The Journal of Open Source Software

AHGestimation: An R package for computing robust, mass preserving hydraulic geometries and rating curves

J Michael Johnson[1], **Shahab Afshari**[2], and **Arash Modaresi Rad**[1]

1 Lynker, NOAA/NWS Office of Water Prediction Affiliate **2** University of Massachusetts Amherst

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: Chris Vernon

Reviewers:

- @mabesa
- @mengqi-z

Submitted: 30 October 2023
Published: unpublished

Summary

Background

In the field of hydrology it is common to express the behavior of a river channel at a given cross section using power law equations relating top width (TW), mean depth (Y), and velocity (V) to a given discharge (Q). Collectively these equations define the "at a station hydraulic geometry" (AHG) (Leopold & Maddock, 1953).

$$TW = a \cdot Q^b$$

$$Y = c \cdot Q^f$$

$$V = k \cdot Q^m$$

AHG relations apply to within-bank flows at a specific cross section and assume the channel characteristics do not significantly change with discharge. Under these assumptions, two continuity conditions determine mass conservation. When either of these is violated, the AHG's equations will cause mass imbalances in computations.

$$Q = TW \cdot Y \cdot V$$

and therefore:

$$\begin{aligned} b + f + m &= & (1)\\ a \times c \times k &= & (2)\\ &= 1 & (3) \end{aligned}$$

Statement of need

Large scale models simulating river flow are critical for forecasting water availability, drought, and flood inundation. These models must represent the size and shape of river channels in some generalized way.

While hydraulic geometry relationships have been extensively studied, they remain unquantified for the majority of stream reaches across the country. Consequently, large-scale models frequently use incomplete approximations that impact the accuracy of streamflow estimates (Heldmyer et al., 2022; J. M. Johnson et al., 2023) and flood forecasting (J. M. Johnson et

The Journal of Open Source Software

al., 2019; Maidment, 2017; Water Prediction, 2023; Zheng et al., 2018). At NOAA, National Water Model characteristics are based on trapezoidal geometries (e.g. (Gochis, 2020)) that are in part derived from hydraulic geometry relationships and drainage area assumptions found in (Bieger et al., 2015, 2016; Blackburn-Lynch et al., 2017).

Other efforts have aimed to calculate, and synthesize river channel data at a large scale in the United States (e.g. (Afshari, 2019; Enzminger et al., 2023)) however each of these relied on traditional Ordinary Least Squares Regression (OLS) fitting methods *and* data preprocessing (Afshari et al., 2017). And while both efforts produced valuable data products, the software is not shared.

This open source package is designed to assist work flows that are challenged by the following characteristics of hydraulic data:

1. Data is often distributed without consistent structure
2. Data is noisy and hard to fit with traditional methods
3. Data is tabular, making the development of efficient regional and continental datasets a challenge.

This software package enhances AHG fitting flexibility, ensuring mass conservation is met in hydraulic systems, and optimal curve fitting. It expands beyond the constraints of singular rating curves by prioritizing the analysis of the entire system while retaining the capability to compute fits based on individual relationships.

Software

AHG estimates have predominately occurred on a location-by-location basis with site specific knowledge guiding the data used and the validation of the outputs. As interest in large scale model applications increases, the need to estimate these relations from disparate and often noisy datasets is increasingly important.

`AHGestimation` is an R package (R Core Team, 2023) providing three capabilities:

1. Tools to estimate single and full system AHG relations using a robust estimation techniques that enforces flow continuity and minimizes total system error (`ahg_estimate`). This is accomplished by introducing a hybrid approach that supplements the traditional OLS approach, with a Nonlinear Least Square (NLS) regression, Evolutionary algorithm (NSGA-2; Mersmann (2020)), and ensemble modeling approach.

2. Methods to filter outliers based on time (`date_filter`), mass conservation (`qva_filter`), and statistical detection (`mad_filter`, `nls_filter`).

3. Formalize, in code, many of the concepts derived in (Dingman & Afshari, 2018) that relate AHG coefficients and exponents to cross-section hydraulics and geometry. These include (`cross_section`, `compute_hydraulic_params`, `compute_n`).

The package documentation includes several examples on the theory, design, and application of this tool set.

The first stable version of `AHGestimation` was made available in 2019 and was applied to an aggregated dataset of USGS manual field measurements. Since then, it has been actively developed to better understand and quantify these fundamental relationships in the face of noisy, large, and disparate data sources. Applications of the software have been used to (1) demonstrate how improved flood forecasts could be delivered from the NOAA/NWS National Water Model (J. M. Johnson et al., 2022) (2) help the NOAA/NWS Office of Water Prediction develop continental scale channel size and shape estimates to improve flood prediction and hydraulic routing and to (3) bolster the co-agency sponsored National Hydrologic Geospatial Fabric (Blodgett et al., 2021, 2023; Bock, 2022).

Example of use

AHGestimation is available on Gtihub and can be installed as follows:

```
#install.packages(remotes)
remotes::install.packages("mikejohnson51/AHGestimation")
```

This example shows how the package can be used to (1) remove data outliers based on time criteria and an NLS envelope (2) fit AHG parameters using a hybrid modeling approach and (3) estimate and plot the shape of the associated cross-section with an area-depth relation. The script to generate the plot can be found here, and the nwis data object is exported with the package to provide sample data and contains the field measurements taken at USGS site 01096500 on the Nashua River at East Pepperell in Massachusetts.

```
nwis

#> siteID       date            Q          Y         V       TW
#> 01096500 1984-11-14   9.7409954 0.5276645 0.652272 28.34640
#> 01096500 1985-01-04  11.8930757 0.6263473 0.682752 27.73680
#> 01096500 1985-05-06  10.8170356 0.5952226 0.563880 32.30880
#> 01096500 1985-06-26   1.9453674 0.2400300 0.332232 24.38400
#> 01096500 1986-01-09  11.3550556 0.5920154 0.606552 31.69920
#> 01096500 1986-02-27  21.4358532 0.8574593 0.902208 27.73680
#> 01096500 1986-05-22   6.9093107 0.4806462 0.454152 31.69920
#> 01096500 1986-07-07   8.9198068 0.5457371 0.512064 32.00400

nrow(nwis)
#>  245

# Keep only those observation made in the most recent 10 years,
# and that fall withing the .5 nls envelope
(data = nwis  |>
  dplyr::rename(Q = Q_cms, Y  = Y_m, V= V_ms, TW = TW_m) |>
  date_filter(10, keep_max = TRUE) |>
  nls_filter(allowance = 0.5) )

# data reduced to 80 observations based on filters
nrow(data)
#>  85
```

The reduced clean data can then be used to fit an AHG relation and compute a set of hydraulic parameters:

```
# Fit AHG relations
ahg_fit = ahg_estimate(data)
t(ahg_fit[1,])

#> V_method  "nls"
#> TW_method "nls"
#> Y_method  "nls"
#> c1        "1.006"
#> c2        "1.001"
#> viable    "TRUE"
#> tot_nrmse "0.3234122"
#> V_nrmse   "0.1337535"
#> TW_nrmse  "0.1009869"
#> Y_nrmse   "0.0886718"
```

```
#> V_coef    "0.2905399"
#> TW_coef   "18.23401"
#> Y_coef    "0.1898499"
#> V_exp     "0.3101059"
#> TW_exp    "0.1756746"
#> Y_exp     "0.5155772"
#> condition "bestValid"

# Use the AHG relations to compute hydraulic parameters
shape = compute_hydraulic_params(ahg_fit[1,])

#> r    p    d    R    bd   fd   md
#> 2.93 0.60 5.70 1.34 0.18 0.51 0.31
```

Finally, the max width and depth, paired with the derived r coefficient can be used to generate a cross section:

```
# Use the max width, max depth, and derived `r` to generate a cross section
# x: is the relative distance from the left bank
# Y: is the associated depth
# A: is the area associated with depth Y

cs = cross_section(r = shape$r,
                   TW = max(data$TW),
                   Ymax = max(data$Y))

#>  ind        x            Y            A
#>    1  0.000000 3.5525892702 1.337069e+02
#>    2  1.040816 3.1514749668 1.136999e+02
#>    3  2.081633 2.7814289966 9.601412e+01
#>    4  3.122449 2.4412395021 8.046464e+01
#>    5  4.163265 2.1296911854 6.687407e+01
#>    6  5.204082 1.8455651416 5.507242e+01
#>    7  6.244898 1.5876386747 4.489715e+01
#>    8  7.285714 1.3546850954 3.619324e+01
#>    9  8.326531 1.1454734964 2.881313e+01
#>   10  9.367347 0.9587685017 2.261685e+01
```

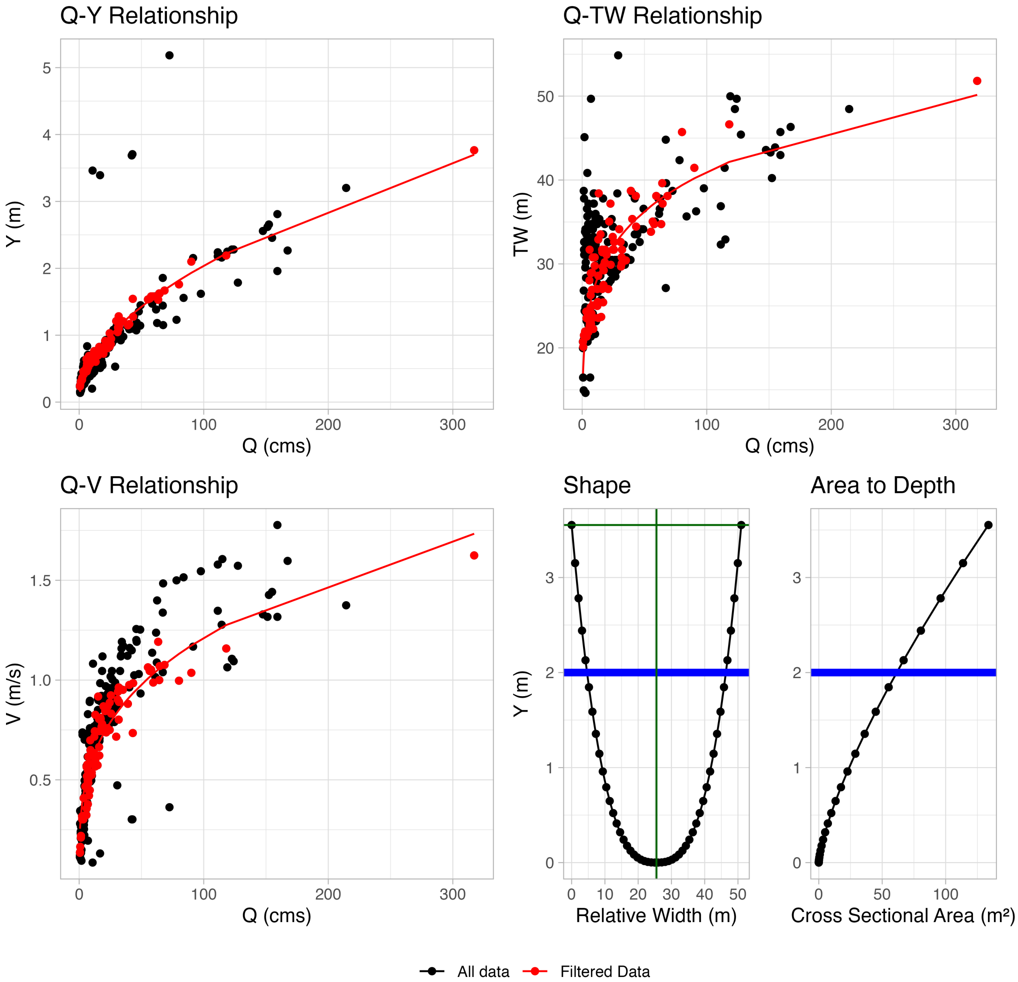

Figure 1: Faceted image with multiple views of the channel estimate.

As a proof of concept, this approach was applied to the synthetic rating curves generated by NOAA/NWS OWP's inundation mapping software ((Water Prediction, 2023)). The package allowed the size of their rating curve database to be reduced by 99.68% and maintained average accuracy within 0.4% nRMSE or the source data (J. ;. C. Johnson J. M.; Coll, 2022). Further, the reduction to a consistent AHG formulation allows them to be more interoperable with efforts like (Enzminger et al., 2023), or (Afshari, 2019) while also beginning to provide key training data for more advanced prediction methods that seek to estimate the shape on non-measured river segments.

Acknowledgements

The development of this package began in 2017 following the NOAA/NWS OWP Summer Institute and clear evidence channel shape may be a limiting factor in National Water Model performance.

The algorithm and implementation began as a graduate school project between friends at UC Santa Barbara and UMass Amherst and has since evolved to provide an open source utility for robust large scale data synthesis and evaluation. Funding from the National Science Foundation (Grants 1937099, 2033607) provided time to draft (J. ;. C. Johnson J. M.; Coll, 2022) and

The Journal of Open Source Software

apply an early version of this software to the Continental Flood Inundation Mapping synthetic rating curve dataset (Liu & Maidment, 2020). Funding from the NOAA/NWS OWP supported the addition of data filtering and hydraulic geometry estimation, improved documentation, and code hardening. We are grateful to all involved.

References

Afshari, S. (2019). *USGS table AHG parameters and supplementary data* (Version v1.0) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.2558565

Afshari, S., Fekete, B. M., Dingman, S. L., Devineni, N., Bjerklie, D. M., & Khanbilvardi, R. M. (2017). Statistical filtering of river survey and streamflow data for improving at-a-station hydraulic geometry relations. *Journal of Hydrology*, *547*, 443–454. https://doi.org/10.1016/j.jhydrol.2017.01.038

Bieger, K., Rathjens, H., Allen, P. M., & Arnold, J. G. (2015). Development and evaluation of bankfull hydraulic geometry relationships for the physiographic regions of the united states. *JAWRA Journal of the American Water Resources Association*, *51*(3), 842–858. https://doi.org/10.1111/jawr.12282

Bieger, K., Rathjens, H., Arnold, J. G., Chaubey, I., & Allen, P. M. (2016). Development and comparison of multiple regression models to predict bankfull channel dimensions for use in hydrologic models. *JAWRA Journal of the American Water Resources Association*, *52*(6), 1385–1400. https://doi.org/10.1111/1752-1688.12460

Blackburn-Lynch, W., Agouridis, C. T., & Barton, C. D. (2017). Development of regional curves for hydrologic landscape regions (HLR) in the contiguous united states. *JAWRA Journal of the American Water Resources Association*, *53*(4), 903–928.

Blodgett, D., Johnson, J. M., & Bock, A. (2023). Generating a reference flow network with improved connectivity to support durable data integration and reproducibility in the coterminous US. *Environmental Modelling & Software*, *165*, 105726.

Blodgett, D., Johnson, J. M., Sondheim, M., Wieczorek, M., & Frazier, N. (2021). Mainstems: A logical data model implementing mainstem and drainage basin feature types based on WaterML2 part 3: HY features concepts. *Environmental Modelling & Software*, *135*, 104927. https://doi.org/10.1016/j.envsoft.2020.104927

Bock, B., A. R. (2022). *National hydrologic geospatial fabric reference and derived hydrofabrics*. U.S. Geological Survey software release. https://www.sciencebase.gov/catalog/item/60be0e53d34e86b93891012b

Dingman, S. L., & Afshari, S. (2018). Field verification of analytical at-a-station hydraulic-geometry relations. *Journal of Hydrology*, *564*, 859–872. https://doi.org/10.1002/aaai.12035

Enzminger, T. L., Minear, J. T., & Livneh, B. (2023). *HyG: A hydraulic geometry dataset derived from historical stream gage measurements across the conterminous United States* [Data set]. Zenodo. https://doi.org/10.5281/zenodo.7868764

Gochis, M. B., D. J. (2020). The WRF-hydro® modeling system technical description, (version 5.2.0). In *NCAR Technical Note*. https://ral.ucar.edu/sites/default/files/public/projects/wrf-hydro/technical-description-user-guide/wrf-hydrov5.2technicaldescription.pdf

Heldmyer, A., Livneh, B., McCreight, J., Read, L., Kasprzyk, J., & Minear, T. (2022). Evaluation of a new observationally based channel parameterization for the national water model. *Hydrology and Earth System Sciences*, *26*(23), 6121–6136. https://doi.org/10.5194/hess-26-6121-2022

Johnson, J. ;. C., J. M.; Coll. (2022). Determining feature based hydraulic geometry and rating curves using a physically based, computationally efficient framework. In *Preprints*. https://doi.org/10.20944/preprints202212.0390.v1

Johnson, J. M., Fang, S., Sankarasubramanian, A., Rad, A. M., Cunha, L. K. da, Clarke, K. C., Mazrooei, A., & Yeghiazarian, L. (2023). *Comprehensive analysis of the NOAA *national water model: A call for heterogeneous formulations and diagnostic model selection.* https://doi.org/10.22541/essoar.167415214.45806648/v1

Johnson, J. M., Munasinghe, D., Eyelade, D., & Cohen, S. (2019). An integrated evaluation of the national water model (NWM)–height above nearest drainage (HAND) flood mapping methodology. *Natural Hazards and Earth System Sciences*, *19*(11), 2405–2420.

Johnson, J. M., Narock, T., Singh-Mohudpur, J., Fils, D., Clarke, K., Saksena, S., Shepherd, A., Arumugam, S., & Yeghiazarian, L. (2022). Knowledge graphs to support real-time flood impact evaluation. *AI Magazine*, *43*(1), 40–45. https://doi.org/10.1002/aaai.12035

Leopold, L. B., & Maddock, T. (1953). *The hydraulic geometry of stream channels and some physiographic implications* (Vol. 252). US Government Printing Office. https://doi.org/10.3133/pp252

Liu, T., Yan Y., & Maidment, D. R. (2020). Height above nearest drainage (HAND) and hydraulic property table for CONUS - version 0.2.1. (20200601). In *Oak Ridge Leadership Computing Facility*. https://cfim.ornl.gov/data/

Maidment, D. R. (2017). Conceptual framework for the national flood interoperability experiment. *JAWRA Journal of the American Water Resources Association*, *53*(2), 245–257. https://doi.org/10.1111/1752-1688.12474

Mersmann, O. (2020). *Mco: Multiple criteria optimization algorithms and related functions.* https://CRAN.R-project.org/package=mco

R Core Team. (2023). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.R-project.org/

Water Prediction, N. O. of. (2023). Inundation mapping. In *Github*. https://github.com/NOAA-OWP/inundation-mapping

Zheng, X., Tarboton, D. G., Maidment, D. R., Liu, Y. Y., & Passalacqua, P. (2018). River channel geometry and rating curve estimation using height above the nearest drainage. *JAWRA Journal of the American Water Resources Association*, *54*(4), 785–806. https://doi.org/10.1111/1752-1688.12661