Title

Data associated with “Different methods of estimating riverbed sediment grain size diverge at the basin scale ” (v2)

Summary

This data package is associated with the publication “Different methods of estimating riverbed sediment grain size diverge at the basin scale” published in Frontiers in Earth Science (Regier et al., 2025).

The distribution of sediment grain size in streams and rivers is often quantified by the median grain size (d50), a key metric for understanding and predicting hydrologic and biogeochemical function of streams and rivers. Manual methods to measure d50 are time-consuming and ignore larger grains, while model-based methods to estimate d50 often over-generalize basin characteristics, and therefore cannot accurately represent site-scale heterogeneity. Here, we apply a machine learning-enabled photogrammetry methodology (You Only Look Once, or YOLO) for estimating d50 for grains > 2 mm based on images collected from streams and rivers throughout the Yakima River Basin (YRB). To understand how such methods may help bridge the gaps in resolution and accuracy between manual and catchment characteristics model-based d50 estimates, we compared YOLO d50 values to manual and model-based estimates across the YRB. We found distinct differences among methods for d50 averages and variability, and relationships between d50 estimates and basin characteristics. Source images can be found at https://data.ess-dive.lbl.gov/view/doi:10.15485/1892052.

This data package was originally published in May 2023. It was updated August 2025 (v2; new and modified files). File and folder names were not revised to indicate changes. See the change history section below for more details.

Brief Overview of Methods

Images of riverbed grains were collected during a 2021 field campaign (171 images across 40 sites in the Yakima River basin, or YRB). A model using the You Only Look Once (YOLO) algorithm was trained to identify and measure individual grains, which was used to calculate the median grain size (d50) for each image. These values were compared to other d50 data sources in the YRB, including USGS manual measurements, and two continental-scale modeling approaches. We used simple spatial statistics and watershed characteristics to understand the similarities and differences between these d50 data sources and explore if the YOLO approach can help integrate between sparse, accurate manual measurements, and spatially resolved, generalized model estimates.

Critical Details

1. Geospatial data downloaded from the National Hydrograph Database (NHD) was used in the associated manuscript. This data was used to visualize the study watershed boundaries and flowlines but was not used for analyses. The data for the HUC8 watershed number 10730001 was downloaded from https://apps.nationalmap.gov/downloader/.
2. The terms catchment and basin are used through this data package. Catchment is defined as the smallest NHDPLUS catchment drainage area for each NHD stream reach. Basin is defined as the total upstream drainage area for each NHD stream reach.

3. Scripts used to construct figures associated the publication can be found at https://github.com/peterregier/d50_computer_vision. Methodology for the YOLO model is described in Chen et al. (2024).

Data Package Structure

In addition to this readme, this data package also includes a file-level metadata (FLMD) file that describes each file and a data dictionary (DD) that describes all column/row headers and variable definitions.

This dataset is comprised of one main data folder containing (1) file-level metadata; (2) data dictionary; (3) readme; (4) and subfolders containing data, figures, and scripts. The data folder contains datasets used for the analyses in the manuscript in image, text-delimited or geospatially-referenced formats. The figures folder contains the figures from the manuscript in different formats. The scripts folder contains all of the scripts used to complete the analyses in the manuscript. All files are .csv, .rds, .dbf, .prj, .shp, .shx, .jpg, .png, .R, .Rproj, or .pdf.

Citations, Acknowledgements, and License

Acknowledgements

This research was supported by the U.S. Department of Energy (DOE) Biological and Environmental Research (BER) Environmental System Science (ESS) program (https://ess.science.energy.gov/) through the Pacific Northwest National Laboratory River Corridor Science Focus Area (SFA). PNNL is operated by Battelle Memorial Institute for the U.S. Department of Energy under Contract No. DE-AC05-76RL01830.

Citations

- Chen, Y., Bao, J., Chen, R., Li, B., Yang,Y., Renteria, L., et al. (2024). Quantifyingstreambed grain size, uncertainty, andhydrobiogeochemical parameters usingmachine learning model YOLO. WaterResources Research, 60,e2023WR036456. https://doi.org/10.1029/2023WR036456
- Fulton S G ; Barnes M ; Borton M A ; Chen X ; Farris Y ; Forbes B ; Garayburu-Caruso V A ; Goldman A E ; Grieger S ; Kaufman M H ; Lin X ; McKever S A ; Myers-Pigg A ; Otenburg O ; Pelly A ; Ren H ; Renteria L ; Scheibe T D ; Son K ; Torgeson J M ; Stegen J C (2022): Spatial Study 2021: Sensor-Based Time Series of Surface Water Temperature, Specific Conductance, Total Dissolved Solids, Turbidity, pH, and Dissolved Oxygen from across Multiple Watersheds in the Yakima River Basin, Washington, USA. River Corridor and Watershed Biogeochemistry SFA, ESS-DIVE repository. Dataset. doi:10.15485/1892052
- Regier P, Chen Y, Son K, Bao J, Forbes B, Goldman A, Kaufman M, Rod KA and Stegen J (2025) Different methods of estimating riverbed sediment grain size diverge at the basin scale. Front. Earth Sci. 13:1529503. doi: 10.3389/feart.2025.1529503

License

This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

Contact

Peter Regier; peter.regier@pnnl.gov

Change History

Change history:

Data Package Version	Changes
Version 1 *May 2023*	Original data package publication
Version 2 *August 2025*	Many changes were made during reprocessing and re-analyzing data to address reviewers' comments prior to publishing the manuscript. See the GitHub commit history for details.