Title

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

Summary

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.

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 manuscript in preparation (led by Y. Chen).

Data Package Structure

This dataset is comprised of one main data folder containing (1) file-level metadata; (2) data dictionary; (3) readme; (4) d50 estimates for USGS sites within the watershed; (5) d50 estimates for images collected across the study basin; (6) digitized distribution information for d50 estimates from Abeshu et al. 2022; (7) study site characteristics; and (8) a subfolder with YOLO model results. All files are .csv, .dat, or .pdf.

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.

Contact

Peter Regier; peter.regier@pnnl.gov

Change History

Version 1	May 2023	Original data package publication