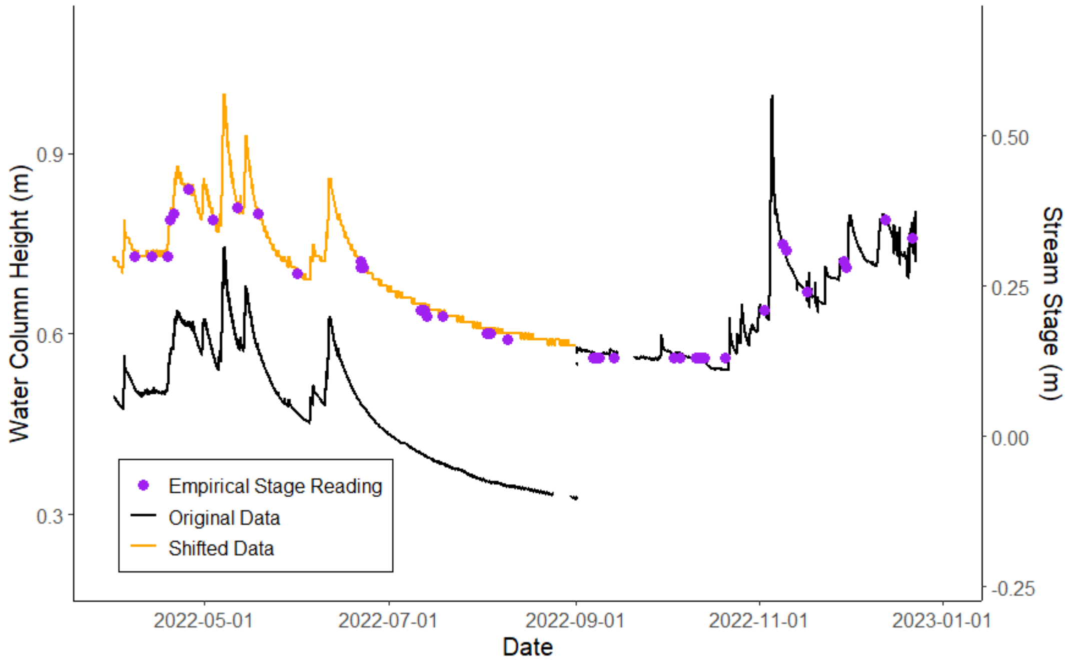

Figure 5. Example of a water column height data corrected using the constant bias shift method. The orange portion of the timeseries has been shifted up to align with the rest of the cleaned water column height data (black) and empirical gaugings (purple dots).

5.1.3 Gap-filling Methods: Regression

Regression gap-filling methods are used when a strong correlation exists between a secondary source of data and the period of water column height or continuous discharge timeseries to be corrected. Linear regression models are developed depending on the step in the algorithm implementation (i.e. during correction of water column height or during correction of continuous discharge) and are specified in the **gapFillMethod** field of the **csd_dataGapToFillMethodMapping** table (**Table 4**).

The coefficients for each regression model are published in the **csd_gapFillingRegression** table in the Continuous discharge data product. Unique regression equations are identified by the **gapFillRegressionID** field in the **csd_gapFillingRegression** table. For each record corrected using a regression method, the gap can be mapped to the regression coefficients by linking **csd_dataGapToFillMethodMapping:gapFillRegressionID** to **gapFillingRegression:gapFillRegressionID**.

Situations may arise when 100% of a data gap cannot be corrected with a regression method. For a given data gap in **csd_dataGapToFillMethodMapping**, if the field **interpolateRemainingGap** is equal to 'Y,' then the remaining percentage of the gap that could not be corrected with the regression method will be corrected using the linear interpolation method. If the field **interpolateRemainingGap** is equal to 'N,' then the remainder of the timeseries data that could not be corrected with the regression method will not be corrected using the linear interpolation method and the value will be set to *NA.*

5.1.3.1 Regression Methods: Relationship with another pressure transducer

Most NEON sites are instrumented with a secondary pressure transducer sensor that can be used to fill gaps in water column height data derived from the primary pressure transducer sensor if the following conditions are met:

1. Data must be available from the secondary pressure transducer within the time range of the gap observed in the timeseries derived from the primary pressure transducer.
2. There must be a strong correlative relationship in water column height between the two transducers before and/or after the gap in the primary timeseries.

If the above conditions are met, a linear regression model is developed using available data from the two pressure transducers within a given date range (records that were flagged for cleaning are not used in the regression model). Gaps are filled by applying coefficients derived from the model to water column height data collected from the secondary transducer to estimate corrected water column height for the primary transducer (**Figure 6**). Data that have been corrected using this gap-filling method will have the **waterColumnHeightGapFilledTransducer** field in the **csd_continuousDischarge** table set to 1.

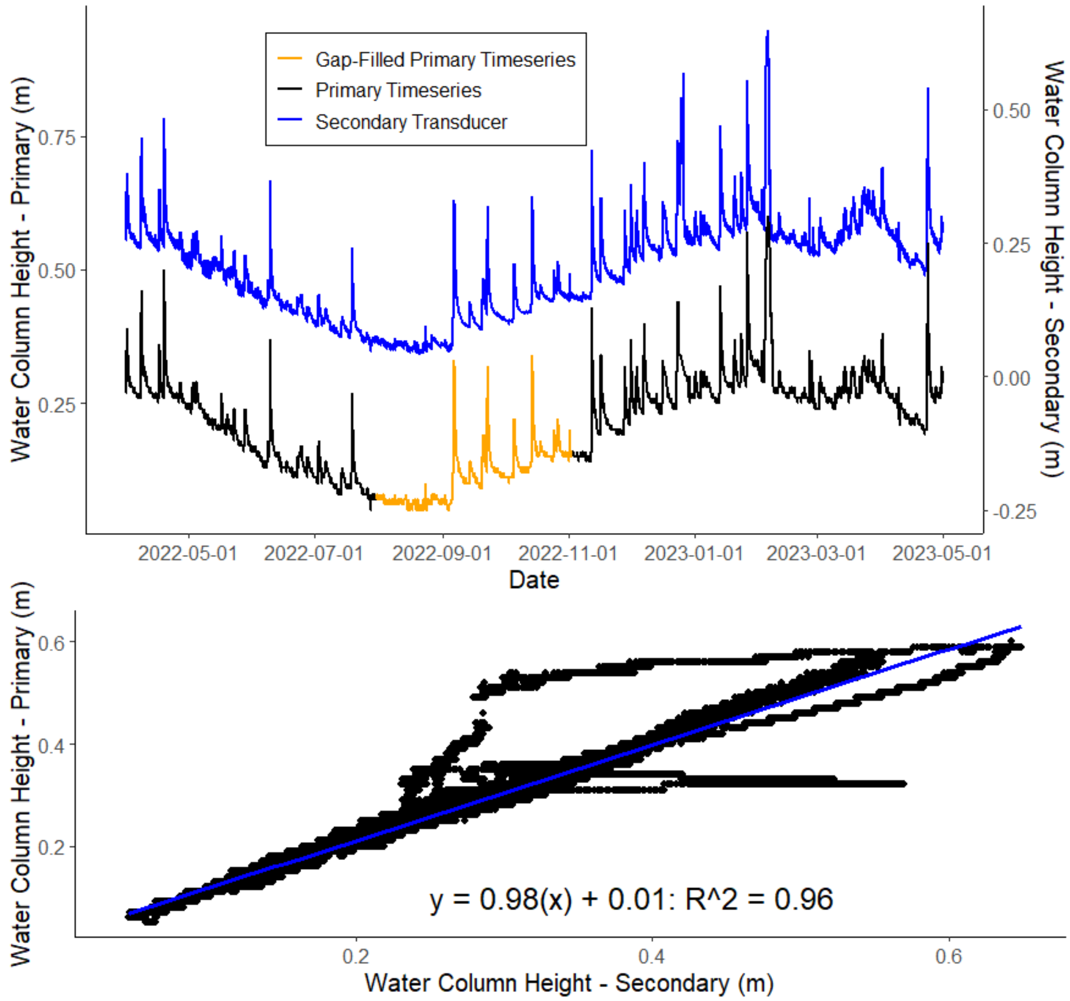

Figure 6. Above: Example of primary water column height data (black line) gap-filled (orange line) using data from secondary pressure transducer (blue line); Below: The linear regression model developed using data from the primary and secondary pressure transducer.

5.1.3.2 Regression Methods: Relationship with a co-located conductivity sensor

The relationship between water column height and specific conductance (derived from a co-located conductivity sensor) may also be used to fill data gaps. A linear regression model is developed with the same methods described in **Section 5.1.3.1** with the co-located specific conductance timeseries as the independent timeseries. This relationship is typically investigated if the pressure transducer regression method is not available or does not meet requirements. Data that have been corrected using this gap-filling method have **waterColumnHeightGapFilledConductivity** in the **csd_continuousDischarge** table set to 1.

5.1.3.3 Gap-filling Methods: Relationship with a co-located USGS Gauging Station

The USGS provides estimates of continuous discharge from thousands of locations throughout the United States. These estimates are derived in a similar manner to NEON; empirical field measurements of stage and discharge (RD[02]) are used to generate a stage-rating curve, which is then applied to a continuous record of stage to generate continuous discharge. The USGS also reviews and adjusts rating curves and continuous discharge estimates in a similar fashion to NEON (RD[03]). USGS data can be accessed from the National Water Information System webpage or using the dataRetrieval R package.

Some NEON sites are located near USGS monitoring stations. At the continuous discharge level, the correlation between NEON and USGS discharge estimates can be used to correct and gap-fill NEON data (**Figure 7**). The same two conditions listed in **Section 5.1.3.1** must be met for paired NEON and USGS discharge timeseries.

To establish a linear relationship between NEON and USGS discharge, both timeseries are first log-transformed:

$$log_{10}(NEON\ Q_{et_i}) = m\left(log_{10}(USGS\ Q_{t_i})\right) + b \tag{5}$$

$$NEON\ Q_{et_i} = 10^{m\left(log_{10}(USGS\ Q_{t_i})\right)+b} \tag{6}$$

Where, at timestamp t_i, *NEON* Q_e is the estimated corrected NEON discharge, *USGS Q* is the USGS discharge value from the co-located monitoring station, *m* is the slope of the linear relationship, and *b* is the intercept of the linear relationship.

Discharge data are published by the USGS at a 15-minute temporal resolution, while NEON discharge data are published at a 1-minute temporal resolution. Due to the differences in resolution, the NEON-USGS discharge regression method can only gap-fill at a 15-minute timestamp. To produce a continuous 1-minute timeseries for the **csd_continuousDischarge** table, records within the 15-minute interval are filled via linear interpolation. Data that have been corrected using this gap-filling method have **dischargeGapFilledUSGS** in the **csd_continuousDischarge** table set to 1.

For each unique regression developed using this method, the USGS monitoring station location ID is published in the **usgsSiteNumber** field of the **csd_gapFillingRegression** table and the date in which the USGS data were accessed is published in the **usgsDownloadDate** field of the **csd_gapFillingRegression** table.

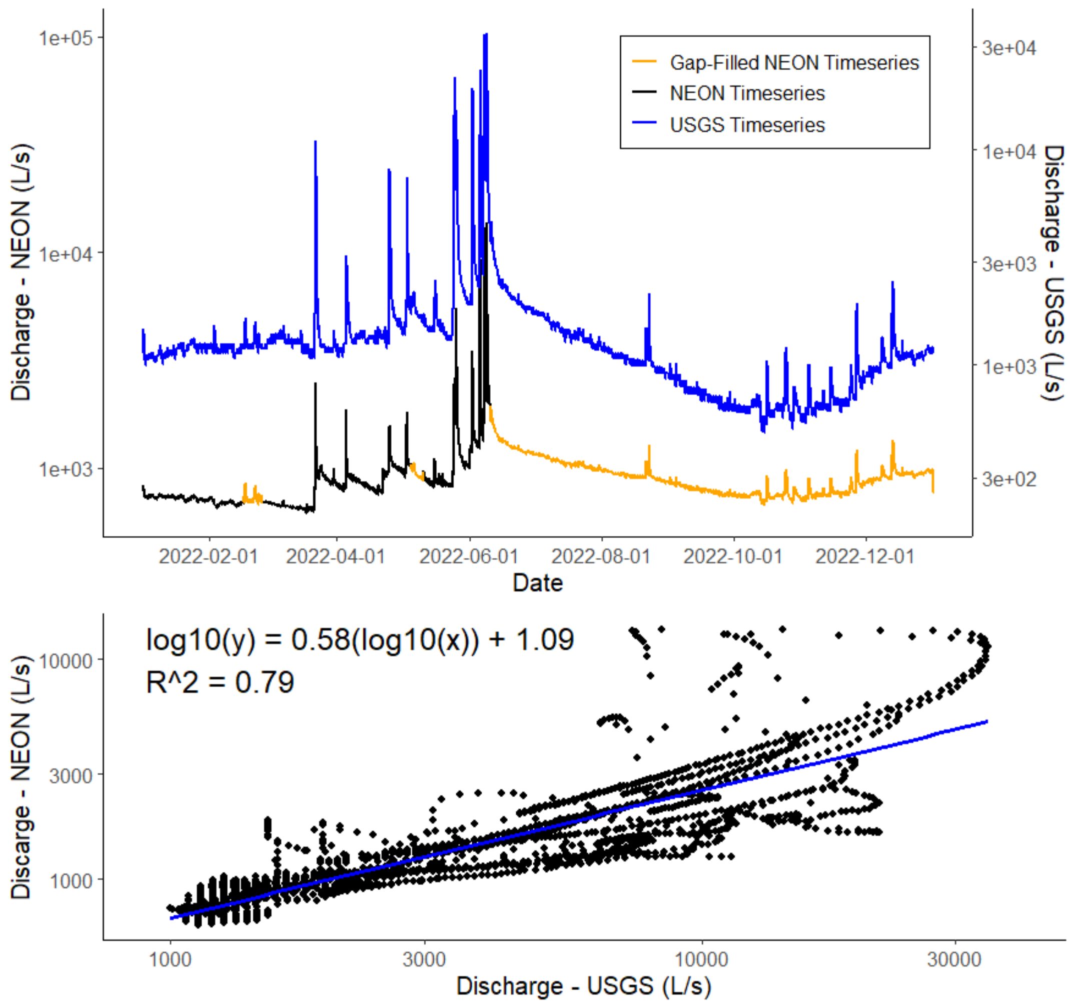

Figure 7. Above: Example of NEON continuous discharge data (black line) gap-filled (orange line) using data from a co-located USGS station (blue line); Below: The linear regression model developed using data from the USGS and NEON stations.

5.1.4 Corrections When No Method Applies

At both the water column height and continuous discharge level there may be situations where no gap-filling method can be applied to adequately fill a data gap. In these situations, timeseries values in the **csd_continousDischarge** table (**waterColumnHeightCorrected** or **continuousDischarge** fields, depending on the step in the algorithm implementation) will be published as *NA*. Users can identify records that fit this situation in the **csd_continuousDischarge** table by observing the following:

- **waterColumnHeightCorrected** is *NA* and **waterColumnHeightCorrectionApplied** = 1
- **continuousDischarge** is *NA* and **dischargeCorrectionApplied** = 1

5.1.5 Quality and Indicator Flags

5.1.5.1 Quality Flags for Corrected and Uncorrected Data

Quality flags (AD[09]) are not published for the Stage-discharge rating curves data product as quality issues are incorporated in the Bayesian model and its uncertainty estimates. The *Null* quality flag is published for the Continuous discharge data product and is contained in the **dischargeNullQF** field of the **csd_continuousDischarge** table. This flag can be applied to both corrected and uncorrected continuous discharge data.

Once data correction is complete for periods of water column height and continuous discharge, any science review quality flags (**dischargeFinalQFSciRvw**) equal to 1 are set to equal 0 as those records are gap-filled in the **continuousDischarge** field of the **csd_continuousDischarge** table. For provisional data, and/or any periods where data correction could not be completed, **maxpostDischarge** fields will retain a science quality review flags (**dischargeFinalQFSciRvw** = 1) if expert review deems that data are suspect.

5.1.5.2 Indicator Flags for Corrected Data

Table 5 and **Table 6** list indicator flags published in the **csd_continuousDischarge** table that describe the gap-filling methods used to correct water column height and continuous discharge data, respectively. The fields are binary, and can be interpreted as follows:

- 0 = Gap-filling method was not applied
- 1 = Gap-filling method was applied
- *NA* = The record has not yet been reviewed for correction

Table 5. Indicator Flags associated with water column height data.

Indicator Flag	Description
waterColumnHeightCorrectedShiftPre	A constant bias shift was applied prior to gap-filling
waterColumnHeightCorrectedShiftPost	A constant bias shift was applied following gap-filling
waterColumnHeightGapFilledInterpolation	Data were gap-filled using linear interpolation
waterColumnHeightGapFilledConstant	Data were gap-filled using a constant value
waterColumnHeightGapFilledTransducer	Data were gap-filled using a linear regression based on a secondary pressure transducer sensor
waterColumnHeightGapFilledConductivity	Data were gap-filled using a linear regression based on a co-located specific conductivity sensor
waterColumnHeightCorrectionApplied	Indicates whether a gap-filling method was applied, if = 0, the water column height data required no correction

Table 6. Indicator Flags associated with continuous discharge data.

Indicator Flag	Description
dischargeCorrectedShiftPre	A constant bias shift was applied prior to gap-filling
dischargeCorrectedShiftPost	A constant bias shift was applied following gap-filling
dischargeGapFilledInterpolation	Data were gap-filled using linear interpolation
dischargeGapFilledConstant	Data were gap-filled using a constant value
dischargeGapFilledUSGS	Data were gap-filled using a linear regression based on a USGS station
dischargeCorrectionApplied	Indicates whether a gap-filling method was applied, if = 0, the discharge data required no correction

6 UNCERTAINTY

The expanded data package includes uncertainties for both **equivalentStage** and the uncorrected **maxpostDischarge** values. For **equivalentStage**, the **stageUnc** field is calculated from the mean difference between observed and predicted stage. The **maxpostDischarge** uncertainties are derived from the Bayesian model and include both parametric and remnant uncertainties, with one and two standard deviations. These model uncertainties apply only to **maxpostDischarge** records and **continuousDischarge** records that have not been corrected using a gap-filling method (cases where **maxpostDischarge** equals **continuousDischarge**). In cases where **continuousDischarge** has been corrected, the uncertainties derived from the Bayesian model no longer apply and are not published.