A simple parameterization for segmented rating curves

T. O. Hodson[1], T. A. Kenney[2], T. M. Over[1], and M. B. Yeheyis[3]

[1]U.S. Geological Survey Central Midwest Water Science Center, Urbana, Illinois
[2]U.S. Geological Survey Water Resources Mission Area, West Valley City, Utah
[3]Environment and Climate Change Canada, Gatineau, Quebec

Key Points:

- Rating curves are widely used in hydrology to generate streamflow timeseries
- However, most rating curves are still fit manually
- This paper proposes a reference parameterization for automating that fitting

Corresponding author: Timothy Hodson, thodson@usgs.gov

Abstract

Streamflow is one of the most important variables in hydrology but is difficult to measure continuously. As a result, nearly all streamflow time series are estimated from rating curves that define a mathematical relationship between streamflow and some easy-to-measure surrogate like water-surface elevation (stage). Most ratings are still fit manually, which is time-consuming and subjective. To improve that process, the U.S. Geological Survey (USGS), among others, is evaluating algorithms to automate that fitting. Several automated methods already exist, and each parameterizes the rating curve slightly differently. Because of the nonconvex nature of the problem, those differences can greatly affect performance. After some trial and error, we settled on reparameterizing the classic segmented power law somewhat like a Bayesian physics-informed neural network. Being physics-informed and Bayesian, the algorithm requires minimal data and also estimates uncertainty. Its implementation is open source and easily modified so that others can contribute to improving the quality of USGS streamflow data.

1 Introduction

Streamflow time series are widely used in hydrologic research, water resource management, engineering design, and flood forecasting, but they are difficult to measure directly. In nearly all time-series applications, streamflow is estimated from rating curves or "ratings" that describe the relation between streamflow and an easy-to-measure surrogate like stage. The shape of the rating is specific to each streamgage and is governed by channel conditions at or downstream from the gage, referred to as controls. Section controls, like natural riffles or artificial weirs, occur downstream from the gage, whereas channel controls, like the geometry of the banks, represent conditions along the stream reach (the upstream and downstream vicinity of the gage). Regardless of the type, the behavior of each control is often well-approximated with standard hydraulic equations that take the general form of a power law with an offset parameter

$$q = C(h - h_0)^b \tag{1}$$

where q is the discharge (streamflow); h is the height of the water above some datum (stage); h_0 is the stage of zero flow (the offset parameter); $(h - h_0)$ is the hydraulic head; b is the slope of the rating curve when plotted in log-log space; and C is a scale factor equal to the discharge when the head is equal to one (ISO 18320:2020, 2020). When multiple controls are present, the rating curve is divided into segments with one power law corresponding to each control resulting in a multi-segment or compound rating.

Although automated methods exist, most ratings are still fit manually using a graphical method of plotting stage and discharge in log-log space. With the appropriate location parameter, each control can be fit to a straight-line segment in log space (Kennedy, 1984; ISO 18320:2020, 2020). Variants of this method have been used for decades, first with pencil and log paper and now with computer-aided software; the fitting is still done by manually adjusting parameters until an acceptable fit is achieved.

Single-segment ratings are relatively easy to fit by automated methods (Venetis, 1970), but compound ratings are more challenging because their solution is nonconvex or multimodal (Reitan & Petersen-Øverleir, 2006). As a result, optimization algorithms can become stuck in local optima and fail to converge to the global optimum. General function approximators, such as natural splines (Fenton, 2018) or neural networks, can be easier to fit but their generality comes at a cost. The form of the power law matches that of the hydraulic equations governing uniform open-channel flow, like the Manning equation (Manning, 1891). Due to that physical basis, power laws are potentially more robust than other generic curve-fitting functions: requiring less data to achieve the same fit and being less prone to overfitting.

This paper describes a basic algorithm for fitting a segmented power law with offset parameters—the classic rating-curve form— and compares its performance against a natural cubic spline; "natural" indicates the second derivatives are set to zero at the endpoints. Several algorithms for fitting segmented power laws already exist. Some are more physical, meaning their structure corresponds to the governing hydraulic equations (Reitan & Petersen-Øverleir, 2008; Le Coz et al., 2014); some are more data-driven with more flexible structures like spline (Fenton, 2018) or local regression (Coxon et al., 2015); and some are a hybrid of the two (Hrafnkelsson et al., 2021). Each approach has different tradeoffs. More physical approaches require less data but may be nonconvex, which makes them challenging to fit, whereas data-driven approaches are easier to fit but require more data to constrain their greater flexibility.

Like our algorithm, the more physical approaches tend to be Bayesian and use sampling (as opposed to optimization; Ma et al., 2019) and priors to help mitigate problems that occur in nonconvex settings. Examples of priors include constraining the exponent b to be around 5/3, constraining the number of rating segments, or constraining the transitions between segments around a particular stage. Being Bayesian, these algorithms inherently estimate uncertainty in the fitted parameters and discharge, which is important for many applications. However, they differ in their exact parameterization, and because of their nonconvex nature, slight differences can greatly affect performance.

Our algorithm distinguishes itself in two ways: its simple implementation, which uses a community-developed open-source probabilistic programming library, and its robust parameterization. These two aspects are in a way interrelated. With a community-developed library, the underlying numerical code is maintained by a broader community of developers, so instead of developing that code, we could focus on testing different parameterizations for the rating curve.

2 Implementation

We implemented our curve-fitting algorithm, along with some additional plotting and evaluation tools, test datasets, and tutorials, as a Python package called `ratingcurve`. The algorithm uses `PyMC` (Salvatier et al., 2016), an open-source Python library for Bayesian statistical modeling and probabilistic machine learning. Using `PyMC`, the core algorithm can be expressed in several lines of code, making it easier to extend or modify, like changing the priors, the parameterization of the rating curve, or the inference algorithm to achieve different tradeoffs of speed and accuracy. This paper demonstrates two such algorithms: Automatic Differentiation Variational Inference (ADVI) (Kucukelbir et al., 2017) and Hamiltonian Monte Carlo with the No-U-Turn Sampler (NUTS) (Hoffman & Gelman, 2014). ADVI is an optimization algorithm, whereas NUTS is a sampling algorithm. In general, sampling is slower but better for nonconvex problems (Ma et al., 2019).

2.1 Usage

Refer to the *Open Research Section* for links to the source code repository and packaged versions of `ratingcurve`. Given observations of discharge (q), stage (h), and, optionally, the standard error of the discharge observations (e), a two-segment rating is fit with

```
rating = PowerLawRating(q, h, e, 2)
trace = rating.fit()
rating.plot(trace)
```

and produces a plot like Figure 1.

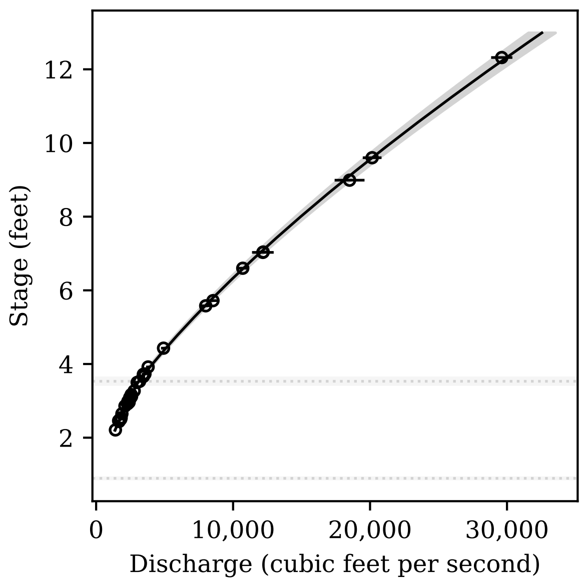

Figure 1. Two-segment rating curve, with 95-percent prediction interval, for the Green River near Jensen, Utah (U.S. Geological Survey streamgage 09261000); generated by `rating.plot(trace)` and fit using ADVI. The circles with error bars show the observations and their uncertainty. Horizontal dotted lines show the segment breakpoints and their prediction intervals.

A rating curve can also be exported as a table for use by other applications, shown in Table 1. In addition to the mean discharge for each stage, the table gives the median and geometric standard error (GSE), which is multiplied and divided by the median to estimate prediction intervals (Limpert et al., 2001).

Table 1. Rating table generated by `rating.table(trace)`. Units are feet (ft) and cubic feet per second (ft^3 s^{-1}); geometric standard error (GSE) is a unitless factor.

Stage	Mean Discharge	Median Discharge	GSE
ft	ft^3 s^{-1}	ft^3 s^{-1}	-
2.20	1376.14	1376.16	1.0107
2.21	1388.27	1388.27	1.0107
2.22	1400.41	1400.40	1.0107
2.23	1412.57	1412.55	1.0106
2.24	1424.74	1424.73	1.0106
…	…	…	…

2.2 Parameterization

Our rating curve algorithm is similar to the segmented power law used in the manual method (Kennedy, 1984; ISO 18320:2020, 2020), as well as in automated approaches (Reitan & Petersen-Øverleir, 2008; Le Coz et al., 2014), but differs in its parameterization. Conceptually, the Reitan and Petersen-Øverleir (2008) parameterization slices the channel cross section horizontally to form each segment: segments stack one on top of the other. Once the stage rises beyond the range of a particular control, that control is "drowned out" and flow through that segment ceases to increase with stage. Our parameterization slices the channel cross-section vertically, so controls never drown out. The Le Coz et al. (2014) parameterization can slice in either direction but differs in that the segments are summed after transforming them back to their original scale; whereas, Reitan and Petersen-Øverleir (2008) sum the segments in log.

After testing several parameterizations, one seemed especially reliable and simple: slicing the channel cross section vertically into control segments and summing them in log, which is somewhat like a ReLU (rectified linear unit) neural network with hydraulic controls as neurons

$$X = \ln(\max(h - h_s, 0) + h_o) \tag{2}$$

$$\ln(q) = a + b^T X + \epsilon + \epsilon_o \tag{3}$$

where h_s are the unknown segment breakpoints; h_o is a vector of offsets, the first is 0 and the rest are 1; max is the element-wise maximum, which returns a vector of size h_s; a is a bias parameter equal to $\log(C)$, the scale factor; b are the slopes of each log-transformed segment; ϵ is the residual error; and ϵ_o is the uncertainty in the discharge observations (optional). The offset vector h_o ensures that $X \geq 0$, so additional segments never subtract discharge.

The default priors and settings are documented in the package; in general, they do not need to be modified. Besides selecting the number of segments, the user can specify a prior distribution on the breakpoints. The default assumes the breakpoints are monotonically ordered and uniformly distributed across the range of the data, $h_{s1} < \min(h) < h_{s2} < \cdots < h_{sn} < \max(h)$. Alternatively, the user can specify approximate locations for each breakpoint and their uncertainty as normal distributions.

Uncertainty in the discharge observations is typically reported as relative standard error (RSE). For convenience, we convert that relative error to a geometric error as $\epsilon_o \sim N(0, \ln(1{+}\mathrm{RSE}/q)^2)$. For small uncertainties, the difference is negligible, and for large uncertainties, it is not known which is better. Like Reitan and Petersen-Øverleir (2008), we assume ϵ is normally distributed with mean zero and variance σ^2, $\epsilon \sim N(0, \sigma^2)$. That simplification can create unaccounted heteroscedasticity (Petersen-Øverleir, 2004) but generally yields a reasonable estimate for the rating and its uncertainty.

3 Results

We compared the performance of the segment power law against a log-transformed natural spline on a simulated 3-segment rating curve. Both models use log transformations, which helps with heteroscedasticity. The power law is strictly increasing; otherwise, both approaches use log transformations, and both are flexible enough to approximate a wide variety of functions. Unlike the spline, the parameters in the power law have physical meaning in that they correspond to parameters in standard hydraulic equations for approximating open-channel flow. The segmented power law is notoriously difficult to calibrate (Reitan & Petersen-Øverleir, 2008), however, and its performance depends, in large part, on its parameterization —we tested several, some mathematically equivalent, some slicing the cross section vertically or horizontally— as well as its priors. If the calibration challenges are overcome, the power law should yield better fits with fewer observations (Reitan & Petersen-Øverleir, 2008).

Figure 2 shows a side-by-side comparison of a spline and power law fit with 6, 12, 24, and 48 stage-discharge observations. For best accuracy, the curves were fit using NUTS. We also specified that the power law had 3 segments and that the spline had 8 degrees of freedom, the same as the power law (1 bias, 3 offsets, 3 slopes, and 1 uncertainty). Otherwise, default settings were used for both.

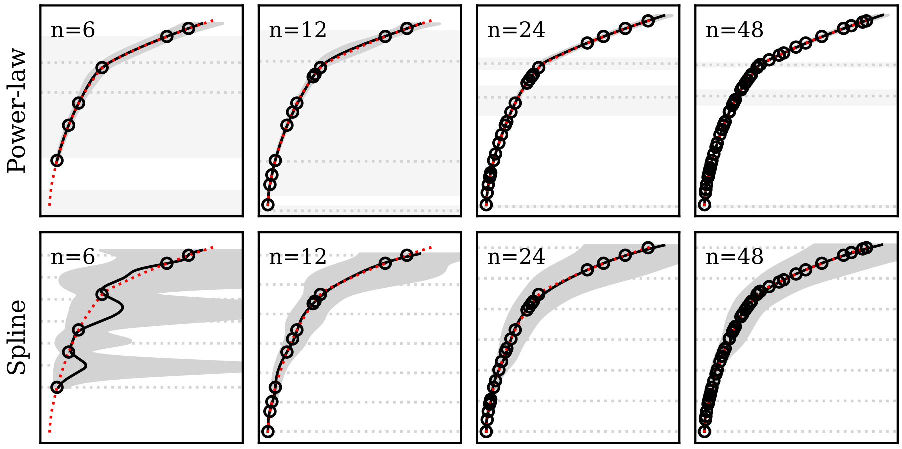

Figure 2. Segmented-power law (top) and natural spline (bottom) fit with different numbers of observations (n). The dashed red line is the true rating and the circles are the simulated observations. Horizontal dotted lines show segment breakpoints and knot locations for the power law and spline, respectively. The shaded regions depict 95 percent prediction intervals for the rating and breakpoints.

Although the natural spline was 5-20x faster, it yielded poorer fits, particularly when $n = 6$. Reducing the degrees of freedom might improve performance when $n = 6$ but also sacrifices flexibility when $n = 48$. By comparison, the power law yielded a good fit with six observations— two fewer than the number of model parameters. Our intent is not to disparage all splines— both parameterizations are technically splines. Rather, we wanted to demonstrate a classic tradeoff between being easy to fit or being accurate, which is a characteristic of data-driven and physical approaches.

In general, the accuracy of data-driven approaches is highly dependent on the avaiability of data. For example, Coxon et al. (2015) recommends a minimum of twenty stage-discharge measurements for their data-driven approach. Taken over the lifetime of a streamgage, twenty may be manageable. However, ratings shift through time from erosion, deposition, vegtation growth, debris/ice jams, etc. (Herschy & Herschy, 2014; Mansanarez et al., 2019), and it may be impracticable to collect twenty measurements between each shift. In that case, a more physical approach like the power law may be a better choice, because they require fewer observations.

This paper focuses on one parameterization of the classic multi-segment power law, but undoubtably more will emerge, which might achieve better tradeoffs of ease and accuracy. For example, our comparison uses NUTS, which is accurate but slow. With 6 observations, NUTS fit the 3-segment power law in around 10 minutes. With 48 observations, NUTS completed in 1 minute; a 10x speedup. In general, stronger priors, more observations, or fewer segments would reduce that time. By comparison, ADVI generally achieved a NUTS-like fit in several seconds, but it occasionally failed to converge on the optimum solution. A better parameterization might yield better convergence with a faster inference algorithm.

4 Conclusions

Despite the existence of automated methods, most stage-discharge rating curves are still fit manually. Although the governing hydraulic equations are relatively simple and well-understood, they are notoriously difficult to solve for multiple controls. Among the automated methods, no parameterization has emerged as the standard, and functionally equivalent parameterizations may vary greatly in performance.

Here, we implement a simple parameterization that works well with minimal data and prior information. Notably, it does not address shifts in the rating curve through time or hysteresis, and the curve is continuous but not smooth (twice differentiable). Such limitations could be addressed, and any such effort will depend, in part, on building from a good starting parameterization. Therefore, our simple-yet-reliable parameterization, use of a community-developed probabilistic programming library, and packaging provide a benchmark for operationalizing automated methods that could promote more widespread use, testing, and refinement by the hydrologic community.

Open Research Section

The latest version of `ratingcurve` is available at `https://github.com/thodson-usgs/ratingcurve`, as well as `https://code.usgs.gov/wma/uncertainty/ratingcurve`. Packaged versions are available via PyPI and conda-forge. A link to the official release of the version used in this paper will appear here in due course.

Acknowledgments
This research was funded by the U.S. Geological Survey National Hydrologic Monitoring Program. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References

Coxon, G., Freer, J., Westerberg, I. K., Wagener, T., Woods, R., & Smith, P. J. (2015, jul). A novel framework for discharge uncertainty quantification applied to 500 UK gauging stations. *Water Resources Research*, *51*(7), 5531–5546. Retrieved from `https://doi.org/10.1002%2F2014wr016532` doi: 10.1002/2014wr016532

Fenton, J. D. (2018, sep). On the generation of stream rating curves. *Journal of Hydrology*, *564*, 748–757. Retrieved from `https://doi.org/10.1016%2Fj.jhydrol.2018.07.025` doi: 10.1016/j.jhydrol.2018.07.025

Herschy, R. W., & Herschy, R. (2014). *Streamflow measurement* (Third ed.). CRC Press. Retrieved from `https://doi.org/10.1201%2F9781482271485` doi: 10.1201/9781482271485

Hoffman, M. D., & Gelman, A. (2014). The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, *15*(47), 1593–1623. Retrieved from `http://jmlr.org/papers/v15/hoffman14a.html`

Hrafnkelsson, B., Sigurdarson, H., Rögnvaldsson, S., Örn Jansson, A., Vias, R. D., & Gardarsson, S. M. (2021, dec). Generalization of the power-law rating curve using hydrodynamic theory and bayesian hierarchical modeling. *Environmetrics*, *33*(2). Retrieved from `https://doi.org/10.1002%2Fenv.2711` doi: 10.1002/env.2711

ISO 18320:2020. (2020). *Hydrometry–measurement of liquid flow in open channels—determination of the stage-discharge relationship.* International Organization for Standardization.

Kennedy, E. J. (1984). *Discharge ratings at gaging stations, U.S. Geological Survey techniques of water-resources investigations 09-a10.* Retrieved from `https://doi.org/10.3133/twri03A10`

Kucukelbir, A., Tran, D., Ranganath, R., Gelman, A., & Blei, D. M. (2017). Automatic differentiation variational inference. *Journal of Machine Learning Research*, *18*(14), 1–45. Retrieved from `http://jmlr.org/papers/v18/16-107.html`

Le Coz, J., Renard, B., Bonnifait, L., Branger, F., & Le Boursicaud, R. (2014, Feb). Combining hydraulic knowledge and uncertain gaugings in the estimation of hydrometric rating curves: A bayesian approach. *Journal of Hydrology*, *509*, 573–587. Retrieved from `http://dx.doi.org/10.1016/j.jhydrol.2013.11.016` doi: 10.1016/j.jhydrol.2013.11.016

Limpert, E., Stahel, W. A., & Abbt, M. (2001). Log-normal distributions across the sciences: keys and clues. *BioScience*, *51*(5), 341. Retrieved from `http://dx.doi.org/10.1641/0006-3568(2001)051[0341:lndats]2.0.co;2` doi: 10.1641/0006-3568(2001)051[0341:lndats]2.0.co;2

Ma, Y.-A., Chen, Y., Jin, C., Flammarion, N., & Jordan, M. I. (2019, sep). Sampling can be faster than optimization. *Proceedings of the National Academy of Sciences*, *116*(42), 20881–20885. Retrieved from `https://doi.org/10.1073%2Fpnas.1820003116` doi: 10.1073/pnas.1820003116

Manning, R. (1891). On the flow of water in open channels and pipes. *Transactions of the Institution of Civil Engineers of Ireland*, *20*, 179–207.

Mansanarez, V., Renard, B., Coz, J. L., Lang, M., & Darienzo, M. (2019, Apr). Shift happens! adjusting stage-discharge rating curves to morphological changes at known times. *Water Resources Research*, *55*(4), 2876–2899. Retrieved from `http://dx.doi.org/10.1029/2018wr023389` doi: 10.1029/2018wr023389

Petersen-Øverleir, A. (2004, jun). Accounting for heteroscedasticity in rating curve estimates. *Journal of Hydrology*, *292*(1-4), 173–181. Retrieved from `https://doi.org/10.1016%2Fj.jhydrol.2003.12.024` doi: 10.1016/j.jhydrol.2003.12.024

Reitan, T., & Petersen-Øverleir, A. (2006, jul). Existence of the frequentistic estimate for power-law regression with a location parameter, with applications for making discharge rating curves. *Stochastic Environmental Research and Risk Assessment*, *20*(6), 445–453. Retrieved from `https://doi.org/10.1007%2Fs00477-006-0037-6` doi: 10.1007/s00477-006-0037-6

Reitan, T., & Petersen-Øverleir, A. (2008, Jun). Bayesian methods for estimating multi-segment discharge rating curves. *Stochastic Environmental Research and Risk Assessment*, *23*(5), 627–642. Retrieved from `http://dx.doi.org/10.1007/s00477-008-0248-0` doi: 10.1007/s00477-008-0248-0

Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. (2016, apr). Probabilistic programming in python using PyMC3. *PeerJ Computer Science*, *2*, e55. Retrieved from `https://doi.org/10.7717%2Fpeerj-cs.55` doi: 10.7717/peerj-cs.55

Venetis, C. (1970, jun). A note on the estimation of the parameters in logarithmic stage-discharge relationships with estimates of their error. *International Association of Scientific Hydrology. Bulletin*, *15*(2), 105–111. Retrieved from `https://doi.org/10.1080%2F02626667009493957` doi: 10.1080/02626667009493957