

Title

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 distributions” submitted to XXX (Regier et al) and a GitHub repository (https://github.com/peterregier/d50_computer_vision). 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. We discuss the advantages and limitations of the YOLO algorithm versus current methods, and explore potential future directions to combine d50 methods to better estimate spatiotemporal variation of d50, and improve incorporation into basin-scale models. This data package is associated with

Critical Details

Public datasets used

We used publicly available shapefiles downloaded from the National Hydrography Database (NHD) for plotting the study watershed boundaries and flowlines. We downloaded shapefiles for the HUC8 watershed numbered 17030001 from <https://apps.nationalmap.gov/downloader>

Data Package Structure

Please see d50_flmd.csv for a list of all files contained in this data package and descriptions for each. The data dictionary for csv files has “_dd” appended to the file name.

Citations and 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.]. Cite this data package with the appropriate DOI. Cite the associated manuscript in any work that that uses analyses or conclusions presented in the manuscript. To cite the paper: XXX.

We acknowledge the U.S. Department of Energy (DOE) Biological and Environmental Research (BER) Environmental System Science (ESS) program (<https://ess.science.energy.gov/>) for funding this research.

Contact

Peter Regier, peter.regier@pnnl.gov

Change History

Version 1	[Insert Date]	Original data package publication