**ESS-DIVE Metadata for grain size methods comparison Data Package**

**Title**:

Data associated with “Machine learning photogrammetric analysis of images provides a scalable approach to study riverbed grain size distributions”

**Alternative Identifiers:**

**Abstract:**

This data package is associated with the publication “Machine learning photogrammetric analysis of images provides a scalable approach to study riverbed grain size distribution” to be submitted to Water Resources Research (Regier et al.).

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 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 photogrammetric methods may help bridge the gaps in resolution and accuracy between manual and 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. Related R scripts can be found at <https://github.com/peterregier/d50_computer_vision>.

This dataset supports a broader study examining the drivers of spatial variability in sediment respiration rates in the Yakima River Basin. We acknowledge the Yakama Nation as owners and caretakers of the lands where we collected the data used in this project. We thank the Confederated Tribes and Bands of the Yakama Nation Tribal Council and Yakama Nation Fisheries for working with us to facilitate sample collection and optimization of data usage according to their values and worldview.

This dataset is comprised of one folder with YOLO model results and one main data folder containing (1) file-level metadata; (2) data dictionary; (3) d50 estimates for USGS sites within the watershed, (4) d50 estimates for images collected across the study basin, (5) digitized distribution information for d50 estimates from Abeshu et al. 2022, and (6) study site characteristics. All files are .csv, .dat, or .pdf.

**Keywords**:

Grain size distribution

Median grain size

Machine learning

Computer vision

Photogrammetry

Freshwater

River

Stream

Riverbed

Hydrology

Biogeochemistry

Respiration

**Data variables:**

Median grain size (d50)

Stream order

Urban land cover

Elevation

Basin area

Stream length

Stream slope

Precipitation as snow

Evapotranspiration

Image suitability

**Pub date**:

**Data usage rights**:

Creative Commons Attribution

**Project**:

River Corridor and Watershed Biogeochemistry SFA

**Funding org:**

U.S. DOE > Office of Science > Biological and Environmental Research (BER)

**DOE Contracts**:

DOE Award #54737

**Related reference**:

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](https://data.ess-dive.lbl.gov/datasets/doi:10.15485/1892052)

**Principal investigator:**

Scheibe

**Contact name**:

Peter Regier

**Contact email**:

peter.regier@pnnl.gov

**Creators**:

Regier

Yunxiang Chen

Son

Bao

Forbes

Goldman

Kaufman

Rod

Stegen

**Start date**:

2021-08-30

**End date**:

2021-09-14

**Location description**:

Refer to metadata spreadsheet

**Coordinates**:

Refer to metadata spreadsheet

**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.