NEON USER GUIDE TO STREAM DISCHARGE (NEON.DP4.00130)

PREPARED BY	ORGANIZATION	DATE
Kaelin M Cawley	AQU	04/12/2018
Nick Harrison	AQU	04/12/2018

CHANGE RECORD

REVISION	DATE	DESCRIPTION OF CHANGE
A	01/11/2018	Initial Release

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 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 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 data product Stream Discharge (NEON.DP4.00130) - the continuous stream discharge calculated from water level measurements using the stage discharge rating curve relationship - 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 Stream Discharge (NEON.DP4.00130) (AD[03]), provided in the download package for this data product.

This document describes the process for performing custom calculations derived from L0 pressure transducer data, L1 Stream discharge field collection data (NEON.DP1.20048), geolocation data, and L4 Stream discharge rating curve data. For information about the raw data that are ingested and processed in the source data product see NEON.DP0.20048.001_dataValidation.csv and NEON User Guide to Stream Discharge (NEON.DP1.20048) available for download with the 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.DP4.00130.001 _variables.csv	NEON Data Variables for Stream Discharge (NEON.DP4.00130)
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.DOC.000927	NEON Calibration and Sensor Uncertainty Values
AD[08]	NEON.DOC.001113	Algorithm Theoretical Basis Document: Quality Flags and Quality Metrics for TIS Data Products
AD[09]	NEON.DOC.011081	Algorithm Theoretical Basis Document: QA/QC Plausibility Testing

2.2 Acronyms

Acronym	**Definition**
Q	Stream discharge
GUI	Graphical User Interface

Table 1: List of input data for Continuous Stream Discharge.

Data Source	DPID	table
L4 Stream Morphology Map	NEON.DP4.00131	geo_controlInfo
L4 Stream Morphology Map	NEON.DP4.00131	geo_priorParameters
L4 Stream Discharge Rating Curve	NEON.DP4.00133	sdrc_sampledParameters
L4 Stream Discharge Rating Curve	NEON.DP4.00133	sdrc_gaugeDischargeMeas
L4 Continuous Stream Discharge	NEON.DP4.00130	csd_pressureGaugeRelationship

3 DATA PRODUCT DESCRIPTION

The stream discharge data product provides calculated discharge values every minute, which are derived from once per minute pressure readings and the stage-discharge rating curve.

3.1 Spatial Sampling Design

Stream discharge is reported at the stream or river level (Horizontal index = 100, Vertical index = 100) using data collected at either sensor set #1 (Horizontal index = 101, Vertical index = 100) or sensor set #2 (Horizontal index = 102, Vertical index = 100), whichever is closer to the staff gauge in wadeable streams, or at the nearshore sensor set (AKA sensor set #1) at large rivers. 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 (S1 = 101, S2 = 102), and **namedLocation** - the configured location of the pressure transducer L0 input data.

3.2 Temporal Sampling Design

Every week, the week of pressure data that ended 21 days ago is transitioned. This is in order to allow 14 days for field operations to enter and transmit gauge height readings followed by a week for QAQC. For example, on March 1st the data for February 1-7th will be transitioned. Data generally appear on the NEON Data Portal the day after they are transitioned.

3.3 Variables Reported

All data and geolocation variables used as inputs for continuous stream discharge are listed in Table 1, Table 2, and Table 3. For more information on calibration and validation assessment of sensors repeatability, see AD[07]. All variables reported in the published data (L4 data) are also provided separately in the file, NEON Data Variables for Stream Discharge (NEON.DP4.00130) (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

Table 2: List of calibration inputs for Continuous Stream 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 Stream Discharge for SITE.AOS.gauge location.

Data Source	field	contents
NEON geolocation database	location startDate	Start date for location
NEON geolocation database	location endDate	End date for location
NEON geolocation database	location xOffset	Flag to indicate whether the readings from this location can be continued from the previous staff gauge using the zoffset (0 = not continued, 1 = continued)
NEON geolocation database	location yOffset	Station horizontal index code for nearest pressure transducer
NEON geolocation database	location zOffset	Vertical offset from previously installed staff gauge

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 spatial data are reported is a configured sensor location at a site.

3.5 Temporal Resolution and Extent

The finest temporal resolution at which gauge relationship data are reported is the date and time of a set of pressure measurements and corresponding staff gauge reading for the csd_pressureGaugeRelationship table. The finest temporal resolution at which discharge data are reported is one minute for the continuous discharge table, csd_continuousDischarge.

3.6 Associated Data Streams

The data from this L4 data product are derived from a level 1 (L1) data product: Gauge Height (NEON.DP1.20367). These data products can be linked by **siteID** in both tables and gag_fieldData:**collectDate** csd_pressureGaugeRelationship:**gaugeCollectDate**.

The data from this L4 data product are 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 are derived from a level 4 (L4) data product: Stream discharge rating curve (NEON.DP4.00133). These data products can be linked by **curveID**.

The data from this L4 data product are derived from a level 4 (L4) data product: Continuous stream discharge (NEON.DP4.00130). These data products can be linked by **siteID** and **calibrationID**.

3.7 Product Instances

The NEON Observatory contains 24 wadeable streams and 3 large rivers.

At each site, Stream Discharge yields approximately 104 gauge and mean pressure readings per year (~2 per week) and 525,600 continuous discharge records per year (~1 per minute).

3.8 Data Relationships

The algorithm used for this L4 data product produces as many records in csd_pressureGaugeRelationship as there are for a site for the gag_fieldData table for the Gauge Height data product (NEON.DP1.20267). One record is created in csd_continuousDischarge for each L1 record.

csd_pressureGaugeRelationship.csv - > One record expected per staff gauge reading collected for a site in the past water year

csd_continuousDischarge.csv - > One record expected per L0 surface water pressure measurement collected (~1 per minute)

siteID and **calibrationID** can be used to link data in the csd_pressureGaugeRelationship and csd_continuousDischarge tables.

4 Algorithm Theoretical Basis

4.1 Theory of measurement

Stream 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 dicharge (Figure 1).

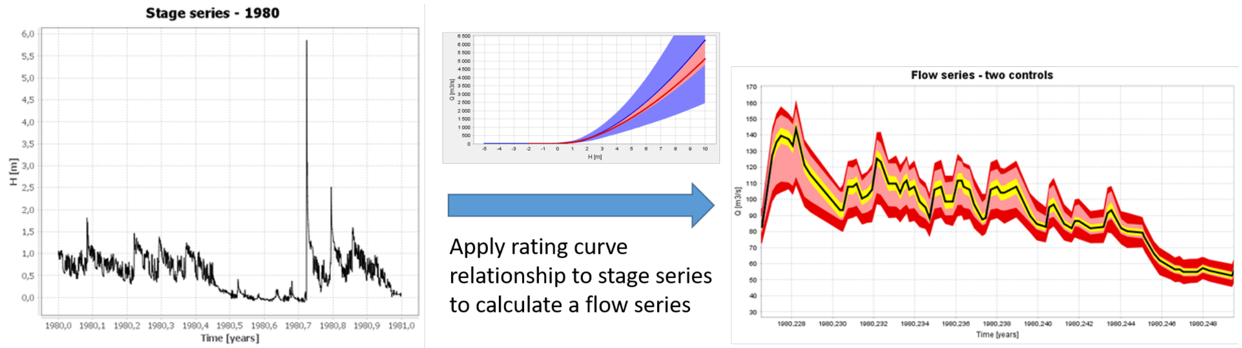

Figure 1: Example stage series and flow series output. Pictures taken from "Stage series" and "Flow series" documentation for BaRatinAGE (Le Coz et al., 2013; Le Coz et al., 2014).

4.2 Theory of Algorithm

Pressure transducers are installed at all NEON sites and are used to develop a continuous discharge record at wadeable streams and large rivers. The pressure measured at the sites can be used to estimate stage by converting to water column height and then applying an offset from the site survey(Equation 1).

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

where,

P_{sw} = calibrated surface water pressure, kPa

ρ = Density of water, 999 kg/m^3

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

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

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

The estimated stage timeseries is converted to discharge using the stage-discharge rating curve (DP4.20133). The NEON stage-discharge rating curve and continuous stream discharge 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 2). 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 2 and text below are taken from "Rating curve equation" documentation for BaRatinAGE (Le Coz et al., 2013).

$$Q(h) = \sum_{r=1}^{N_{segment}} \left(1_{[K_{r-1};K_r]}(h) \times \sum_{j=1}^{N_{control}} M(r,j) \times a_j(h-b_j)^{c_j} \right) \quad (2)$$

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).

4.3 Algorithm Implementation

The NEON OS transition system runs a Docker container containing R code to estimate continuous discharge (similar infrastructure to Metzger, 2017 without the use of HDF5 file formats).

Within the Docker container:

1. 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) is queried from the NEON database.
4. Configuration files for BaM are created.
5. BaM executable is run.
6. 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). See the Data Product User Guide and Validation file associated with the Gauge Height data product 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, 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 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.

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[08] and AD[09] 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.