January XX, 2023

Dear XXX editorial team,

Please find attached our original manuscript entitled "Machine learning photogrammetric analysis of images provides a scalable approach to study riverbed grain size distributions" for your consideration for publication in *XXX*. Accurate, spatiotemporally resolved information on riverbed grain size distributions (GSD), often quantified as the median GSD (d50) across fluvial networks is important for modeling watershed functions, including the fate and transport of solutes and particulates, and ecosystem respiration. However, while there are multiple methods for measuring or estimating d50, they each have significant limitations.

Here, we present a novel photogrammetric method to estimate d50 values across the Yakima River basin (YRB), which we compare to publicly available d50 measurements and estimates across the basin. We made machine-learning enabled photogrammetric d50 estimates for images collected across 40 sites using using the You Only Live Once (YOLO) algorithm, which utilizes machine learning to identify and quantify grains in images we collected at 40 sites across the YRB. We then compared YOLO-based estimates to manual measurements and estimates made using two different model-based approaches. Our results indicate YOLO provides rapid and accurate characterization of d50 across a wide variety of sites and has potential to bridge the gap between manual measurements, which are ground-truthed and site-specific, but generally spatially limited, and modeled estimates, which are spatially continuous, but may over-generalize. In addition, we explore several potential avenues for improving the spatiotemporal resolution of d50 from reach-to-basin scales, including opporutnities to construct living models, and inclusion of uncrewed platforms to automate image collection.

Because d50 values are so important for accurately modeling and predicting watershed hydrologic and biogeochemical function, accurate, spatially-resolved, and scalable measurements are an urgent need to improve our understanding of watershed function. We believe that the work herein represents a needed step towards harnessing the power of machine learning and growing “big data” resources to help integrate manual and modeled d50 information into a cohesive, scalable approach to rapidly and accurately characterize ensuring responsible and appropriate incorporation of machine learning into analysis of the growing “big data” resources being collected and used by aquatic scientists.

P.R., Y.C, and J.S. designed the study approach, P.R., Y.C., and K.S., conducted data collection and processing, Y.C. and J.B. led modeling, and P.R. led statistical analyses of model outputs. P.R. led drafting of the manuscript, and all authors contributed to editing. All authors have read the manuscript, agree with the journal’s data deposition requirements (all data and code are publicly available), and declare no conflicts of interest.

Sincerely,

Peter Regier and co-authors