NEON USER GUIDE TO STAGE-DISCHARGE RATING CURVES (NEON.DP4.00133)

PREPARED BY	ORGANIZATION	DATE
Kaelin M Cawley	AQU	11/16/2020
Nick Harrison	AQU	11/16/2020
Zachary Nickerson	AQU	11/16/2020

CHANGE RECORD

REVISION	DATE	DESCRIPTION OF CHANGE
A	01/11/2018	Initial Release
B	10/19/2020	Updated to reflect algorithm changes and additional input and output data tables

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 1 OS data, and associated metadata. In the NEON data products framework, the raw data collected in the field (i.e. staff gauge and discharge measurements from a single collection event) 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 L1 or higher, 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 data product Stage-discharge rating curves (NEON.DP4.00133) - the parameters that describe the relationship between staff gauge measurements and discharge measurements made over a range of dates and flow conditions - and associated metadata from input data and calculations. This document also provides details relevant to the publication of the data products via the NEON data portal, with additional detail available in the file, NEON Data Variables for Stage-Discharge Rating Curves (NEON.DP4.00133) (AD[03]), provided in the download package for this data product.

This document describes the process for performing custom calculations derived from L1 discharge field collection data (NEON.DP1.20048), L4 wadeable stream morphology data (DP4.00131.001), L4 bathymetric and morphological maps (DP4.00132.001), geolocation data, and previous L4 Stage-discharge rating curves data. For information of the raw data that are ingested and processed for the source data product see the following: for discharge field collection, NEON User Guide to Discharge Field Collection (AD[07]) and NEON.DP0.20048.001_dataValidation.csv (AD[08]); for wadeable stream morphology, NEON User Guide to Wadeable Stream Morphology (AD[09]); for bathymetric and morphological maps, NEON User Guide to Bathymetric and Morphological Maps (AD[10]). Documents are available for download with the respective L1 data package. 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]	NEON.DP4.00133.001 _variables.csv	NEON Data Variables for Stage-Discharge Rating Curves (NEON.DP4.00133)
AD[04]	NEON.DOC.001152	NEON Aquatic Sampling Strategy
AD[05]	NEON.DOC.000008	NEON Acronym List
AD[06]	NEON.DOC.000243	NEON Glossary of Terms
AD[07]	NEON_streamDischarge _userGuide.pdf	NEON User Guide to Discharge Field Collection (NEON.DP1.20048)
AD[08]	NEON.DP1.20048.001 _validation.csv	NEON Data Validations for Disharge Field Collection (NEON.DP1.20048)
AD[09]	NEON_streamMorphology _userGuide.pdf	NEON User Guide to Wadeable Stream Morphology (NEON.DP4.00131.001)
AD[10]	NEON_bathymetry _user-Guide.pdf	NEON User Guide to Bathymetric and Morphological Maps (NEON.DP4.00132.001)

2.2 Acronyms

Acronym	Definition
Q	Discharge
H	Stage height
m	Meters
m3/s	Cubic meters per second
L/s	Liters per second
ADCP	Acoustic Doppler Current Profiler
MCMC	Markov Chain Monte Carlo
BaM	Bayesian Modeling
BaRatinAGE	Bayesian Rating Curve Advanced Graphical Environment
GUI	Graphical User Interface
UI	User Interface

3 DATA PRODUCT DESCRIPTION

The Stage-discharge rating curves (DP4.00133.001) data product provides a rating curve that describes the relationship between staff gauge and discharge measurements at stream, river, and lake inflow/outflow locations each water year (October 1^{st} - September 30^{th}). The relationship developed for the rating curve is used for computing the Continuous discharge (NEON.DP4.00133) data product. An example rating curve that was plotted for a range of discharge (y-axis) and stage values (x-axis) is plotted below (Figure 1).

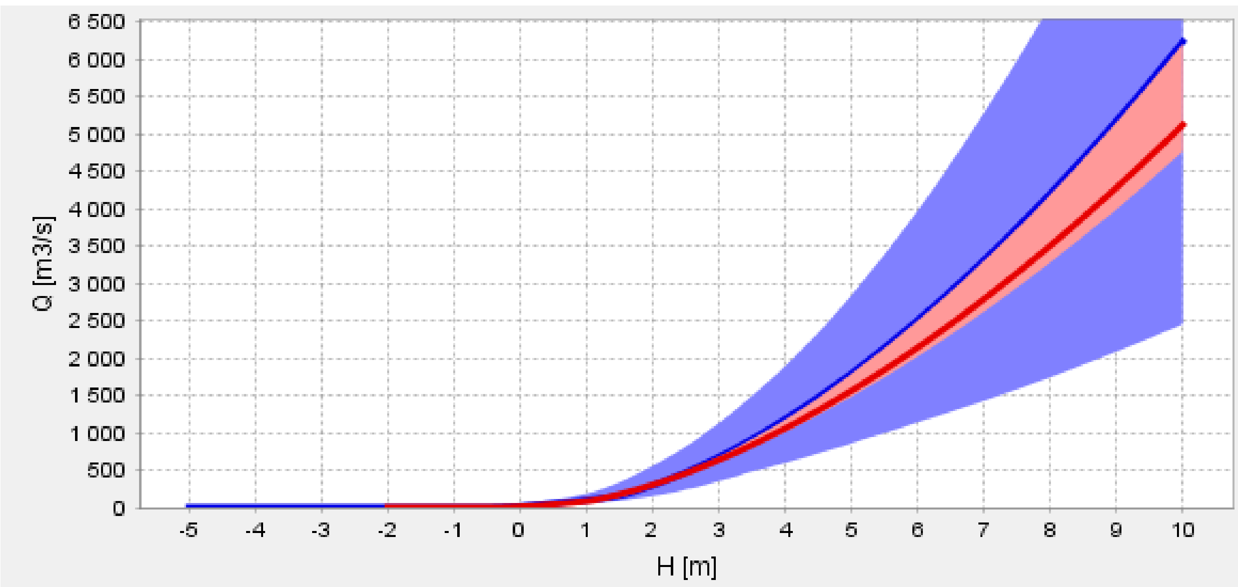

Figure 1: An example rating curve (red line) with parametric uncertainty (light red shading). Blue line and shading represent "prior" rating curve and parametric uncertainty using information based on hydraulic controls only, respectively. Picture taken from "Rating Curve" documentation for BaRatinAGE (Le Coz, 2013; Le Coz, 2014).

3.1 Spatial Sampling Design

A stage-discharge rating curve is developed at the site level (except for Toolik Lake, where two stations will have rating curves produced, the inflow and outflow to the lake) using data collected at the same spatial scale as the discharge and stage measurements, which are measured at all wadeable stream sites on a run or riffle near sensor set #1 or sensor set #2 and near the buoy at river sites. Point measurements of water depth and velocity are made along the transect from one bank to the other using a wading rod and attached velocity meter (at wadeable stream sites and Toolik Lake inflow and outflow) or acoustic Doppler current profiler instrumentation (at wadeable streams, non-wadeable rivers, and Toolik Lake inflow and outflow). If a discharge cross-section or staff gauge is re-located, a new rating curve may be developed for a site if the hydrologic data is no longer directly comparable to the existing set of stage and discharge measurements.

Table 1: List of input data for Stage-Discharge Rating Curves.

Data Source	DPID	table
L1 Discharge field collection data	NEON.DP1.20048	dsc_fieldData
L1 Discharge field collection data	NEON.DP1.20048	dsc_individualFieldData
L1 Discharge field collection data	NEON.DP1.20048	dsc_fieldDataADCP
L4 Stream morphology map	NEON.DP4.00131	geo_controlInfo
L4 Stream morphology map	NEON.DP4.00131	geo_priorParameters
L4 Stream morpgology map	NEON.DP4.00131	geo_curveIdentification
L4 Bathymetric and morphological maps	NEON.DP4.00132	bat_controlInfo
L4 Bathymetric and morphological maps	NEON.DP4.00132	bat_priorParameters
L4 Bathymetric and morphological maps	NEON.DP4.00132	bat_curveIdentification

Table 2: List of geolocation inputs for Stage-Discharge Rating Curves.

Data Source	field	location
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

3.2 Temporal Sampling Design

A stage-discharge rating curve for a site or station is created following the conclusion of the water year (October 1st to September 30th).

3.3 Variables Reported

All data and geolocation variables used as inputs for the development of a stage-discharge rating curve are listed in Table 1 and Table 2. All variables reported in the published data (L4 data) are also provided separately in the file, NEON Data Variables for Stage-Discharge Rating Curves (NEON.DP4.00133) (AD[03]).

Field names have been standardized with Darwin Core terms (http://rs.tdwg.org/dwc/; accessed 16 February 2014), the Global Biodiversity Information Facility vocabularies (http://rs.gbif.org/vocabulary/gbif/; accessed 16 February 2014), the VegCore data dictionary (https://projects.nceas.ucsb.edu/nceas/projects/bien/wiki/VegCore; accessed 16 February 2014), where applicable. For AOS, Earth Gravitational Model 96 (EGM96) is the reference gravitational ellipsoid. Latitudes and longitudes are denoted in decimal notation to six decimal places, with longitudes indicated as negative west of the Greenwich meridian.

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

3.4 Spatial Resolution and Extent

The finest spatial resolution at which data are reported is a site, except at Toolik Lake where data is published from two stations.

3.5 Temporal Resolution and Extent

The finest temporal resolution at which data are reported is the interval of a rating curve not to exceed one water year (October 1st to September 30th). When there is a single stage-discharge rating curve for a water year, data will be published for September of said water year. When changes in channel morphology, hydrology, or geolocation data necessitate more than one stage-discharge rating curve for a water year, data will be published for multiple months of a water year. Users should download stage-discharge rating curve data from the NEON data portal for the time interval of October - September of a water year to ensure they are downloading the extent of the data available.

3.6 Associated Data Streams

The data from this L4 data product is derived from a level 1 (L1) data product: Discharge field collection (NEON.DP1.20048). These data products can be linked by **siteID**.

The data from this L4 data product is derived from a level 4 (L4) data product: Stream morphology map data product (NEON.DP4.00131). These data products can be linked by **siteID**.

The data from this L4 data product is derived from a level 4 (L4) data product: Bathymetric and morphological maps data product (NEON.DP4.00132). These data products can be linked by **namedLocation**.

The data from this L4 data product is used to develop a level 4 (L4) data product: Continuous discharge (NEON.DP4.00130). These data products can be linked 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.

Stage-discharge rating curves yields approximately 29 rating curves (1 per stream and river site, 2 at Toolik Lake) per water year comprised of approximately 702 stage and discharge pair records per year during stage-discharge relationship development and 324 stage and discharge records per year once a relationship has been developed.

3.8 Data Relationships

The algorithm used for this L4 data product produces as many records in sdrc_gaugeDischargeMeas as is used to develop a rating curve, which will include approximately as many records as in the past water year in the dsc_fieldData table for discharge field collection (NEON.DP1.20048), and may include records from previous water years if the measurements are relatable to the past water year. The values may differ in the two tables because, during rating curve development, the gauging record for each water year is evaluated, and outliers in the stage/discharge gauging record are removed from individual gauging records if deemed appropriate (due to measurement errors or other apparent problems associated with the gauging). Gaugings from previous years that match the stage-discharge relationship observed in the current water year and have relatable staff gauge geolocations are included in the new rating curve. Discharge values are recalculated from point measurements in the dsc_individualFieldData table for discharge field collection (NEON.DP1.20048) and stage values are offset using information from the NEON geolocation database. Stage offset values correct staff gauge height data after having changed location. First, a total reference elevation is calculated for a staff gauge location as the sum of a staff gauge's elevation above sea level (m) and zOffset (vertical correction needed in order for the staff gauge location to be relatable to previous staff gauge locations; m). Second, the total reference elevation for the initial location of a staff gauge is subtracted from each subsequent location to obtain the stage offset value for each subsequent location. Third, every stage measurement between the start date and end date of a staff gauge location will have the appropriate stage offset applied. One record is created for each hydraulic control at a site for the sdrc_posteriorParameters table. One record is created for each rating curve (usually one per year) for the sdrc_stageDischargeCurveInfo table. 500 records per hydraulic control at a site are created for the sdrc_sampledParameters, which represent the Markov Chain Monte Carlo (MCMC) samples of the range of model parameters. One record per staff gauge/discharge measurement set used to develop the rating curve is created in the sdrc_resultsResiduals table. Hydrologic control data and stage-discharge rating curve identification data are used to develop this L4 data product. For stream sites, information on the control and curve identification tables used to develop stage-discharge rating curves (geo_controlInfo, geo_priorParameters, geo_curveIdentification) can be found in the user guide for stream morphology (AD[09]). For large river sites and the Toolik Lake inflow and outflow stations, information on the control and curve identification tables used to develop stage-discharge rating curves (bat_controlInfo, bat_priorParameters, bat_curveIdentification) can be found in the user guide for bathymetry and morphological maps (AD[10]).

sdrc_gaugeDischargeMeas.csv - > One record expected per staff gauge and discharge measurement used to develop the rating curve for the current water year

sdrc_posteriorParameters.csv - > One record expected per hydraulic control at a site

sdrc_stageDischargeCurveInfo.csv - > One record expected per rating curve developed for a site for the current water year

sdrc_sampledParameters.csv - > 500 records expected per hydraulic control at a site

sdrc_resultsResiduals.csv - > One record expected per staff gauge and discharge measurement used to develop the rating curve for the current water year

geo_controlInfo.csv - > One record per the square of the number of hydraulic control(s) at a site, e.g. there will be 4 records if there are 2 controls. This table will be downloaded with stream sites, is an exact copy of the table of the same name in the download package for the stream morphology map data product (NEON.DP4.00131), and is include here for its value in evaluating Equation 1.

bat_controlInfo.csv - > One record per the square of the number of hydraulic control(s) at a site, e.g. there will be 4 records if there are 2 controls. This table will be downloaded with NEON large river sites and the Toolik Lake inflow and outflow stations, is an exact copy of the table of the same name in the download package for the bathymetric and morphological maps data product (NEON.DP4.00132), and is include here for its value in evaluating Equation 1.

gaugeEventID can be used to link data in the sdrc_gaugeDischargeMeas and sdrc_resultsResiduals tables. There may be multiple entries for a given **gaugeEventID** in the sdrc_resultsResiduals table if the staff gauge and discharge readings were used in more than one rating curve, which is common if the stream or river does not undergo major geomorphic change between water years.

allEventID in the sdrc_stageDischargeCurveInfo table contains a concatenated string, which can be parsed using a "|" delimiter to determine which **gaugeEventIDs** were used for developing a specific rating curve. Parsing **allEventID** will enabe linking the sdrc_stageDischargeCurveInfo table with the sdrc_gaugeDischargeMeas and sdrc_resultsResiduals tables.

curveID can be used to link the sdrc_posteriorParameters, sdrc_stageDischargeCurveInfo, sdrc_sampledParameters, and sdrc_resultsResiduals tables.

4 Algorithm Theoretical Basis

4.1 Theory of measurement

Discharge is the volume of water flowing through a stream or river over time and is a function of the height of the water column, cross-sectional area, and water velocity. In a natural stream, riffles can act as weirs and the stream velocity is often a function of channel roughness where natural weirs are not present. The stream channel's physical dimension controls the cross-sectional area as the water column height changes (Figure 2). Staff gauges are used to measure the surface water elevation relative to a fixed reference. The staff gauge does not measure absolute water depth. Staff gauge measurements are collected at the start and end of discharge measurements where wading surveys are utilized in wadeable streams and acoustic Doppler current profiler instrumentation (ADCP) are utilized in non-wadeable rivers. With these paired measurements and the physical hydraulic characteristics, a stage-discharge rating curve is developed using the Bayesian Modeling (BaM) algorithm, described below.

4.2 Theory of Algorithm

The NEON Stage-discharge rating curves data product is 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 1). 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)

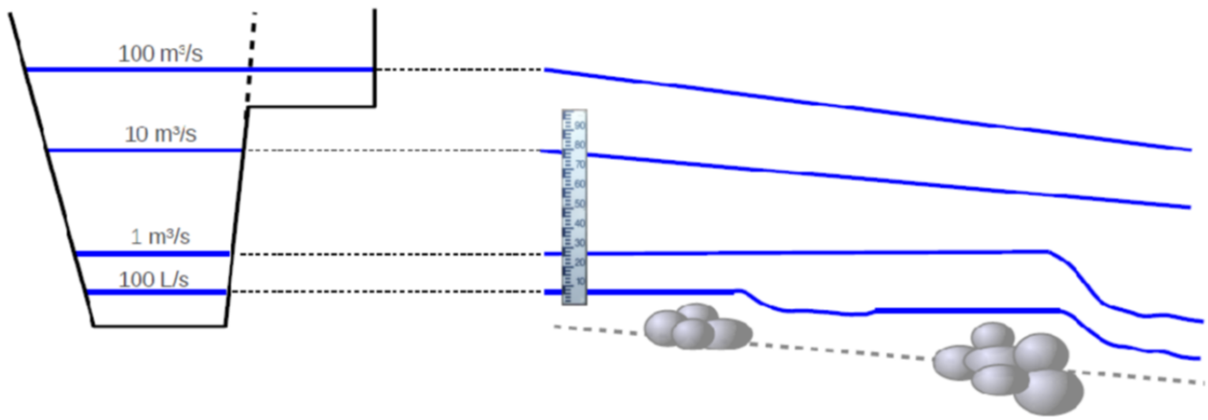

Figure 2: Diagram of natural riffles acting like weirs at low water and channel controlled at higher water level. Picture taken from "Hydraulic Analysis" documentation for BaRatinAGE (Le Coz et al., 2013; Le Coz et al., 2014).

sampling (Le Coz, 2014). Equation 1 and text below are taken from "Rating curve equation" documentation for BaRatinAGE (Le Coz et al., 2013).

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

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).

Users are directed to the information in the **geo_controlInfo.csv** table for the data needed for the $1_I(h)$ function.

4.3 Algorithm Implementation

The NEON OS transition system runs a Docker container containing R code to create a rating curve every 366 days (to account for leap years) to produce the annual rating curve (similar infrastructure to Metzger, 2017 without the use of HDF5 file formats).

Within the Docker container:

1. Data for the site and water year is queried from the NEON database.
2. Configuration files for BaM are created.
3. BaM executable is run.
4. BaM 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 (UI).

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, 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 Change Log section of the data product readme, provided with every data download, contains a history of major known errors and revisions.

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 liklihood 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.

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. Please see below for an explanation of **dataQF** codes specific to this product.

In addition to the **dataQF** field in each table there is also a **recalculatedL1QF** that indicates if the recalculated discharge value is different than the L1 published discharge value (1 indicates they are different and 0 indicates that they are not different) and a **L1DataQF** that indicates a quality flag associated with the L1 input 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
recalculatedL1QF	1 or 0	1 indicates L1 and recalculated L4 discharge values are different and 0 indicates that they are not different
L1DataQF	varies	This field contains the L1 quality flag associated with input L1 data, if applicable

6 REFERENCES

Le Coz, J., B. Renard, L. Bonnifait, F. Branger, R. Le Boursicaud (2014) Combining hydraulic knowledge and uncertain gaugings in teh 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.