

OPEN
ANALYSIS

User-focused evaluation of National Ecological Observatory Network streamflow estimates

Spencer Rhea¹  , Nicholas Gubbins², Amanda G. DelVecchia^{1,3}, Matthew R. V. Ross² & Emily S. Bernhardt¹

Accurately estimating stream discharge is crucial for many ecological, biogeochemical, and hydrologic analyses. As of September 2022, The National Ecological Observatory Network (NEON) provided up to 5 years of continuous discharge estimates at 28 streams across the United States. NEON created rating curves at each site in a Bayesian framework, parameterized using hydraulic controls and manual measurements of discharge. Here we evaluate the reliability of these discharge estimates with three approaches. We (1) compared predicted to observed discharge, (2) compared predicted to observed stage, and (3) calculated the proportion of discharge estimates extrapolated beyond field measurements. We considered 1,523 site-months of continuous streamflow predictions published by NEON. Of these, 39% met our highest quality criteria, 11% fell into an intermediate classification, and 50% of site-months were classified as unreliable. We provided diagnostic metrics and categorical evaluations of continuous discharge and stage estimates by month for each site, enabling users to rapidly query for suitable NEON data.

Introduction

The flow regime of a river is a fundamental driver of its biology, biogeochemistry, and geomorphology¹. Streamflow, or stream discharge, influences many ecological processes, including ecosystem respiration², greenhouse gas concentrations and flux^{3,4}, and nutrient and element transport⁵. Streamflow is also a fundamental measurement in hydrology used for rainfall-runoff models⁶, fluvial geomorphological applications⁷, and understanding impacts of land use and climate change on hydrologic systems⁸. Streamflow estimates are directly used to estimate gas exchange rates applied in modeling stream metabolism^{9,10} and to calculate greenhouse gas evasion¹¹.

Broad scale and continuous estimates of streamflow that correspond to other ecological data are rare. The United States Geological Survey (USGS) collects continuous stream discharge estimates at approximately 13,000 stations nationwide, but gauges are not necessarily collocated with any ecological data, and routine, consistent measurement of any associated variables is generally not a priority¹². In contrast, many studies have quantified ecosystem processes that are either related to or dependent on direct estimates of streamflow, but at local or temporally limited scales depending on when sampling is possible¹³, which tends to be in warmer months^{3,11,14}.

The National Ecological Observatory Network (NEON) provides an unprecedented opportunity to compare up to five years of corresponding, continuously and consistently collected terrestrial, aquatic, and subsurface data from 28 stream and river (referred to as 'stream') sites across the United States (Fig. 1). Until 2021, ecological analyses of these data were somewhat limited by a lack of continuous discharge estimates. The NEON Continuous Discharge data product was released in 2021 as DP4.00130.001¹⁵ and provided both the continuous estimates and associated uncertainty ranges.

Streamflow is traditionally estimated by converting a continuous measurement of water height (stage) to streamflow based on the relationship between stage and discharge¹⁶. Stage is converted to streamflow using different methods including: empirical equations (such as Manning's Equation)¹⁷, empirical rating curves that relate discharge to stage measured across a range of flow conditions¹⁶, or modeled estimates¹⁸. One of the most common methods for calculating a time series of streamflow for a particular site is the rating curve method. Although it is widely accepted that rating curves have uncertainty, data is often used as an errorless value in

¹Duke University, Durham, NC, 27705, USA. ²Colorado State University, Fort Collins, CO, 80523, USA. ³The University of North Carolina at Chapel Hill, Chapel Hill, NC, 27514, USA. ✉e-mail: spencer.rhea@duke.edu

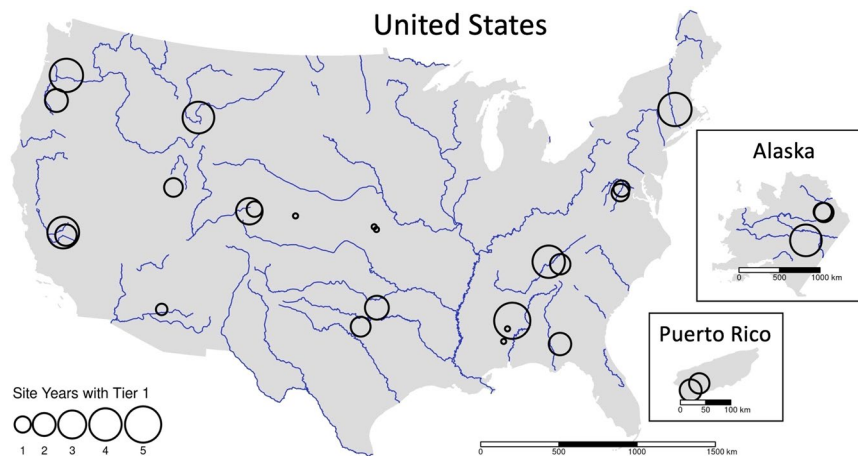


Fig. 1 Map of all NEON aquatic sites where NEON estimates streamflow, sized by the total amount of Tier 1 data available at the site.

applications¹⁹. Uncertainty in rating curves can come from many sources including: error in flow measurements²⁰, extrapolations beyond measured streamflow²¹, and a poor relationship between stream stage and streamflow.

In order to quantify error in rating curves, NEON used a newer approach that generates streamflow estimates with corresponding uncertainty ranges²². This approach, described in detail by Le Coz *et al.*, 2017, used a Bayesian framework that combined information from measured streamflow (the likelihood functions) and hydrologic controls (priors) into a “MaxPost” rating curve (the rating curve that relates to the maximum-posterior parameter values)²². Using the Bayesian approach provided two benefits over traditional rating curves. One, information on hydrologic controls, such as riffles, large obstructions, and other geomorphic structures which constrain flow could be incorporated into the rating curves²². Because different hydrologic features control streamflow at different flow conditions, this method allowed the final rating curve to account for these inflection points. Second, Bayesian methods allowed for quantification of uncertainty associated with rating curves. Having uncertainty associated with a streamflow value is important as many ecohydrologists use streamflow values as absolute although it is known there is almost always uncertainty associated with rating curves¹⁹. Throughout this paper we will use “predicted streamflow” to refer to the NEON Continuous Discharge dataset.

Manual measurements of streamflow used to construct NEON’s rating curves were collected via acoustic Doppler current profilers (ADCP)²³, conservative salt tracer releases²⁴, and flow meter measurements²⁵. It is important to note, each of these methods are used under different conditions and are subject to unique types of error, see citations for details on how the methods are used to calculate streamflow. NEON then used streamflow measurements to construct unique rating curves for each hydrologic year (starting on October 1st). Continuous inputs to these rating curve models were generated by converting continuous pressure transducer data to continuous stage (stream surface elevation) using linear regressions with up to 26 manually measured gauge heights per year^{15,26}.

To facilitate informed use of NEON streamflow estimates, we evaluated the reliability of discharge estimates for every site-month available by analyzing the quality of continuous stage estimates (rating curve inputs), rating curve fits, and the proportion of streamflow estimates extrapolated beyond field measurements. Using this information, we created a categorical classification set that binned site-months into tiers of reliability, and flagged data that was of lesser quality. Although NEON’s Bayesian approach to modeling discharge reports uncertainty, the Continuous Discharge product contained sixteen columns that related to various sources of uncertainty and nine columns of quality control flags in the expanded data product, making using this information effectively complicated.

Our evaluation dataset includes categorical classifications of NEON’s Continuous Discharge product that allows users to easily query tabular data for reliable discharge estimates by month at all 28 sites where NEON provides estimates of streamflow using commonly used metrics like Nash-Sutcliffe coefficients. The classification dataset includes flags for potential sensor drift at sites (“drift flag”), flags for weak relationships between stage and gauge height (“regression flag”), and a final three tier rating for the reliability of streamflow estimates. “Tier 1” corresponds to the data that is ready to use for most hydrologic analyses, “Tier 2” indicates data that may be unreliable at high streamflow values or has a weaker fit of stage and streamflow overall, and “Tier 3” identifies data that is likely unreliable and should only be used after close inspection. We hope this dataset can increase the informed use of NEON streamflow estimates for ecological questions at a continental scale.

Results

To facilitate use of the NEON Continuous Discharge dataset, we evaluated and classified site-months into various categories to inform users on potential uncertainty in stage measurements and rating curve fits. Based on the availability of NEON discharge data, we were able to consider 1,523 site-months of continuous streamflow. Our evaluation included multiple tests to determine the quality of streamflow data. The first part of our

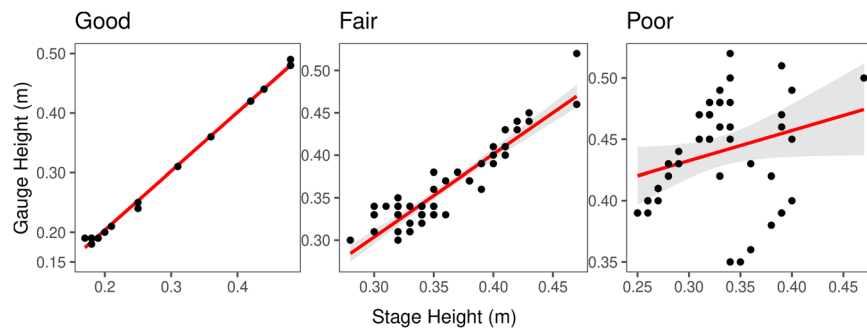


Fig. 2 Examples of NEON stage-gauge height regression falling into each of our classifications: Teakettle Creek (TECR) 2019 relationship (good), Lewis Run (LEWI) 2017-1 relationship (fair), and LeConte Creek (LECO) 2017-2 relationship (poor).

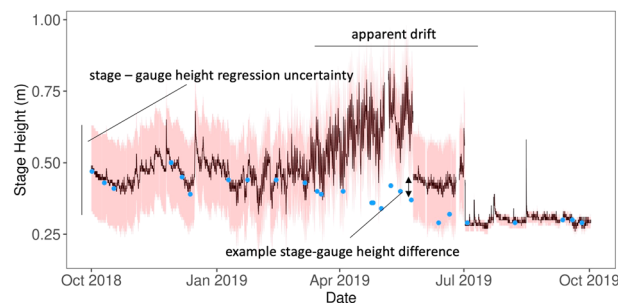


Fig. 3 Continuous gauge height at the Lewis Run (LEWI) site, continuous gauge height (black), uncertainty (red), and manual stage height (blue). Pressure transducers are subject to occasional drift, here we use manual measurements of stage to ensure pressure data is consistent with real stream stage. Uncertainty bounds are produced by NEON and are available in the Elevation of Surface Water product.

evaluation considered the quality of the continuous stage datasets which is used as the inputs for NEON's rating curves. NEON predicted continuous stage by converting continuous in-stream pressure transducer readings to water height using linear regressions parameterized with field measurements of water height. We calculated the Nash-Sutcliffe model efficiency coefficients (NSE) for these regressions and tested if continuous pressure transducer measurements of stage deviated from manual measurements of gauge height. Next, we evaluated the fit of rating curves used to convert this continuous stage to continuous discharge, by calculating a NSE for each model and the proportion of reported continuous discharge values extrapolated beyond field measurements.

Stage classification. Regression classification found 56% of site months with good regression ($NSE > 0.9$) between continuous stage and gauge height, 8% with fair regressions ($0.75 < NSE < 0.90$), 31% with poor regressions ($NSE < 0.75$), and 5% were not classified due to inadequate data. Figure 2 shows an example of a regression relationship falling into each of the three categories.

Drift detection. Comparing manual measurements of stage height to continuous gauge height and uncertainty ranges, allowed us to determine periods of potential sensor drift (Fig. 3). Drift detections found 57% of site-months likely had no drift, 8% of sites had potential drift, and the remaining 35% of site-months could not be assessed for drift because there were no manual measurements of stage height. If a site could not be assessed, it did not affect a site-month's classification but the inability to assess drift over the month is noted in the evaluation dataset.

Rating curve classification. Rating curves were classified based on the efficiency of the rating curves (NSE) and the amount of continuous discharge values extrapolated beyond field measurements. Rating curve classification found 67% of site-months with Tier 1 rating curves, 19% with Tier 2 rating curves, and 17% with Tier 3 rating curves (see Table 1 and methods for description of the three tiers). Figure 4 shows an example of rating curves falling into each of the three tiers.

To determine the final classification of a site-month's quality, our workflow first asked if the input data to the rating curve was reliable. If the relationship between the stage and gauge height was unreliable (a fair or poor classification), we report a regression flag. Second, we checked if a site-month had potential drift. If the data did have potential drift, we reported a drift flag. Finally, if these tests found the input stage data were reliable and there was no detectable drift, then the rating curve classification is reported as the final quality assessment (Fig. 5, Panel A).

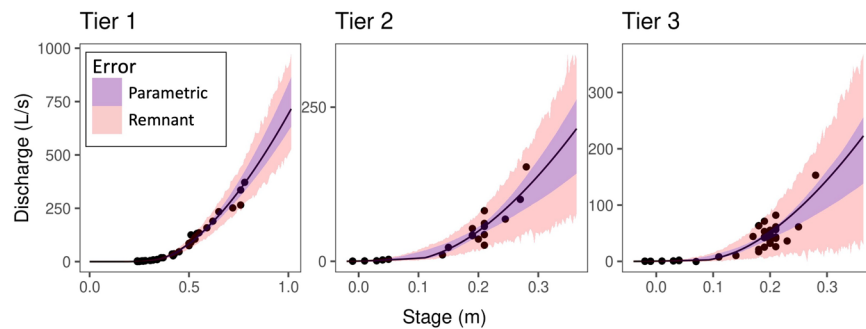


Fig. 4 Examples of NEON rating curves falling into each of our classifications: Como Creek (COMO) 2018 curve (Tier 1, NSE = 0.97), Arikaree River (ARIK) 2016 curve (Tier 2, NSE = 0.87), and ARIK 2017 curve (Tier 3, NSE = 0.74). Parametric error in purple and remnant error in red produced by NEON.

Of the 1,523 site-months of continuous streamflow predictions published by NEON, 38% were classified in our highest quality category of Tier 1. 11 of the 28 NEON sites had at least one full water year of Tier 1 streamflow data. Of all data classified in our analysis, site-months with unreliable inputs to the discharge models (predicted stage values) comprised the largest proportion of site-month classifications, with 38% of site months resulting in a regression flag (Fig. 5).

Figure 5 outlines how data moves through our classification scheme and the final results for each site-month.

Discussion

Our goal with this formal evaluation of the NEON Continuous Discharge product was to enable researchers to rapidly identify high quality data that could be used in subsequent analyses and to reduce the likelihood that consumers of this dataset invest effort in conducting analyses with unreliable data. Streamflow data that was classified as Tier 1 met the high professional standards of the hydrology community and can be used in all subsequent analysis with confidence. This accounted for 39% of the data released by NEON. More than half of the data provided by NEON should be used with caution.

About 11% of the published NEON streamflow data were classified as Tier 2 which means that more than 15% of reported values are beyond the maximum manual measurement of streamflow, or that the fit between the rating curve and stage measurements was weaker than for Tier 1 (Nash Sutcliffe coefficients 0.75–0.9). For data in this tier, low to intermediate flow data is likely reliable, but users are discouraged from calculating annual fluxes of solutes or gases from data where stormflows may be significantly over or underestimated, or modeling metabolism or hydrologic exchange during periods of higher flow. These cautions are repeated and made stronger for the ~8% of the dataset that was classified as Tier 3 because more than 30% of their reported streamflow values were beyond the maximum manual estimate of discharge or because the fit between the rating curve and stage measurements were poor (Nash Sutcliffe <0.75).

Nearly half (42% of the data) were not ranked into one of these three tiers because of poor fits between stage and gauge height measurements that caused us to question the quality of the data. There are a variety of reasons that might lead to the large uncertainty around both readings (e.g., field error, sensor failure, movement of sensors or gauges without record, channel reshaping following floods) but we have insufficient information with which to identify or correct for any of these potential errors. We have imposed Tier rankings in our evaluation, but the quality of the NEON discharge product spans a continuous gradient. Users are encouraged to set their own criteria for use, and all of the evaluative metrics are available to assist them in doing so.

Although we determined that a large proportion of the streamflow data currently provided by NEON are unreliable, some sites provide more than a year of reliable data that can be used, and the extent of reliable data has been increasing over time. We urge NEON site managers and field crews to devote increased attention to collecting manual discharge estimates at high and low flows to improve their rating curves and increase the availability of highly reliable data in the future. We encourage all prospective users of NEON discharge data to engage in efforts to improve upon our current effort. For example, efforts to estimate discharge in ungauged basins may prove useful in supplementing and refining the unreliable data collected for the early years of NEON. Improvements in the NEON Continuous Discharge product will offer many exciting opportunities to combine an unprecedented set of hydrologic, biogeochemical, biological, and ecological data from streams across the United States. Our intention with this paper is to enable those exciting analyses to occur.

Methods

As part of the streamflow data release, NEON released four relevant data products: Gauge Height²⁶, Elevation of Surface Water²⁹, Stage-discharge Rating Curves³⁰, and Continuous Discharge¹⁵. Data users are able to download this full suite of information and protocols to inform decisions on data usage and applicability. We evaluated the quality of the Continuous Discharge product using all four relevant NEON data products, considering the validity of model inputs as well as the goodness-of-fit of final streamflow estimates. We analyzed 1) the fit of the regression between manual stage height readings and continuous pressure transducer data used to estimate continuous stream surface elevation, 2) the fit of rating curves transforming stream surface elevation to streamflow, and 3) the proportion of streamflow estimates over the maximum manually-measured streamflow.

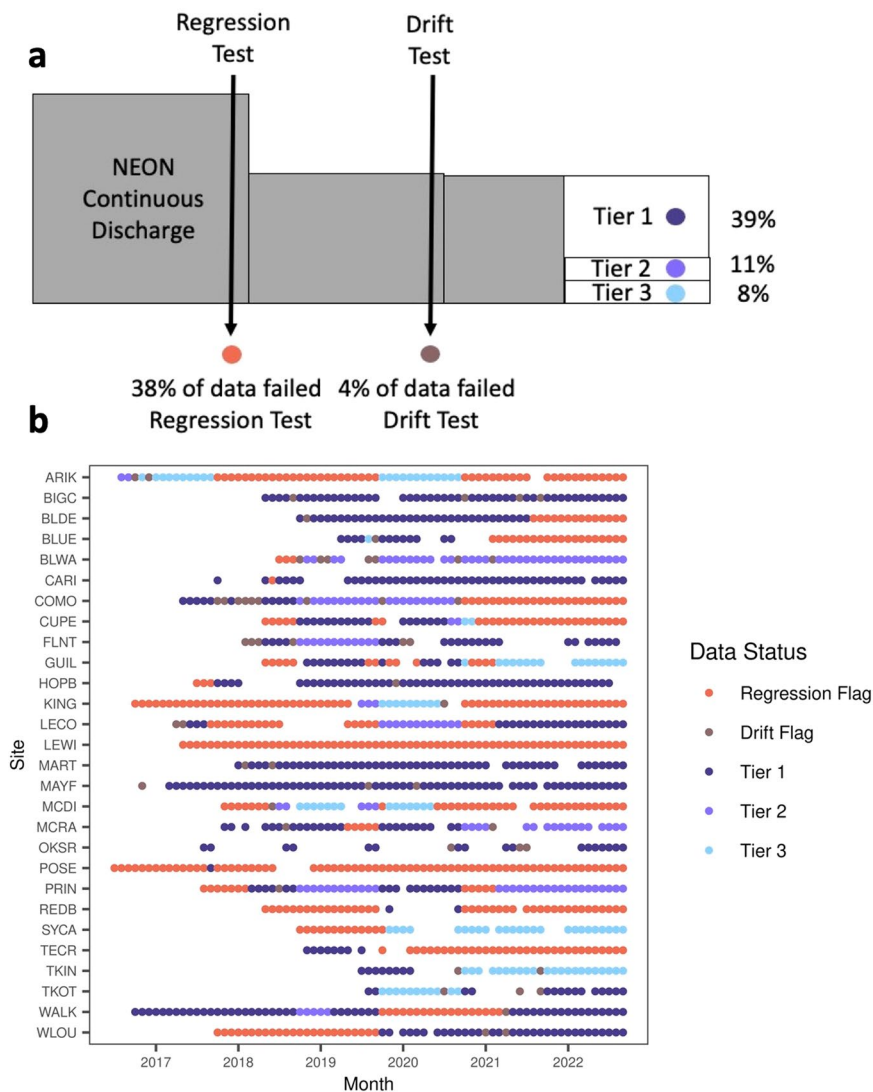


Fig. 5 Panel a, conceptual diagram depicting how data is partitioned into final evaluation category with the percentage of site-months with each flag or classification. Panel b, NEON monthly site classification for stage-gauge regressions, drift, and rating curves. Points colored by rating curve classification (Tier 1, 2, 3) if the regression status is good and no drift was detected in that month, otherwise colored red or brown to indicate flags. Missing points indicate no data is available for that month.

Stage classification. The rating curve models predicting streamflow required continuous stream stage estimates as model inputs. NEON predicted continuous gauge height with a two step approach. First, continuous in-stream transducer readings were converted to water height by applying an offset between the transducer elevation and the staff gauge (Eq. 1). This offset is derived from the NEON geolocation database as the difference between the location of the pressure transducer and the staff gauge²⁷. The offset changes only when the location of either the staff gauge or transducer moves.

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

Conversion of pressure data to water height used by NEON²⁷ where h_{wc} is the estimated water column height (m), P_{sw} is calibrated surface water pressure (kPa), p is the density of water (999 kg/m^3), g is the acceleration due to gravity (9.81 m/s^2), and h_{stage} is the offset between the pressure transducer and the staff gauge (m).

Then, NEON uses a linear regression between manually-measured reference stage height and the calculated gauge height from Eq. 1, yielding final predictions of continuous stream gauge height²⁷. In an ideal setting, stage and gauge height should correlate perfectly²⁸. In the field, sensor uncertainty, manual reference measurement error, and shifting conditions in the stream can convolute the relationship. We tested the goodness of fit between continuously estimated stream gauge height values and manual stage measurements using the Nash-Sutcliffe model efficiency coefficient (Eq. 2). Nash-Sutcliffe coefficient is a commonly used metric in hydrology used to

Classification	Nash-Sutcliffe Coefficient	Extrapolated beyond maximum gauging
Tier 1	$NSE > 0.9$	<15%
Tier 2	$0.75 < NSE < 0.9$	<30%
Tier 3	$NSC < 0.75$	>30%

Table 1. Rating curve classification criteria.

evaluate how well a model performed relative to observed values (manually measured stage and calculated gauge height). For the purposes of this discussion, manual reference measurements will be referred to as ‘stage’ and automated, sensed readings as ‘gauge height’.

$$NSE = 1 - \frac{\sum(Q_o - Q_m)^2}{\sum(Q_o - \bar{Q}_o)^2} \quad (2)$$

Equation 2 presents Nash-Sutcliffe model efficiency coefficient, where Q_o is an observed value (streamflow or stage height), Q_m is a modeled value, and \bar{Q}_o is the mean of observed values.

Stage, gauge height, and regression data were sourced from the NEON Continuous Discharge product, representing what was directly applied to streamflow estimation. Up to 26 stage measurements were available per year. We examined every regression between stage and gauge height (one per site year in which data was available) and classified each as either ‘good’, ‘fair’, or ‘poor’ quality based on their goodness of fit. Regressions with a NSE (Eq. 2) of 0.90 or greater were considered good, those with a NSE of less than 0.90 but greater than or equal to 0.75 were considered fair, and those with an NSE of less than 0.75 were considered poor (Fig. 2).

Drift detection. Because electronic instruments, such as pressure transducers, can have systematic directional drift, referred to as ‘drift’, during deployment, we developed an approach to detect periods of time when NEON’s Elevation of Surface Water product drifted. We used two methods to assess and flag the potential for instrument drift at monthly time steps. First, we flagged any period the manually measured stage fell outside NEON’s uncertainty bound for gauge height made at the same time. From this, we calculated the proportion of stage measurements outside of the gauge height uncertainty bounds per month. This proved to be a relatively lenient filter that missed periods of manually identified drift. We found adding a second filter that flagged any month where the difference between the manually measured stage and gauge height exceeded 6 cm, was effective in catching the majority of periods where drift was identified. Second, we calculated the average differences between stage and gauge height for each month (Fig. 3). To determine appropriate cut-off values to classify areas of potential drift, we manually audited and flagged periods of observable directional drift. Our goal was to set a maximum cut-off difference which retained as much usable data as possible while still capturing 70% of the manually flagged directional drift periods. Applying this method, we determined a cut-off value of 6 cm average monthly deviation between observed and predicted stage values.

Using these two filters in combination, we again classified data into three groups: ‘likely no drift’, ‘potential drift’, and ‘not assessed’. Site-months with no more than 50% of stage measurements outside of the gauge height time series uncertainty and an average difference between stage and gauge height less than 6 cm were considered to have ‘likely no drift’. Site-months with either more than 50% of stage readings outside of the gauge height time series uncertainty or an average difference between stage and gauge height more than 6 cm were deemed to have ‘potential drift’. Site-months with no stage measurements could not be evaluated and were considered ‘not assessed’. Although this approach to identify drift is imperfect, in that slight drift could be missed and times without manual measurements are not possible to assess, we believe this is a helpful method given the data available from NEON and the fact drift has been observed when visually inspecting data (Fig. 3).

Rating curve classification. To evaluate how well rating curves predicted streamflow, we assessed each rating curve used to convert stage to discharge. NEON prepares a new rating curve for each site’s water year (beginning on October 1st)²⁷. In cases where NEON reported multiple rating curves for a site’s water year each curve was assessed separately across the time series which it was used. We classified rating curves into three tiers based on two metrics: the Nash-Sutcliffe coefficient (Eq. 2) between observed and predicted streamflow, and the percentage of continuous discharge values above the maximum manually measured gauging used to construct the rating curve.

First, we calculated the Nash-Sutcliffe coefficient for each rating curve to estimate how well rating curves captured the variation in the stage-streamflow relationship. We used the reported values for modeled and manually measured streamflow from the ‘Y1simulated’ and ‘Y1observed’ columns in the ‘sdrcl_resultsResiduals’ table of the Stage-discharge rating curves product. NEON generally conducts between 12 and 24 manual gaugings per year to build and maintain the stage-discharge relationship.

Second, we calculated the percentage of continuous streamflow values outside the range of manually measured estimates of streamflow. This was useful to assess if the stage-discharge relationship is representative of observed flow conditions. The relationship between discharge and stage is often nonlinear, with inflection points around changes in channel morphology making gauging the stream at high and low flow conditions critical to building a reliable rating curve¹⁶. A rating curve based on a large number of direct field measurements all taken during a narrow range of baseflows, for example, could generate a rating curve with a high Nash-Sutcliffe coefficient that is unreliable when extrapolated to high or low flow events. Using these two metrics, we were able to

classify rating curves into categories of relative quality. To calculate the percentage of values in the continuous streamflow product that fall outside the range of manually gauged streamflow values, we extracted the maximum and minimum gauging values from the 'sdr_c_resultsResiduals' table in the Stage-discharge Rating Curve product. We then compared the predicted values derived from each rating curve (as reported in the 'csd_continuousDischarge' table) to the extracted range and calculated the proportion of values which fell outside of it.

We used the Nash-Sutcliffe coefficient and percentage of streamflow values over the maximum observed field measurements to classify rating curves into three categories outlined in Table 1.

To integrate stage-gauge regressions, drift detections, and rating curve classification, we produced a summary table with classifications for all three tests and the corresponding metrics used in each classification (Fig. 5). The table is grouped by month and site so users can query sites and determine which months have the appropriate data for their needs.

Data availability

The results of this analysis are available at the cited HydroShare repository³¹.

All data evaluated in this publication are collected by the National Ecological Observation Network and made available through NEON's web portal. Data was accessed using the loadByProduct function in the R Package neonUtilities³². We evaluated the continuous discharge product¹⁵ and Stage-discharge rating curve product³⁰ for use in various stream ecology analyses. The RELEASE-2022 portions of the evaluation datasets are available at the cited refs. ^{15,30}. The PROVISIONAL portions of the evaluation datasets were accessed on 10/12/2022 from <https://data.neonscience.org/data-products/DP4.00130.001> and <https://data.neonscience.org/data-products/DP4.00133.001>, however these are no longer available as per NEON policy.

Code availability

All code used in this analysis is available in the cited github repository³³. This has been annotated to make it easy for users to see what portion of the NEON stream discharge data product is being used and to easily alter the code to analyze subsets of the data specific to their intended use. Code was run using the neonUtilities R package version 2.1.2³² and R version 4.1.2³⁴.

Received: 21 April 2022; Accepted: 24 January 2023;

Published: 11 February 2023

References

- Poff, N. L. *et al.* The Natural Flow Regime. *BioScience* **47**, 769–784 (1997).
- Summers, B. M. *et al.* Long-term data reveal highly-variable metabolism and transitions in trophic status in a montane stream. *Freshwater Science* **39**, 241–255 (2020).
- Stanley, E. H. *et al.* The ecology of methane in streams and rivers: patterns, controls, and global significance. *Ecological Monographs* **86**, 146–171 (2016).
- Liu, S. & Raymond, P. A. Hydrologic controls on pCO₂ and CO₂ efflux in US streams and rivers. *Limnology and Oceanography Letters* **3**, 428–435 (2018).
- Lloyd, C. E. M., Freer, J. E., Johnes, P. J., Coxon, G. & Collins, A. L. Discharge and nutrient uncertainty: implications for nutrient flux estimation in small streams. *Hydrological Processes* **30**, 135–152 (2016).
- Peel, M. C. & McMahon, T. A. Historical development of rainfall-runoff modeling. *WIREs Water* **7**, e1471 (2020).
- Restrepo, J. C. *et al.* Sediment Transport and Geomorphological Change in a High-Discharge Tropical Delta (Magdalena River, Colombia): Insights from a Period of Intense Change and Human Intervention (1990–2010). *Journal of Coastal Research* **32**, 575–589 (2016).
- Guzha, A. C., Rufino, M. C., Okoth, S., Jacobs, S. & Nóbrega, R. L. B. Impacts of land use and land cover change on surface runoff, discharge and low flows: Evidence from East Africa. *Journal of Hydrology: Regional Studies* **15**, 49–67 (2018).
- Appling, A. P., Hall Jr., R. O., Yackulic, C. B. & Arroita, M. Overcoming Equifinality: Leveraging Long Time Series for Stream Metabolism Estimation. *Journal of Geophysical Research: Biogeosciences* **123**, 624–645 (2018).
- Hall, R. O. Metabolism of Streams and Rivers. In *Stream Ecosystems in a Changing Environment* 151–180, <https://doi.org/10.1016/B978-0-12-405890-3.00004-X> (Elsevier, 2016).
- Raymond, P. A. *et al.* Global carbon dioxide emissions from inland waters. *Nature* **503**, 355–359 (2013).
- United States Geological Survey. Water Data for the Nation.
- Appling, A. P., Leon, M. C. & McDowell, W. H. Reducing bias and quantifying uncertainty in watershed flux estimates: the R package loadflex. *Ecosphere* **6**, 1–25 (2015).
- Hotchkiss, E. R. *et al.* Sources of and processes controlling CO₂ emissions change with the size of streams and rivers. *Nature Geosci* **8**, 696–699 (2015).
- National Ecological Observatory Network. Continuous Discharge (DP4.00130.001), *National Ecological Observatory Network (NEON)* <https://doi.org/10.48443/xz4k-5j04> (2022).
- Kennedy, E. J. *Discharge ratings at gaging stations: U.S. Geological Survey Techniques of Water Resources Investigations.* (1989).
- Tuozzolo, S. *et al.* The impact of reach averaging Manning's equation for an *in-situ* dataset of water surface elevation, width, and slope. *Journal of Hydrology* **578**, 123866 (2019).
- Kratzert, F., Klotz, D., Brenner, C., Schulz, K. & Herrnegger, M. Rainfall-runoff modelling using Long Short-Term Memory (LSTM) networks. *Hydrol. Earth Syst. Sci.* **22**, 6005–6022 (2018).
- Domeneghetti, A., Castellarin, A. & Brath, A. Assessing rating-curve uncertainty and its effects on hydraulic model calibration. *Hydrol. Earth Syst. Sci.* **16**, 1191–1202 (2012).
- Schmidt, A. R. Analysis of stage-discharge relations for open-channel flows and their associated uncertainties. (University of Illinois at Urbana-Champaign).
- Baldassarre, G. D. & Claps, P. A hydraulic study on the applicability of flood rating curves. *Hydrology Research* **42**, 10–19 (2011).
- Le Coz, J., Renard, B., Bonnifait, L. & Branger, F. & Le Boursicaud, R. Combining hydraulic knowledge and uncertain gaugings in the estimation of hydrometric rating curves: A Bayesian approach. *Journal of Hydrology* **509**, 573–587 (2014).
- Moore, S. A., Jamieson, E. C., Rainville, F., Rennie, C. D. & Mueller, D. S. Monte Carlo Approach for Uncertainty Analysis of Acoustic Doppler Current Profiler Discharge Measurement by Moving Boat. *Journal of Hydraulic Engineering* **143**, 04016088 (2017).
- Tazioli, A. Experimental methods for river discharge measurements: comparison among tracers and current meter. *Hydrological Sciences Journal* **56**, 1314–1324 (2011).

25. Pantelakis, D., Doulgeris, C., Hatzigiannakis, E. & Arampatzis, G. Evaluation of discharge measurements methods in a natural river of low or middle flow using an electromagnetic flow meter. *River Research and Applications* **38**, 1003–1013 (2022).
26. National Ecological Observatory Network. Gauge height (DP1.20267.001). *National Ecological Observatory Network (NEON)* <https://doi.org/10.48443/80nt-8v02> (2022).
27. Cawley, K. M., Harrison, N., Nickerson, Z. & Hensley, B. Neon user guide to continuous discharge (DP4.00130.001). *National Ecological Observatory Network (NEON)* 16.
28. Freeman, L. A. *et al.* Use of submersible pressure transducers in water-resources investigations. <https://pubs.usgs.gov/twri/twri8a3/>.
29. National Ecological Observatory Network. Elevation of surface water (DP1.20016.001). *National Ecological Observatory Network (NEON)* <https://doi.org/10.48443/1p13-c355> (2022).
30. National Ecological Observatory Network. Stage-discharge rating curves (DP4.00133.001). *National Ecological Observatory Network (NEON)* <https://doi.org/10.48443/02pm-1t67> (2022).
31. Rhea, S. NEON Continuous Discharge Evaluation. *HydroShare* <https://doi.org/10.4211/hs.03c52d47d66e40f4854da8397c7d9668> (2023).
32. Lunch, C. neonUtilities: Utilities for Working with NEON Data. (2021).
33. Rhea, S. NEON Continuous Discharge Evaluation, code. *Zenodo* <https://doi.org/10.5281/zenodo.7563088> (2023).
34. R Core Team. R: A language and environment for statistical computing. (2021).

Acknowledgements

We thank the National Ecological Observatory Network, Bobby Hensley, and Kaelin Cawley for assistance in interpreting data products. We also thank Mike Vlah, Audrey Thellman, and Nicholas Marzolf for their guidance with this manuscript. This work was supported by three NSF grants administered through the Macrosystems program: 2139710 (to ESB), 1926420 (to ESB and MVR), and 2106071 (to ESB and AGD).

Author contributions

S.R. contributed to the conceptualization, analysis, and writing. N.G. contributed to the conceptualization, analysis, and writing. A.G.D. contributed to the conceptualization, analysis, and writing. M.R.R. contributed to the conceptualization and writing. E.S.B. contributed to the conceptualization and writing.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023, corrected publication 2023