# Quantifying streambed grain sizes and hydro-biogeochemistry using YOLO and photos

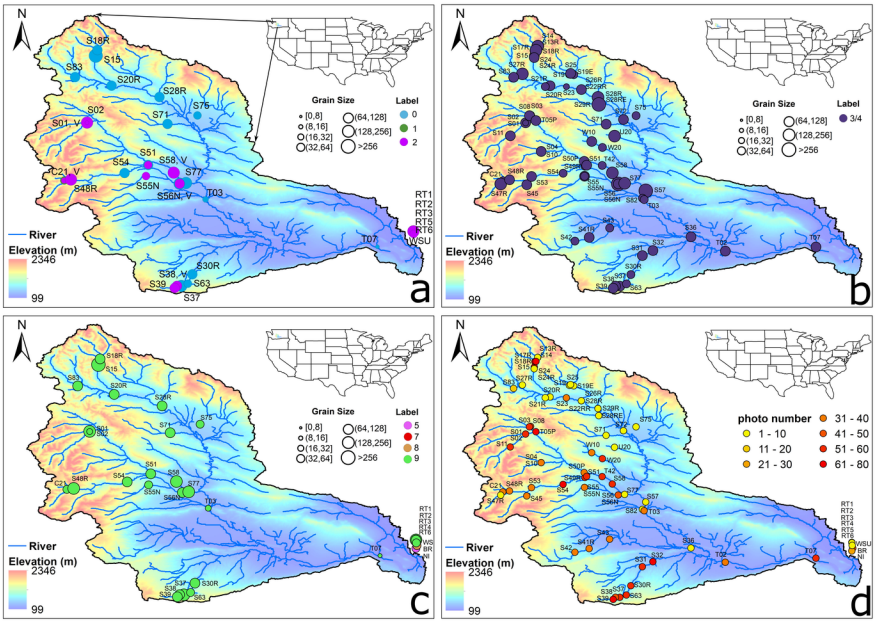
Yunxiang Chen[1], Jie Bao[2], Rongyao Chen[3], Bing Li[4], Yuan Yang[5], Lupita Renteria[2], Dillman Delgado[1], Brieanne Forbes[1], Amy E. Goldman[2], Manasi Simhan[6], Morgan Barnes[1], Maggi Laan[1], Sophia McKever[1], Zhangshuan Hou[2], Xingyuan Chen[2], Timothy D. Scheibe[2], and James C Stegen[2]

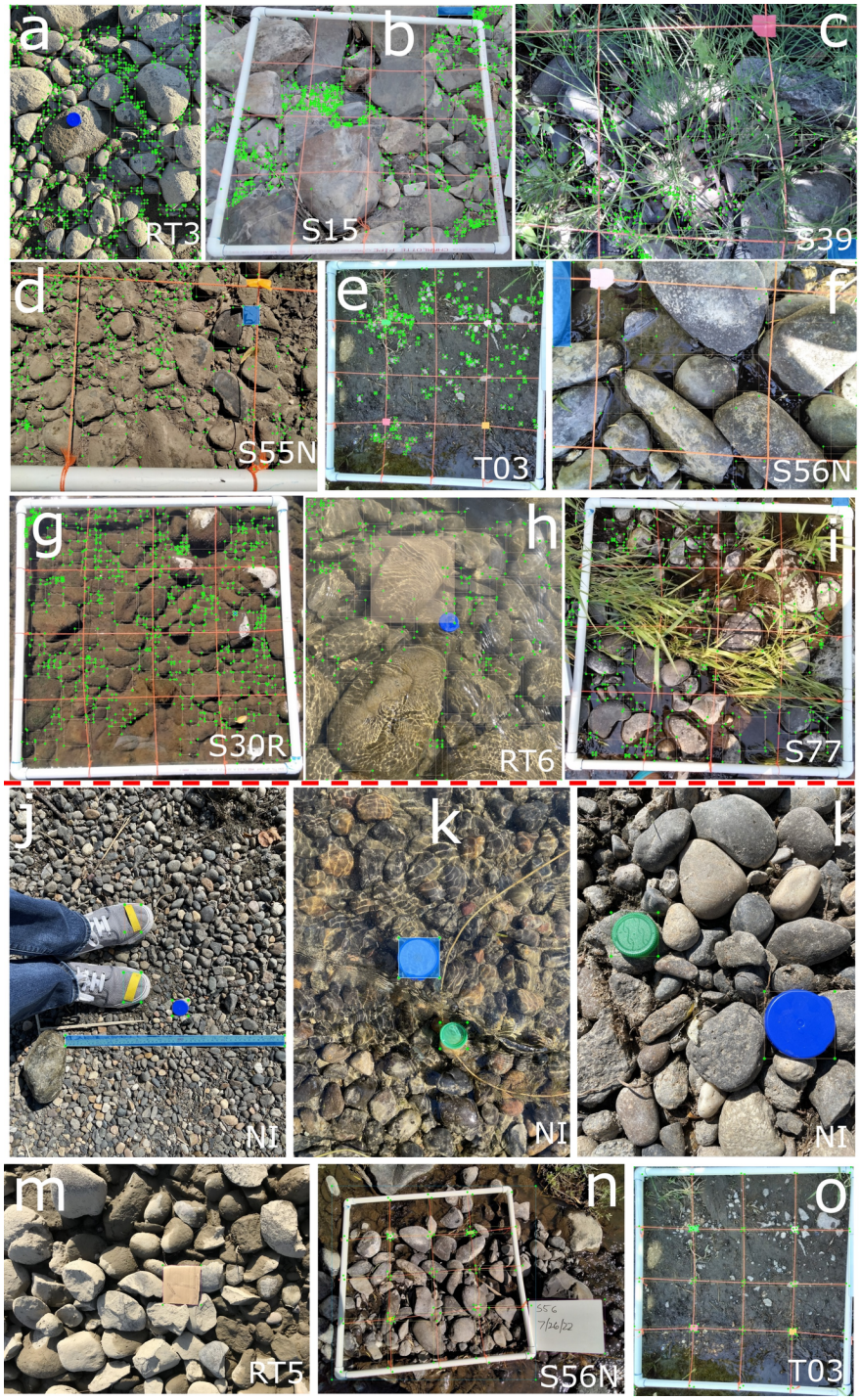
[1]Pacific Northwest National Laboratory
[2]Pacific Northwest National Laboratory (DOE)
[3]School of Petroleum Engineering, Yangtze University
[4]PNNL
[5]Center for Western Weather and Water Extremes, Scripps Institution of Oceanography, University of California San Diego
[6]Department of Geoscience, University of Wisconsin - Madison

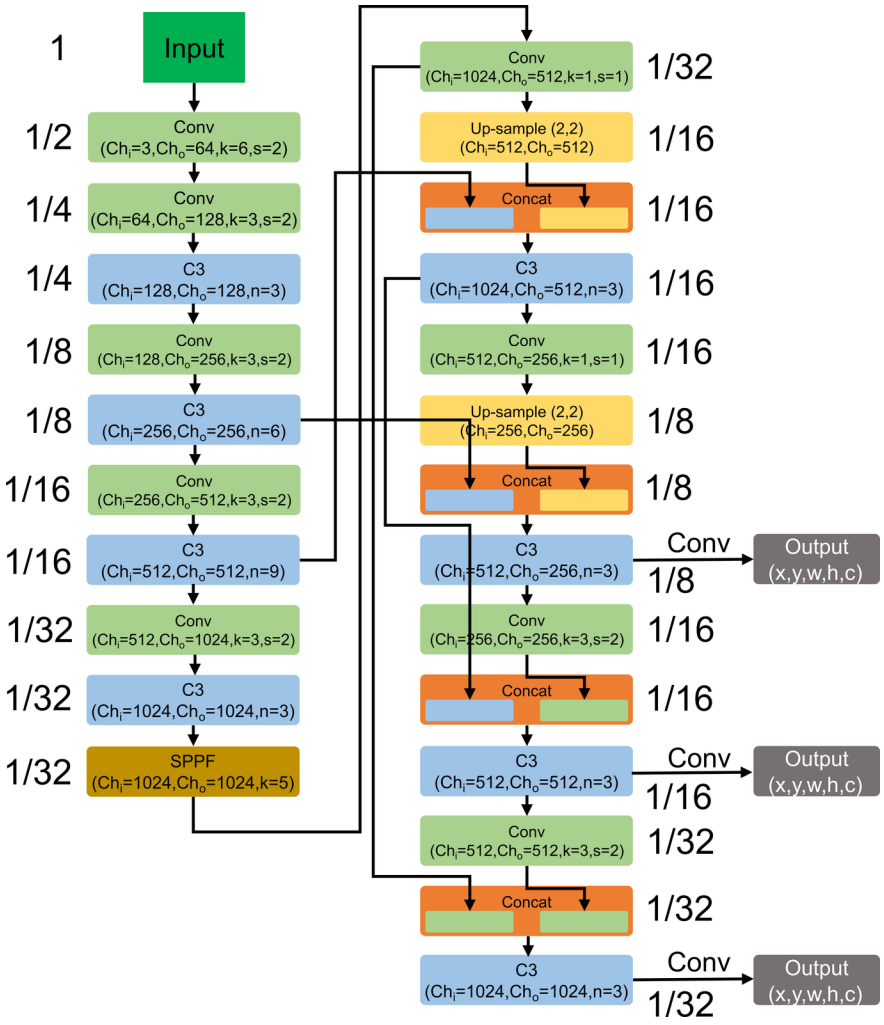
October 19, 2023

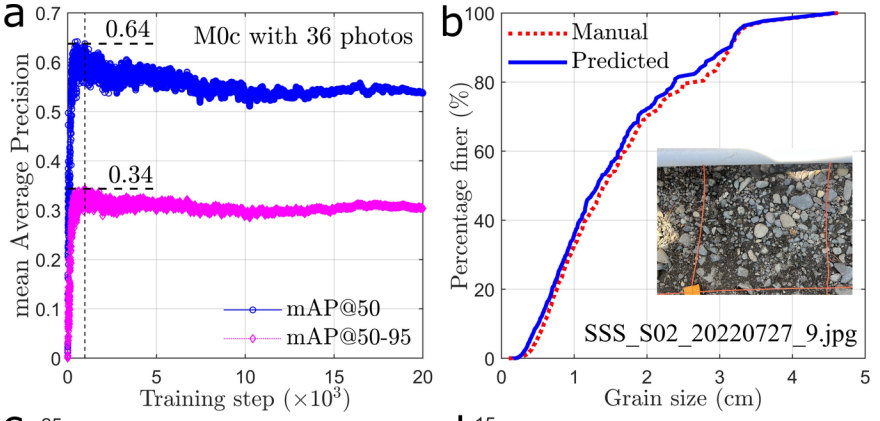
**Abstract**

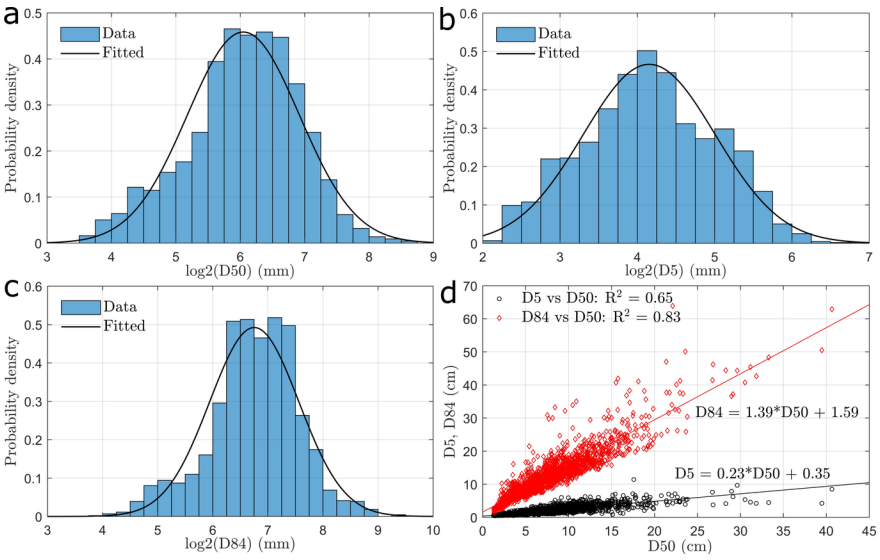
Streambed grain sizes and hydro-biogeochemistry (HBGC) control river functions. However, measuring their quantities, distributions, and uncertainties is challenging due to the diversity and heterogeneity of natural streams. This work presents a photo-driven, artificial intelligence (AI)-enabled, and theory-based workflow for extracting the quantities, distributions, and uncertainties of streambed grain sizes and HBGC parameters from photos. Specifically, we first trained You Only Look Once (YOLO), an object detection AI, using 11,977 grain labels from 36 photos collected from 9 different stream environments. We demonstrated its accuracy with a coefficient of determination of 0.98, a Nash–Sutcliffe efficiency of 0.98, and a mean absolute relative error of 6.65% in predicting the median grain size of 20 testing photos. The AI is then used to extract the grain size distributions and determine their characteristic grain sizes, including the 5th, 50th, and 84th percentiles, for 1,999 photos taken at 66 sites. With these percentiles, the quantities, distributions, and uncertainties of HBGC parameters are further derived using existing empirical formulas and our new uncertainty equations. From the data, the median grain size and HBGC parameters, including Manning's coefficient, Darcy-Weisbach friction factor, interstitial velocity magnitude, and nitrate uptake velocity, are found to follow log-normal, normal, positively skewed, near log-normal, and negatively skewed distributions, respectively. Their most likely values are 6.63 cm, 0.0339 s·m-1/3, 0.18, 0.07 m/day, and 1.2 m/day, respectively. While their average uncertainty is 7.33%, 1.85%, 15.65%, 24.06%, and 13.88%, respectively. Major uncertainty sources in grain sizes and their subsequent impact on HBGC are further studied.

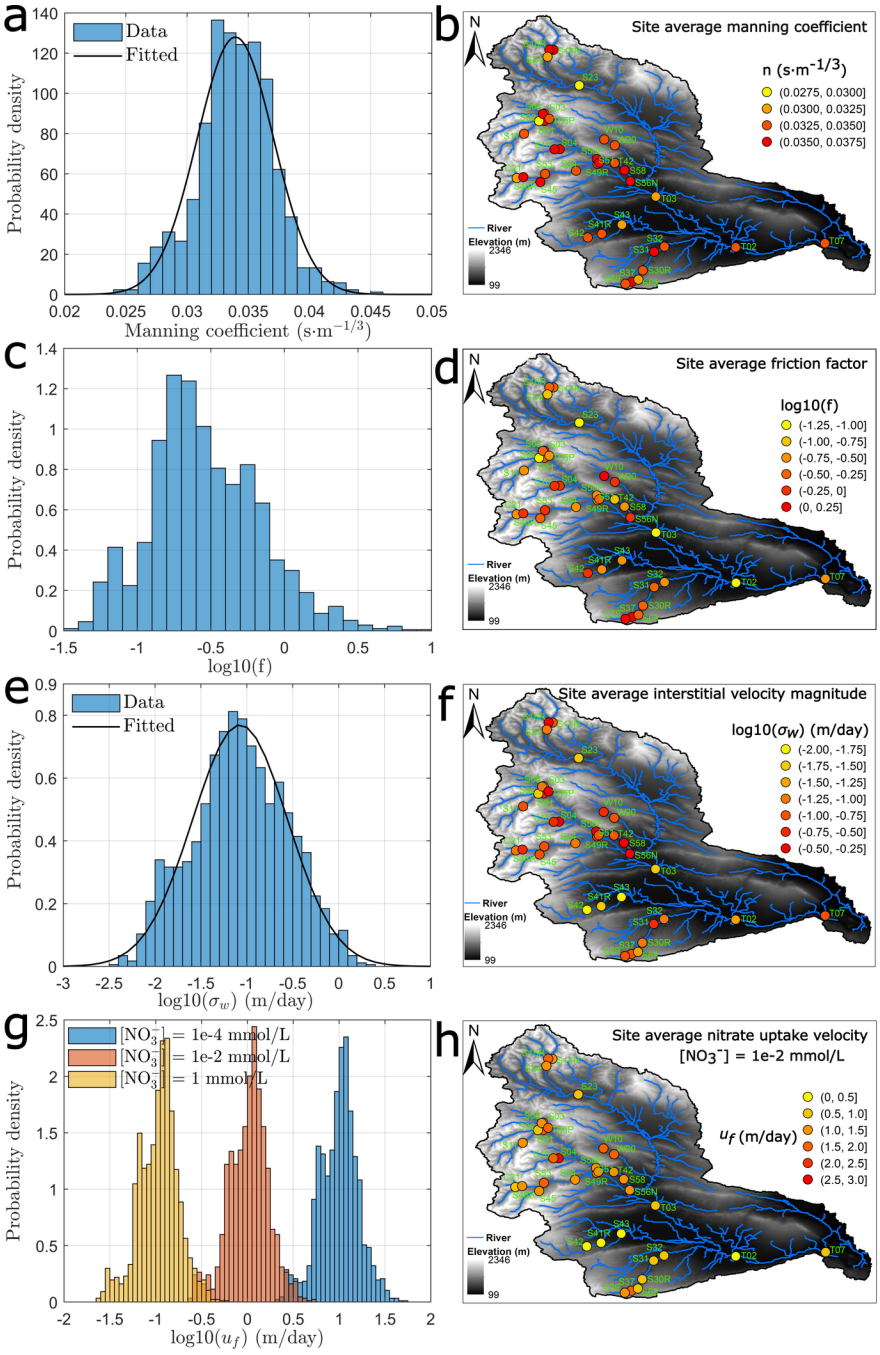


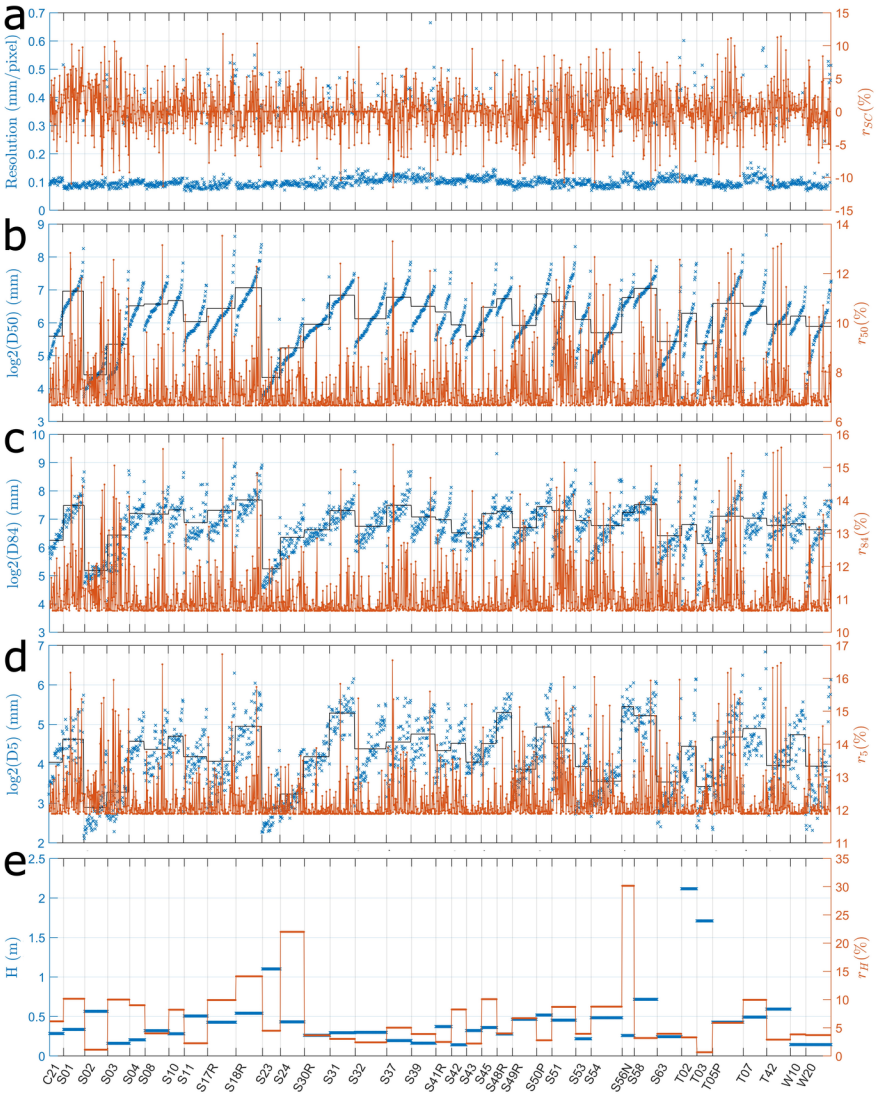


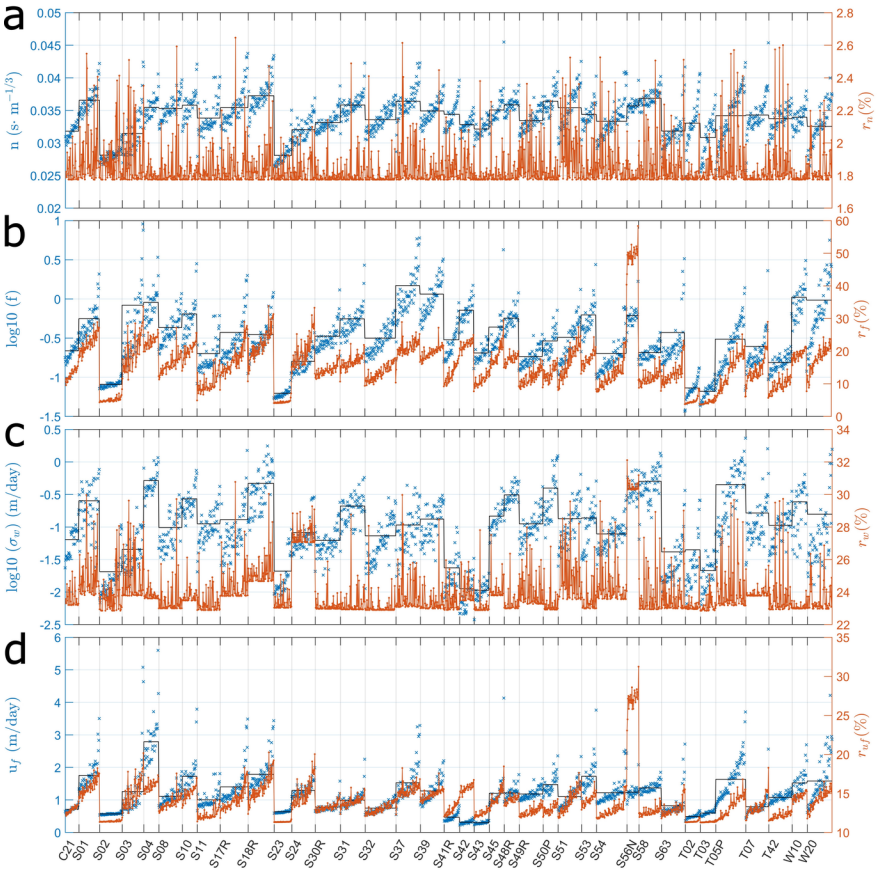


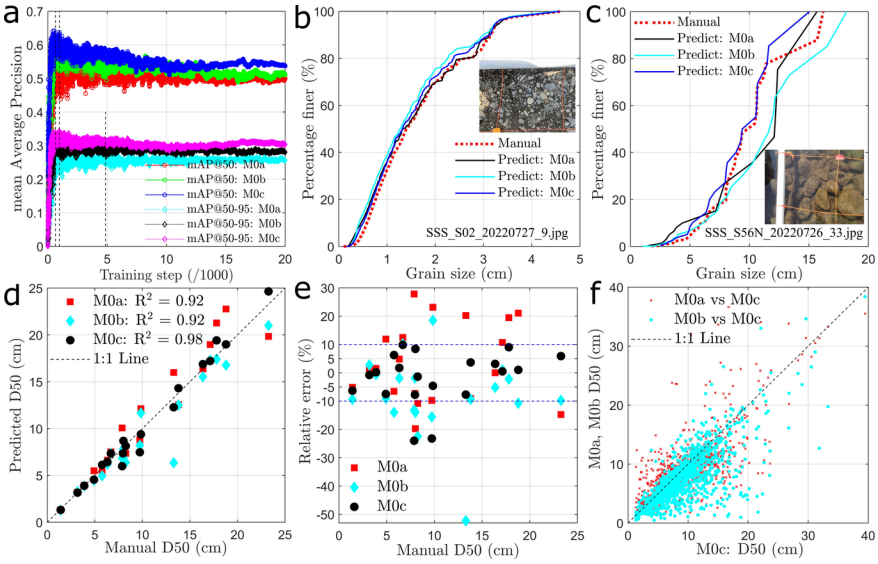


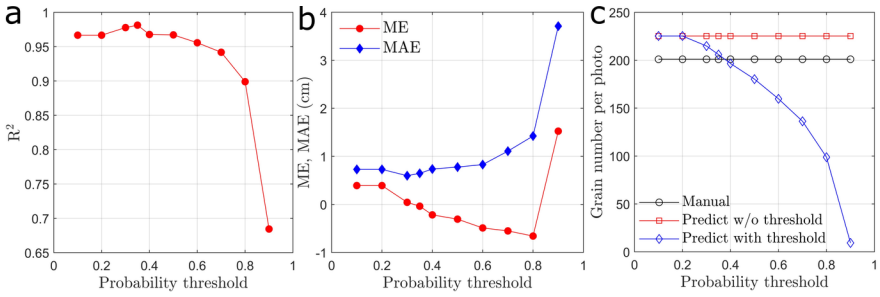



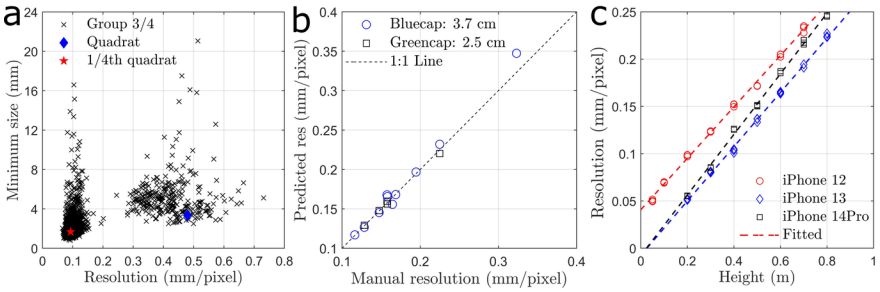


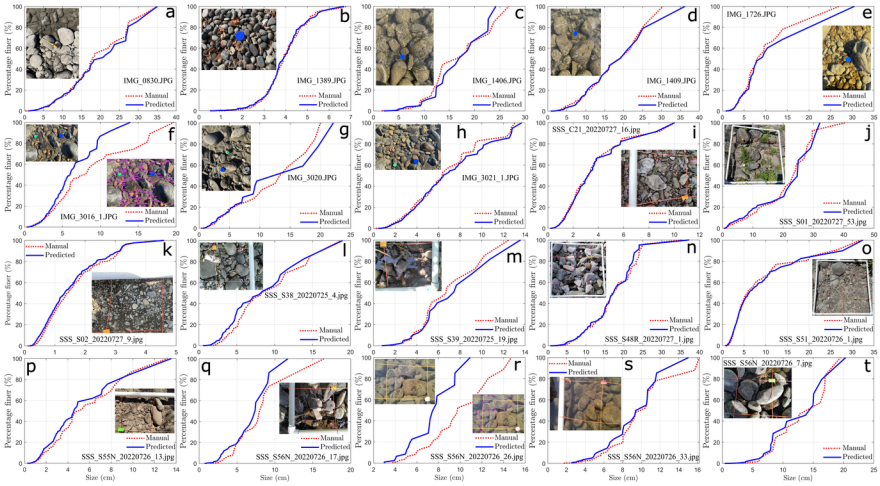


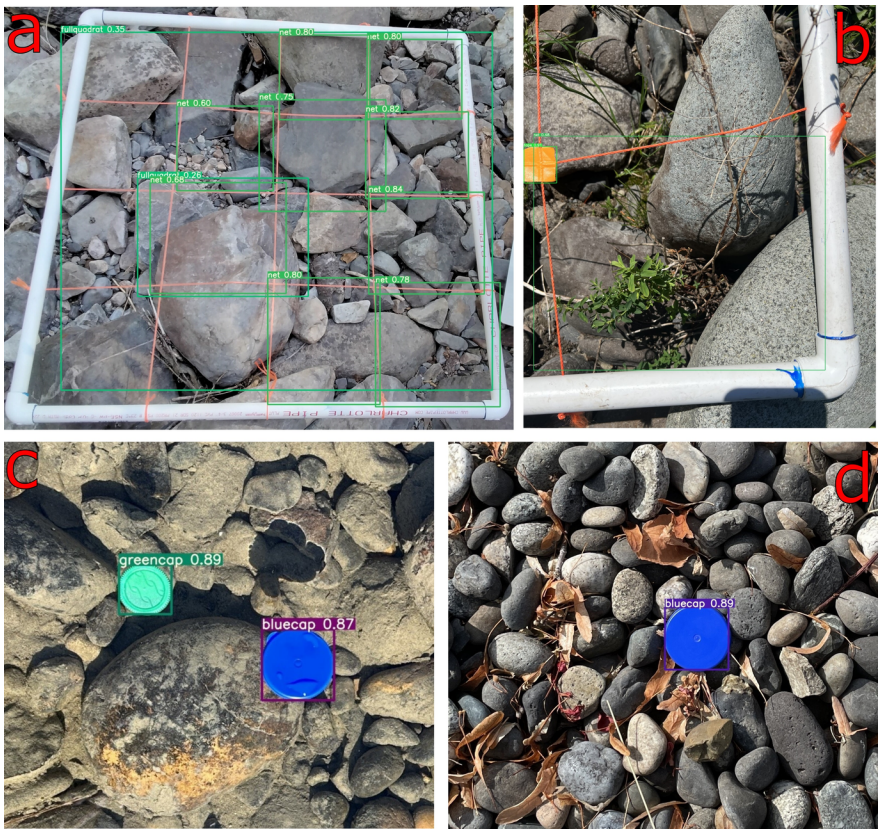






# Quantifying streambed grain sizes and hydro-biogeochemistry using YOLO and photos

**Yunxiang Chen[1], Jie Bao[1], Rongyao Chen[2], Bing Li[1], Yuan Yang[3], Lupita Renteria[1], Dillman Delgado[1], Brieanne Forbes[1], Amy E. Goldman[1], Manasi Simhan[4], Morgan E. Barnes[1], Maggi Lanan[1], Sophia McKever[1], Z. Jason Hou[1], Xingyuan Chen[1], Timothy Scheibe[1], James Stegen[1,5]**

[1]Pacific Northwest National Laboratory, Richland, WA, USA
[2]School of Petroleum Engineering, Yangtze University, Wuhan, Hubei, China
[3]Center for Western Weather and Water Extremes, Scripps Institution of Oceanography, University of California San Diego, CA, USA
[4]Department of Geoscience, University of Wisconsin - Madison, Madison, WI, USA
[5]School of the Environment, Washington State University, Pullman, WA, USA

**Key Points:**

- Stream sediments bigger than 0.45 mm can be detected from smartphone photos by YOLO with a Nash–Sutcliffe efficiency of 0.98.
- Quantities, distributions, and uncertainties of streambed hydro-biogeochemistry can be determined from photos.
- We have identified sources of uncertainty in grain size measurements and proposed approaches to reduce this uncertainty.

Corresponding author: Yunxiang Chen, yunxiang.chen@pnnl.gov

## Abstract

Streambed grain sizes and hydro-biogeochemistry (HBGC) control river functions. However, measuring their quantities, distributions, and uncertainties is challenging due to the diversity and heterogeneity of natural streams. This work presents a photo-driven, artificial intelligence (AI)-enabled, and theory-based workflow for extracting the quantities, distributions, and uncertainties of streambed grain sizes and HBGC parameters from photos. Specifically, we first trained You Only Look Once (YOLO), an object detection AI, using 11,977 grain labels from 36 photos collected from 9 different stream environments. We demonstrated its accuracy with a coefficient of determination of 0.98, a Nash–Sutcliffe efficiency of 0.98, and a mean absolute relative error of 6.65% in predicting the median grain size of 20 testing photos. The AI is then used to extract the grain size distributions and determine their characteristic grain sizes, including the 5th, 50th, and 84th percentiles, for 1,999 photos taken at 66 sites. With these percentiles, the quantities, distributions, and uncertainties of HBGC parameters are further derived using existing empirical formulas and our new uncertainty equations. From the data, the median grain size and HBGC parameters, including Manning's coefficient, Darcy-Weisbach friction factor, interstitial velocity magnitude, and nitrate uptake velocity, are found to follow log-normal, normal, positively skewed, near log-normal, and negatively skewed distributions, respectively. Their most likely values are 6.63 cm, 0.0339 $s \cdot m^{-1/3}$, 0.18, 0.07 m/day, and 1.2 m/day, respectively. While their average uncertainty is 7.33%, 1.85%, 15.65%, 24.06%, and 13.88%, respectively. Major uncertainty sources in grain sizes and their subsequent impact on HBGC are further studied.


## Plain Language Summary

Streambed grain sizes control river hydro-biogeochemical function by modulating the resistance, speed of water exchange, and nutrient transport at water-sediment interface. Consequently, quantifying grain sizes and size-dependent hydro-biogeochemical parameters is critical for predicting river's function. In natural streams, measuring these sizes and parameters, however, is challenging because grain sizes vary from millimeters to a few meters, change from small creeks to big streams, and could be concealed by complex non-grain materials such as water, ice, mud, and grasses. All these factors make size measurements a time-consuming and high-uncertain task. We address these challenges by demonstrating a workflow that combines a computer vision artificial intelligence (AI),

smartphone photos, and new uncertainty quantification theories. The AI performs well across various sizes, locations, and stream environments as indicated by an accuracy metric of 0.98. We apply the AI to extract the grain sizes and their characteristic percentiles for 1,999 photos. These characteristic grain sizes are then input into existing and our new theories to derive the quantities, distributions, and uncertainties of hydro-biogeochemical parameters. The high accuracy of the AI and the success of extracting grain sizes and hydro-biogeochemical parameters demonstrate the potential to advance river science with computer vision AI and mobile devices.

## 1 Introduction

Streambed grain size is a crucial factor controlling streambed hydro-biogeochemistry (HBGC). In hydrology, hydraulics, and geomorphology, streambed flow resistance, which is parameterized by the Manning coefficient or Darcy–Weisbach friction factor, is directly linked to characteristic grain sizes such as the median, 84th, and 90th percentiles of grain size distributions (Strickler, 1923; S. Lang et al., 2004; Chaudhry, 2008; Ferguson, 2010, 2007; Rickenmann & Recking, 2011; Powell, 2014; Ferguson, 2022). In stream-groundwater interactions, the speed of water exchange through the porous sediment interface, quantified as streambed interstitial velocity, is controlled by pressure variation and subsurface permeability, both of which depend on characteristic grain sizes of streambeds (Kenney et al., 1984; Shepherd, 1989; Elliott & Brooks, 1997; Y. Chen et al., 2021). In biogeochemistry, grain sizes exert direct control over turbulent mass transfer that determines the upper limit of the total nitrate uptake velocity from streams by benthic algae, microbes, and turbulence (O'Connor & Hondzo, 2008; Mulholland et al., 2009; Grant et al., 2018). Despite the importance, measuring streambed grain sizes and size-dependent HBGC is challenging due to the multiscale and heterogeneous nature of grain size, the diversity of stream environments, and consequently the high labor costs associated with grain size quantification and HBGC estimation.

Over the past seven decades, large efforts have been made to address the aforementioned challenges. These efforts can be categorized into traditional sieve methods, grid- or area-based sediment counting or weighting methods (Wolman, 1954; Leopold, 1970; Kellerhals & Bray, 1971; Anastasi, 1984; Fehr, 1987; Fripp & Diplas, 1993), manual photo sieving method (Adams, 1979; Ibbeken & Schleyer, 1986), automated or semi-automated photo sieving methods (Butler et al., 2001; Graham et al., 2005; Detert & Weitbrecht, 2012; Purinton & Bookhagen, 2019), image texture statistics methods (Carbonneau et al., 2004; Rubin, 2004; Verdú et al., 2005; Carbonneau et al., 2005a, 2005b; Buscombe & Masselink, 2009; Buscombe et al., 2010; Buscombe & Rubin, 2012; Buscombe, 2013; Black et al., 2014), machine learning (ML) methods (Z. Chen et al., 2020; Soloy et al., 2020; N. Lang et al., 2021; Ermilov et al., 2022), point cloud methods (Vázquez-Tarrío et al., 2017; Steer et al., 2022), and ML-based in-direct grain size regression methods (Gomez-Velez et al., 2015; Ren et al., 2020; Abeshu et al., 2022). The sieve method is the oldest and most reliable approach for fine sediment characterization, however, it is not feasible for field sampling of coarse sediments due to the requirement to transport a large

number of rocks to the laboratory for drying, sieving, and weighing (Leopold, 1970). Although the grid and area based methods avoid the need to move heavy rocks, they suffer from poor reproducibility along with significant time and labor costs, due to the necessity of manually measuring and recording grain sizes in the field (Wohl et al., 1996; Bunte & Abt, 2001).

The manual photo-sieve approach was therefore developed in the late 1970s to circumvent the need for direct measurements of grains in the field, however, it remains time-consuming as it involves manual identification and digitization of grains from images (Graham et al., 2005). Consequently, automated and semi-automated techniques were developed. These approaches are based on a series of image processing algorithms such as converting colored images to grayscale, applying simple or double thresholds, edge detection, bottom-hat transformation, and finally using watershed segmentation or k-means clustering to generate individual grains (Graham et al., 2005; Detert & Weitbrecht, 2012; Purinton & Bookhagen, 2019). These methods significantly reduce the time required to generate reliable grain size distributions, but usually need considerable time to adjust key parameters used in the image processing techniques (Graham et al., 2005; Purinton & Bookhagen, 2019). Instead of directly detecting individual grains, statistical methods approximate key grain size metrics, such as the median size, by relating grain sizes to characteristic quantities of image texture derived from auto-correlation (Rubin, 2004), one-dimensional (1D) and two-dimensional (2D) semi-variance (Carbonneau et al., 2004; Verdú et al., 2005), co-occurrence matrix-derived entropy (Carbonneau et al., 2005b), spectrum decomposition (Buscombe et al., 2010), wavelets (Buscombe & Rubin, 2012; Buscombe, 2013), and their combinations (Buscombe & Masselink, 2009; Black et al., 2014). Among these methods, the spectrum decomposition and the global wavelet approaches are especially important because they provide good estimates for the median size (with root-mean-square relative errors of 9.5% to 16%) and the full grain size distribution without the need for calibration (Buscombe et al., 2010; Buscombe, 2013). Despite these successes, it is worth noting that mean sizes obtained from statistical methods are conceptually similar but different from the sizes obtained from sieve or photo-sieve approaches.

In addition to image processing and statistical methods, machine learning methods implicitly learn the relationship between input images and desired targets using data and neural networks. Examples include learning median size and grain size distribution

(N. Lang et al., 2021), individual grains (Soloy et al., 2020; Z. Chen et al., 2020), and clustered grains (Ermilov et al., 2022) using convolutional neural networks (CNNs), Mask regional CNN (R-CNN) (He et al., 2017), and atrous separable convolution (L.-C. Chen et al., 2018), respectively. The Mask R-CNN is the most similar to the traditional sieve and photo-sieve methods, however, its accuracy, which stands at approximately a 50% detection rate in predicting overlapping rocks, needs further improvement before being deployed for practical applications (Soloy et al., 2020). All of the image-based methods mentioned above use images as input, therefore, the grain sizes are three dimensional (3D)-sediment projected 2D sizes. The point-cloud based grain size characterization is more similar to actual 3D grain sizes (Steer et al., 2022), but obtaining accurate 3D point cloud poses a larger challenge than grain size quantification. There also exist ML-based in-direct methods to estimate grain sizes by learning the relationship between median grain size and large-scale geomorphological and hydrological attributes such as elevation, slope, depth, velocity, etc. (Gomez-Velez et al., 2015; Ren et al., 2020; Abeshu et al., 2022). These estimates, however, are not actual measurements and require careful validation against direct measurements before their use in large-scale models.

In summary, past efforts have tackled challenges related to accuracy, reproducibility, cost, multi scales, and heterogeneity. These methods are expected to yield satisfactory results when applied to streambeds primarily composed of granular sediments, such as sand, cobble, gravel, and boulders (Buscombe, 2013). However, they may encounter challenges in stream riparian zones where non-granular materials like grass, mud, ice, wood, and both static and flowing water overlie granular sediments. New methods that can detect sediments hidden beneath these non-granular and non-sediment objects are needed. Another aspect that is not well resolved by previous efforts is photo resolution estimation. Though photo resolution can be manually measured from reference scales, this process is usually time-consuming when dealing with a large number of images. Therefore, there is a need for fully automated photo resolution estimation method.

Our first goal is to address these needs by developing two ML models, one for grain detection and one for scale detection, using the You Only Look Once (YOLO) version 5 framework (Redmon et al., 2016) with 11,977 and 121 labels of grains and reference scales. The YOLO framework is selected because it is a general, real-time, object detection algorithm (Redmon et al., 2016) with the capability to detect hidden grains covered by non-sediment objects with much higher detection rate, compared to regional CNN

approach (He et al., 2017; Soloy et al., 2020). Our second goal is to estimate streambed hydro-biogeochemical parameters based on YOLO-derived characteristic grain sizes and empirical equations for Manning coefficient (Rickenmann & Recking, 2011), Darcy–Weisbach friction factor (Ferguson, 2007, 2022), streambed interstitial velocity magnitude (Kenney et al., 1984; Y. Chen et al., 2021), and nitrate uptake velocity (Grant et al., 2018). Our third goal is to quantify uncertainties in both characteristic grain sizes and their propagation to the estimated HBGC parameters as well as the dominant sources of uncertainties in grain sizes and HBGC.

To achieve these goals, the paper is organized as follows: Section 2 introduces the study site, photo collection and grouping, training label generation, YOLO framework setup, as well as the equations used for HBGC and uncertainty calculation; Section 3 evaluates the YOLO model accuracy and reports the distributions and uncertainties of grain sizes and HBGC parameters; a thorough discussion covering the accuracy of grain sizes and HBGC, their major sources of uncertainty, the effects of photo number and probability threshold on model accuracy, potential automated photo resolution estimation strategy, as well as the limitations and future directions, is included in Section 4; the major results and implications are summarized in Section 5.

## 2 Methods

### 2.1 Photo acquisition and grouping

We obtained 2,121 photos from 75 sites at the Yakima River Basin (YRB) and the Columbia River section near the Port of Benton (Figure 1d) during 2021 – 2023. In 2021, we collected 383 photos from 47 sites; in 2022, we obtained 1,688 photos across 41 sites; in 2023, we took 50 photos from 3 sites near the Boat Ramp (BR) of the Leslie Groves Park. 6 camera types were used, including Samsung's SM-T500 tablet and Apple's iPhone 7, 12, 13, 13 Pro Max, and 14 Pro.

From these photos, we selected 61 photos as our training (36), validation (5), and testing (20) datasets. These datasets are mutually exclusive and labeled as 0, 1, and 2, respectively, for convenience (Figure 1a). To study the effects of the number of photos on model accuracy, we further divided the 36 training photos into three mutually inclusive groups, each having 11, 21, and 36 photos, respectively. For convenience, models trained with these groups are termed as model M0a, M0b, and M0c, respectively. In addition,

we trained a fourth model for scaling, termed as model Msc, to convert pixel size to real-world size using 50 photos (23 photos are from the 2,121 photos).

The 4 trained AI models were applied to predict both individual grains and reference scales for 2,143 photos. These photos were divided into 7 groups, labeled as 3 to 9, and each had 144, 1855, 24, 20, 21, 21, and 58 photos, respectively. Their roles are described as follows: the photos in group 3 and 4 are used to predict grain sizes of photos obtained in 2021 and 2022 (Figure 1b); the 20 photos in group 6 (same photos as group 2 in Figure 1a) are used to the test the accuracy of model M0a – M0c for predicting grain sizes; the photos in groups 5 (from iPhone 12), 7 (iPhone 13), 8 (iPhone 14 Pro), and 9 (Figure 1c) are used to evaluate the sensitivity of grain sizes and scaling to camera types and height as well as the accuracy of model Msc in predicting scales, respectively. The number of photos taken at each site is visualized in Figure 1d for reference. Details of site coordinates, grain sizes, and photo number can be found from our accompanying data package (Y. Chen et al., 2023).

### 2.2 Label generation

We manually generated labels (see label definition in Section 2.3) for both the grain detection AI models (M0a - M0c) and the scale detection AI model (Msc). For the grain detection models, we manually generated 16,951 labels from 61 photos, resulting in an average of 278 labels per photo (with a minimum of 19 and a maximum of 3,315). Out of these labels, 5,272 were used for training M0a, 10,154 for M0b, and 11,977 for M0c, respectively. For the scale detection model, we generated 121 labels from 50 photos representing 10 types of scales. These photos represent diverse flow, vegetation, and geological conditions in natural streams. 9 photos for the grain detection models and 6 photos for the scale detection model are illustrated in Figure 2 to visualize the environmental conditions and manually-generated labels (green dots bounded boxes). Photos a to i represent the following 9 conditions: dry bed, dry bed with high grain size ratio, dry bed with grass, dry bed with mud, partial-dry partial-wet mud, dry bed with ice, submerged bed with static water, submerged bed with flowing water and waves, hybrid rock/water/grass bed. Photos j to o represent 10 reference scales with known sizes, including, yellow tape 1, yellow tape 2, blue cap, green cap, tape measure, yellow paper board, quadrat net, color tapes, full quadrat, and white paper board. Their sizes are 7.05 cm × 1.7 cm, 7.1 cm × 2 cm, 3.7 cm, 2.5 cm, readable from tape measure, 11 cm × 11 cm, 20 cm × 20

cm, 2.54 cm in width, 80 cm × 80 cm, and 30.48 cm × 22.86 cm, respectively. The rest 52 photos for grain detection AI models and 44 photos for scale detection AI model and their labels can be found in the accompanying data package (Y. Chen et al., 2023).

### 2.3 YOLO framework

You Only Look Once (YOLO) is an object detection AI algorithm that is widely used for computer vision tasks (Redmon et al., 2016). In this study, the fifth major updated version was used and called YOLOv5. The Python implementation of YOLOv5 algorithm was open-sourced in 2020 by Ultralytics on GitHub (Ultralytics, 2020). YOLOv5 is a state-of-the-art real-time object detection system that is faster and more accurate than its predecessors.

A brief sketch of the YOLOv5 network flowchart is shown in Figure 3, which is summarized from GitHub (Ultralytics, 2020). Generally, it is constructed by a series of convolutional layers (Conv in Figure 3) (W. Zhang et al., 1990), modified bottleneck cross stage partial network layers (C3 in Figure 3) (Wang et al., 2020), a spatial pyramid pooling-fast layer (SPPF in Figure 3) (He et al., 2014), concatenate layers (Concat in Figure 3), and up-sampling layers. The fractional numbers, such as 1/2, 1/4, 1/8 and so on, in Figure 3 represent the relative image resolutions to the input image. For the convolutional layers in Figure 3, $Ch_i$, $Ch_o$, $k$, and $s$ stand for input image's number of channels, output image's number of channels, kernel size, and stride size, respectively. For the C3 layer, it reduces the number of convolutional layers from 4 to 3 in bottleneck cross stage partial network, which is originally connected to the output of bottleneck block (Wang et al., 2020). The value $n$ in Figure 3 stands for the number of bottleneck blocks in C3 layer. The spatial pyramid pooling-fast layer is a modified spatial pyramid pooling layer specifically designed for YOLOv5 with higher computational efficiency (Ultralytics, 2020). It concatenates several MaxPool layers (PyTorch, 2022) with different sizes for resolving the difficulties of detecting objects with various sizes.

The final outputs of YOLO, also called as labels, are the centroid ($x$ and $y$ in Figure 3), width ($w$ in Figure 3), height ($h$ in Figure 3), and class ($c$ in Figure 3) of the anchor box and the probability of the detected object in each class. The centroid and sizes of the anchor box are all normalized by the dimension of the original input image. In this study, we have 10 classes for reference scales (Section 2.2) and only one class for grain.

The network input is the image and the outputs are the corresponding labels. To avoid over-fitting, 5 labeled images were used for validation. During training, the optimizer does not consider the loss between the prediction of the validation images and true labels. The loss for the validation images is only used as a training termination criterion. With the predicted width and height of individual grains, we define the diagonal length of the grain, i.e., $D_p = \sqrt{w^2 + h^2}$, as the final grain size in pixel length, which can be converted to real size ($D$) by multiplying it with the estimated photo resolution.

### 2.4 Streambed hydro-biogeochemistry estimation equations

With given water depth ($H$) and flow velocity ($U$) as well as the photo-derived characteristic grain sizes, e.g., 5th ($D_5$), 50th ($D_{50}$), and 84th ($D_{84}$) percentiles of grain size distributions, key streambed hydro-biogeochemical parameters, including Manning's coefficient ($n$), Darcy–Weisbach friction factor ($f$), shear velocity ($u_\tau$), streambed interstitial velocity magnitude ($\sigma_w$), and streambed nitrate uptake velocity ($u_f$) can be estimated by Equations 1 (Rickenmann & Recking, 2011), 2 (Ferguson, 2007, 2022), 3 (Y. Chen et al., 2021; Kenney et al., 1984), and 4 (Grant et al., 2018), respectively. The water depth is a reach average depth, which was estimated using a wading-based depth transect procedure. The details of such a procedure can be found in the field protocol described in our data package published in US DOE's Environmental System Science Data Infrastructure for a Virtual Ecosystem (ESS-DIVE) (Delgado et al., 2023). The velocity is the average velocity for August between February 1979 and December 2020, which was computed by Kaufman et al. (2023) from the National Oceanic and Atmospheric Administration's National Water Model version 2.1 (NOAA, 2023).

$$n = \frac{D_{84}^{1/6}}{20.4} \tag{1}$$

$$\sqrt{\frac{8}{f}} = \frac{U}{u_\tau} = \frac{c_1 c_2 H/D_{84}}{\sqrt{c_1^2 + c_2^2 (H/D_{84})^{5/3}}} \tag{2}$$

$$\sigma_w = c_3 \frac{g k_I}{\nu} \frac{U^2}{g D_{50}} (\frac{H}{D_{50}})^{c_4}, k_I = c_5 D_5^2 \tag{3}$$

$$u_f = k_m \phi, k_m = 0.17 u_\tau Sc^{-2/3}, Sc = \frac{\nu}{D_m}, \phi = c_6 [\mathrm{NO_3^-}]^{c_7} \tag{4}$$

The constants used in the above equations are: $c_1$ = 6.5, $c_2$ = 2.5 (Ferguson, 2022); $c_3$ = 0.88 (range 0.62 – 1.11), $c_4$ = -0.66 (Y. Chen et al., 2021); $c_5 = 1{\times}10^{-9}$ (Kenney et al., 1984); $c_6$ = 0.0032, $c_7$ = -0.49 (Grant et al., 2018); gravity acceleration $g$ = 9.81 m/s$^2$, water viscosity $\nu = 1 \times 10^{-6}$ m$^2$/s, nitrate molecular diffusion in water $D_m$ =

$1.7\times10^{-9}$ $m^2/s$ (Picioreanu et al., 1997). Non-constant parameters include subsurface intrinsic permeability $k_I$ ($m^2$), hydrogeology-biochemistry interaction efficiency $\phi$, Schmidt number $Sc$, and stream nitrate concentration $[NO_3^-]$ ($mol/m^3$, equivalent to 62 mg/L). Our field survey in 2021 shows that the nitrate concentration in YRB varies between 0.0005 and 0.1 with a mean of 0.008 $mol/m^3$ (Grieger et al., 2022). In 2022, stream nitrate concentrations are not available for all locations where depth were measured, therefore, we select three values, 0.0001, 0.01, and 1 $mol/m^3$, to represent the typical magnitudes reported at the YRB and in the literature (Mulholland et al., 2008; Grant et al., 2018; X. Zhang et al., 2021; Sadayappan et al., 2022).

### 2.5 Uncertainty quantification for grain sizes and hydro-biogeochemistry

Uncertainties occur in grain detection, scaling, and the propagation from grain sizes to hydro-biogeochemical parameter estimations. For any given photo, the real grain size $D_x$ ($x$ = 5, 50, and 84) are calculated by $D_x = D_{xp}SC$ with the $D_{xp}$ and $SC$ denoting the grain size measured by pixel number and the photo resolution measured by real size per pixel. The $D_{xp}$ is determined by YOLO and its uncertainty $r_{xp}$, quantified by the average absolute relative error of testing photos, can be directly estimated by comparing the YOLO-predicted and manually measured grain sizes. For photo-resolution uncertainty, we manually draw two straight lines for all photos following the scales showing in Figure 2 and then calculate the relative error ($r_{SC}$) between the photo resolution calculated from the two lines. With the estimation of pixel-based grain size uncertainty and scale uncertainty, the real-world grain size uncertainty and its propagation to HBGC parameters can be estimated by Equations 5 – 9 based on the law of propagation of uncertainty (Ku, 1966). The detailed mathematical derivation of these equations can be found in Appendix. The $r_H$ is the mean absolute relative difference between the measured water depth ($H$) and its time-average value over the observation period (around 1 month in August 2022). The uncertainty measurement for flow velocity ($r_U$) and stream nitrate concentration ($r_N$) are not available for the study sites. However, existing literature report that velocity measurement uncertainty by Acoustic Doppler current profilers (ADCPs) could range 1% to 25% depending on the distance away from the ADCPs (Mueller et al., 2007) and stream nitrate concentration uncertainty is 12% on average across 7 watersheds in US (Jiang et al., 2014). Therefore, we choose 10% as a rough

estimation of the typical measurement uncertainty for stream velocity and nitrate concentration in this work.

$$r_x = \sqrt{{r_{xp}}^2 + {r_{SC}}^2}, x = 5, 50, 84 \tag{5}$$

$$r_n = r_{84}/6 \tag{6}$$

$$r_f = 2\{1 - \frac{5}{6}[1 + \frac{c_1^2}{c_2^2}(\frac{H}{D_{84}})^{-5/3}]^{-1}\}\sqrt{r_H^2 + r_{84}^2} \tag{7}$$

$$r_w = \sqrt{4r_U^2 + (1 - c_4)^2 r_{50}^2 + c_4^2 r_H^2} \tag{8}$$

$$r_{uf} = \sqrt{r_U^2 + r_f^2/4 + c_7^2 r_N^2} \tag{9}$$

## 3 Results

### 3.1 YOLO performance

We evaluate the performance of YOLO through four metrics: the mean average precision (mAP) of the YOLO training, the accuracy of grain size distribution, median grain sizes, and their relative error (Figure 4). The mAP@50 and mAP@50-95 are two typical metrics used to quantify the accuracy of object detection AI algorithm. The symbol @50 means the prediction is correct if the intersection over union (IoU) larger than 50%. The IoU stands for the relative overlapping area between the predicted object bounding box and the ground truth object bounding box. Similarly, the symbol @50-95 means the prediction is correct if the IoU larger than 50% to 95% with 5% increase interval. Additional 5 photos with 954 labeled grains are used as validation data set. The accuracy of the prediction on the 5 validation photos are not seen by the optimizer, and it is only used to track the model accuracy during training and helps on determination of the best model, as shown in Figure 4(a). The weighted mAP (10% of mAP@50 and 90% of mAP@50-95) is used as final accuracy metric, and it reaches the maximum at 968 steps (Figure 4a: vertical dashed line). The corresponding mAP@50 and mAP@50-95 at this step is 0.64 and 0.34, respectively (Figure 4a: horizontal dashed lines). After 968 training steps, both mAP@50 and mAP@50-95 decrease, with no indication that the accuracy can increase within 20,000 training steps. Therefore, the trained model, which is used for all the results in the study, is the model stored at 968 training steps. For Microsoft Common Objects in Context (COCO) dataset, a commonly used benchmark dataset for object detection AI, typical values for mAP@50 and mAP@50-95 fall in the range

0.46 – 0.73 and 0.28 – 0.56, respectively (Ultralytics, 2020). In our case, the shape, sizes, color, transparency, lighting, and environmental conditions are more complex than those photos used in COCO (Figure 2), however, the model still achieves 0.64 and 0.34 values for mAP@50 and mAP@50-95 on the validation photos, respectively (Figure 4a). This means the YOLO training achieved a good performance.

To illustrate the model's capability in extracting grain size distributions (GSDs), Figure 4b shows a comparison of the area-weighted GSD between the model prediction (blue line) and manual labels (red line). The cumulative probability in calculated by $P_i = \sum A_i(D \leq D_i)/\sum A_i$ with $A_i$ and $D_i$ denoting the area and size of each grain. The minimum difference between the two lines demonstrates that the area-weighted GSD is accurately reproduced by the trained model. Similar comparisons for the remaining 19 photos used for testing are not included here for simplicity, however, can be found in Figure 12. These comparisons demonstrate that the GSDs can be well reproduced by YOLO algorithms for most (18 of 20) photos.

Based on the GSD curves, the median grain size D50, defined as the grain size corresponding to 50% finer grain sizes, can be calculated from the GSDs of the 20 testing photos. Figure 4c shows a one-to-one plot between the predicted D50 and manually estimated D50. The result shows that YOLO predicts D50 with an accuracy of 0.98, 0.98, -0.037 cm, and 0.91 cm in terms of R-squared, Nash–Sutcliffe efficiency (NSE), mean error, and root-mean-square between the prediction and manual measurements. To further examine such accuracy, Figure 4d shows the relative error between the predicted D50 and manually estimated D50. The result shows 90% (18 dots) of the data points demonstrate a relative error less than 10% and 10% (2 dots) show a relative error larger than 20%. On average, the mean absolute relative error is 6.65% for the 20 testing photos. The result also shows the relative error does not correlate with the grain size, which suggests the accuracy of YOLO is stable for both small and large grains.

### 3.2 Characteristic grain size distributions

With the confirmed high accuracy of the YOLO model, we apply the model to extract the grain size distributions (GSDs) from 1,999 photos (66 sites) in groups 3 and 4, and then calculate the characteristic grain sizes, e.g., D5, D50, and D84, from the GSDs. As valid water depth measurements are available at only 41 sites, Figure 5 shows only

the results of characteristic grain sizes from 1,745 photos obtained at the 41 sites to make a consistent evaluation for HBGC parameters in Section 3.3. In general, the three grain size distributions follow log-normal distributions (black solid lines in Figure 5a-c are fitted Gaussian distributions) with the log2-transformed mean of 4.15, 6.05, 6.75 and standard deviation of 0.86, 0.87, and 0.81 for D5, D50, and D84, respectively. This means the most likely sizes of D5, D50, and D84 are around 1.78 cm, 6.63 cm, and 10.76 cm, respectively. As D5, D50, and D84 represent different importance of grain sizes in controlling HBGC, Figure 5d further shows the relationship between D5 and D50 and that between D84 and D50. The result shows that D5 and D84 increase linearly with D50, although there are some large residuals.

### 3.3 Streambed hydro-biogeochemistry distributions

With the photo-derived characteristic grain sizes (D5, D50, and D84), measured water depth, extracted velocity, and assumed typical stream nitrate concentration (see details in Section 2.4), the HBGC parameters can be estimated using Equations 1 - 4. To mitigate the uncertainty resulting from an insufficient number of photos, we show results only from sites with more than 3 photos. Consequently, we are showing the results from 1,737 photos at 37 sites (refer to site locations in Figure 6b).

Overall, HBGC parameters demonstrate different distribution patterns compared to grain sizes. Specifically, the Manning coefficient follows a normal distribution (black line in Figure 6a) with a mean and standard deviation of 0.0339 and 0.0031 $s \cdot m^{-1/3}$, respectively. The log10-transformed friction factor, log10(f), shows a positively skewed distribution (Figure 6c) with its skewness (defined as the adjusted Fisher-Pearson skewness coefficient), mean, median, mode, and standard deviation of 0.43, -0.54, -0.58, -0.75, and 0.37, respectively. This suggests the friction factor has the most likely value of 0.18 ($=10^{-0.75}$), which falls in the range of 0.13 – 0.32 calculated from high-resolution computational fluid dynamics simulations for natural gravel bed rivers with median grain size of 6 cm (Y. Chen et al., 2019). The log10-transformed streambed interstitial velocity magnitude, log10($\sigma_w$), follows a near-Gaussian distribution (Figure 6e) with skewness, mean, median, mode, and standard deviation of -0.03, -1.07, -1.08, -1.15, and 0.52, respectively. This suggests the streambed interstitial velocity magnitude has a high likelihood at the scale of 0.07 ($=10^{-1.15}$) m/day for the study region, which is close to the value (0.11 m/day) estimated by a temperature-based data assimilation approach applied at the Hanford reach of the

Columbia River (K. Chen et al., 2023). The distribution of the nitrate uptake velocity is more complex. Firstly, the distribution is strongly affected by the concentration of stream nitrate. It may decrease 3 orders of magnitude if the nitrate concentration increases from 1e-4 mmol/L (=0.0062 mg/L) (Figure 6g blue histogram) to 1 mmol/L (=62 mg/L) (Figure 6g gold histogram). The median and mean values of stream nitrate concentration were reported at the order of 1e-2 mmol/L (=0.62 mg/L) over 72 agriculture and urban sites in US (Grant et al., 2018). The mean nitrate concentration in the YRB was also reported at a similar magnitude of 0.008 mmol/L (Grieger et al., 2022). Therefore, it is reasonable to use 0.01 mmol/L as the most likely magnitude of nitrate concentration in US. Using such a concentration, the nitrate uptake velocity varies between 0.23 and 5.6 m/day and shows a negatively skewed distribution with the skewness, mean, median, mode, and standard deviation of -0.23, 0.013, 0.036, 0.075, and 0.22, respectively (Figure 6g gold histogram). This means the nitrate uptake velocity has a high chance to be 1.2 ($=10^{0.075}$) m/day with a US median or mean nitrate conditions. This value is in the range between measured median (0.6 m/day) and mean (2.5 m/day) uptake velocity across the US (Grant et al., 2018).

The left panels of Figure 6 illustrate the overall distributions of HBGC parameters but not their spatial variations. To visualize the spatial variations, the right panels show the spatial distributions of site average HBGC parameters. The number of photos at each site can be found on Figure 1d. Figure 6b shows that the site average Manning coefficient mostly clusters at red (0.035 - 0.0375 $s{\cdot}m^{-1/3}$) and light red (0.0325 - 0.035 $s{\cdot}m^{-1/3}$), which means the site average Manning coefficient has a low spatial heterogeneity. Such a behavior can also be observed in Figure 8a where the site average value (black line) of Manning coefficient shows small variation across the sites. In contrast, the site-average friction factor exhibits greater heterogeneity, as indicated by the diverse range of colors in Figure 6d. The highest log10-transformed friction factor values (0 – 0.25) occur at site S37, S39, and W10, followed by 8 sites (W20, S04, S03, S42, S10, S53, S56N, and S48R) in the group -0.25 – 0. The lowest values (yellow dots at group -1 – -0.75) occur at S02, T02, T03, and S23, and the rest of the data points share similar colors. This behavior can also be observed in Figure 8b (see black line). Different from the friction factor, the log10-transformed interstitial velocity magnitude has maximum values at sites S04, S58, S18R, T05P, S50P, and S56N (Figures 6f dark red and 8c black line), followed by the value group -0.75 – -0.25 (red) at 5 sites (S48R, S10, S01, W10, and S31). The

lowest interstitial velocity occurs at the sites S42 and S43 with a value of around -2 (Figures 6f yellow and 8c black line). Compared to the friction factor and interstitial velocity, the uptake velocity distribution demonstrates obvious hot spot at site S04 (dark red) and cold spots (yellow) at sites T02, S41R, S42, and S43 with a value of 2.8 m/day and a range of 0.3 – 0.5 m/day, respectively. Interestingly, the cold spots are all within or downstream of the Yakama Indian Reservation region. It is also interesting to mention that the hot (S04) and cold (S42 and S43) spots in nitrate uptake velocity are also the hot and cold spots in the interstitial velocity. This suggests the hot/cold spots in denitrification are likely affected by the water exchange between stream and groundwater in the YRB. This is consistent with the work of Son et al. (2022) that shows hyporheic exchange flux is the most important factor controlling nitrate removal based on data from basin-scale numerical simulations and random forest relative importance analyses.

### 3.4 Uncertainty in characteristic grain sizes

With the uncertainty quantification equations introduced in Section 2.5, the uncertainty or variability associated with manually-measured photo resolution, YOLO-derived grain sizes, and water depth observations can be estimated for each photo. Figure 7a shows the manually-measured photo resolution (blue cross) and the relative error $r_{SC}$ (yellow line) associated with each resolution. The results shows that around 90% of the photos have a resolution of around 0.1 mm/pixel (corresponding to 1/4 of the quadrat in Figure 2n, o), and 10% of the photos have a resolution between 0.2 and 0.7 mm/pixel (corresponding to the full quadrat in Figure 2n, o). The relative error for these scales, however, are mostly in the range -10% – 10% and have an overall mean and mean absolute error of 0.13% and 2.3%, respectively. This means the photo resolution estimation has no systematic bias and the manual measurement uncertainty is low enough for further grain size quantification.

With the photo resolution uncertainty ($r_{SC}$), the uncertainty in D50, D84, and D5 can be calculated by Equation 5 with the YOLO-associated grain size uncertainty $r_{50p}$ (=6.65%), $r_{84p}$ (=10.65%), and $r_{5p}$ (=11.88%) directly estimated from the average absolute relative error of testing photos as discussed in Section 3.1. Figures 7b,c,d show the combined effects of photo resolution uncertainty and YOLO accuracy uncertainty for D50, D84, and D5, respectively. The result shows the uncertainty of D50 varies between 6.65% and 13.53% with a mean value of 7.33%. For D84 uncertainty, its minimum,

maximum, and mean are 10.65%, 15.88%, and 11.11%, respectively. For D5 uncertainty, these values are 11.88%, 16.73%, and 12.30%, respectively.

The water depth is estimated every 1 minute during July 28 and August 31 2022 (see details in data package (Delgado et al., 2023)). With these data, the depth ($H$) is calculated as the time averaged depth over the whole measurement period. The uncertainty or variability ($r_H$) of such a depth is calculated as the average absolute relative difference between the actual depth and the calculated mean depth. Figure 7e shows the variations of the mean depth and its variability at each site. The result shows the depth varies between 0.14 m and 2.11 m, with a mean of 0.45 m across all the sites. Highest depth occurs at sites T02 and T03 while depth less than 0.25 m are found at 9 sites (S63, S53, S04, S37, S39, S03, W10, W20, and S42). The depth variability varies between 0.66% and 30.2% with a mean 6.6%. High depth uncertainty is observed at sites S56N, S24, and S18R.

### 3.5 Uncertainty in hydro-biogeochemistry

With the quantification of uncertainties for grain sizes, depth, and assumed typical measurement uncertainty in velocity and nitrate concentration (see details in Section 2.5), Figure 8 shows all calculated values (blue cross dots), site-average values (black lines), and estimated uncertainty (yellow lines) for Manning's $n$, friction factor $f$, streambed interstitial velocity magnitude $\sigma_w$, and streambed nitrate uptake velocity $u_f$. It is observed that the Manning coefficient varies in a range 0.0245 – 0.0455 $\mathrm{s{\cdot}m^{-1/3}}$ with low uncertainty range of 1.78% – 2.61% (Figure 8a). The friction factor, by contrast, spans over 2 order of magnitude (0.04 – 9) and its uncertainty has minimum, maximum, and average of 3.63%, 58.36%, and 15.65%, respectively. The highest uncertainty occurs at site S56N (Figure 8b yellow line). The interstitial velocity magnitude spans even larger ranges from 0.0038 to 2.31 m/day. However, its uncertainty range is lower than the friction factor, which has minimum, maximum, and average of 22.84%, 32.11%, and 24.06%, respectively. The highest uncertainty is observed at site S56N (Figure 8c yellow line). The nitrate uptake velocity shows a lower variation range between 0.23 and 5.6 m/day. The highest uptake velocity occurs at site S04 while the lowest values occur at sites S42 and S43 (Figure 8d black line). The highest uncertainty occurs at site 56N (Figure 8d yellow line), which is similar to those observed for friction factor and interstitial velocity magnitude. Overall, the uptake velocity uncertainty is estimated as 11.28%, 31.23%,

and 13.88% in terms of the minimum, maximum, and average value. It is worth noting that the results for uptake velocity are based on US mean nitrate concentration (0.01 mmol/L). Therefore, the uptake velocity variation range will change with nitrate concentration at other sites, however, its uncertainty may be similar if the depth and grain size conditions are similar.

# 4 Discussion

## 4.1 Accuracy of grain sizes and hydro-biogeochemistry parameters

To apply the present approach to other rivers, it is important to evaluate the accuracy of the YOLO-derived grain sizes and grain size-based HBGC estimations. As percentile-based grain sizes are derived from the grain size distribution (GSD) curve, the accuracy of GSD determines the accuracy of characteristic grain sizes, e.g., D50, D84, and D5. As demonstrated in Figure 4b and Figure 12, the pre-trained YOLO can reproduce the GSDs with high accuracy for 90% (18 out of 20) of the testing photos that represent 9 different streamed conditions. Under these diverse conditions, the median grain sizes calculated from these GSDs demonstrate relative errors less than 10% (Figure 4d). These results indicate that GSDs and subsequently derived characteristic grain sizes are accurate, at least, for the majority (90%) of the photos. Even though two (10%) testing photos (Figure 12(f,r)) show larger error in GSD, the overall accuracy of all the testing photos, as indicated by an R2 value of 0.98, an NSE value of 0.98, and a mean absolute relative error of 6.65%, is still suitable for practical applications. A closer examination of the two photos (Figure 12(f,r)) with higher error shows that the error is likely caused by the unclear boundaries between the largest grains and ambient smaller sediments, due to light reflection and flocculation on wet grain surface and water surface. Future work may be needed to address these challenges to further improve grain size accuracy.

With the YOLO-derived characteristic grain sizes, using the equations introduced in Section 2.4 to estimate the streambed HBGC parameters will undoubtedly bring errors, partially from the limitation of the equations themselves, and partially from the propagation of uncertainties in input parameters. Though it is challenging to measure HBGC at all study sites, we are able to identify measured or calibrated data for HBGC from existing literature, and can evaluate the accuracy of the photo-driven, AI-enabled, and theory-based estimations for HBGC. Firstly, the well-calibrated Manning's coeffi-

cients from a two-dimensional hydraulic model for the Columbia River vary between 0.027 – 0.038 s/m$^{1/3}$ (Niehus et al., 2014), which is close to the range calculated from all photos (Figure 6a: 0.0245 – 0.0455 s/m$^{1/3}$) and site average value (Figure 6b: 0.0281 – 0.0373 s/m$^{1/3}$). Secondly, the flow resistance from 2,890 field measurements vary between 0.02 and 200 for rivers with $H/D_{84} < 200$ (Rickenmann & Recking, 2011), which covers the range derived from all photos (Figure 6c: 0.04 – 9) and site-average values (Figure 6d: 0.06 – 1.5). Meanwhile, the maximum likelihood of friction factor occurs at 0.18 (=$10^{-0.75}$) (Figure 6c), which falls in the range of 0.13 – 0.32 computed from high-resolution computational fluid dynamics simulations for natural gravel bed rivers with a median grain size of 6 cm (Y. Chen et al., 2019), a value very close to the most likely median size (6.63 cm) observed in our study area (Section 3.2). Regarding the interstitial velocity, direct field measurements are rare. However, by using a temperature-based data assimilation approach, K. Chen et al. (2023) were able to estimate the time series of vertical hydrological exchange flux at the Hanford Reach of the Columbia River. Using their data (Figure S5a in K. Chen et al. (2023)), the interstitial velocity magnitude is estimated as 0.11 m/day by calculating the ratio of the standard deviation of estimated hydrological exchange flux time series to the subsurface porosity (0.43) reported in their work. As demonstrated in Section 3.3, the most likely value of interstitial velocity is around 0.07 m/day (Figure 6e). This suggests most of the estimated interstitial velocity magnitude falls in the observation range. For the streambed nitrate uptake velocity, if the stream nitrate concentration is at the US mean or median level, i.e., 0.01 mmol/L (Grant et al., 2018), the estimated uptake velocity is most likely at the scale of 1.2 m/day, which is between the median (0.6 m/day) and mean (2.5 m/day) uptake velocity measured at 72 sites in US (Grant et al., 2018). The above comparisons, therefore, suggest that photos can be used to make reasonable estimates of HBGC parameters, using AI and empirical equations.

### 4.2 Major sources of uncertainty

Though Section 4.1 demonstrates the accuracy of estimating grain sizes and HBGC, it is still important to quantify potential uncertainties in these estimations. This is necessary to reduce measurement uncertainties in field work and evaluate their impacts on large-scale watershed models. With the use of explicit mathematical formulas, the uncertainties in grain sizes and HBGC can be mathematically accurately derived as shown

in Equations 5 - 9. From these equations, we can see that the uncertainty of YOLO model ($r_{xp}$) and photo resolution ($r_{SC}$) are propagated to the characteristic grain sizes ($r_x$). As demonstrated in Section 3.4, the overall uncertainty for YOLO model is 6.65%, 10.65%, and 11.88% in predicting D50, D84, and D5 pixel sizes, while that for photo resolution is 2.32%. Therefore, the average compounding uncertainty (based on Equation 5) in D50, D84, and D5 are 7.33%, 11.11%, and 12.30%, respectively. Such grain size uncertainties are further propagated to Manning coefficient through $r_n = r_{84}/6$, which results in low uncertainty (mean value 1.85%) in estimating Manning coefficient. The uncertainty in friction factor is more complex because it depends on not only input parameter uncertainty (depth uncertainty $r_H$ and grain size uncertainty $r_{84}$), but also the ratio of water depth to grain size. Despite such complexity, its uncertainty should vary between 1/3 to 2 times of the compounding uncertainty of water depth and D84 ($r_{H_{D84}}$) because the depth/grain size dependent term reduces to 1/3 and 2 for very deep ($H \gg D_{84}$) and shallow water ($H \ll D_{84}$). As the average uncertainty in depth and D84 are 6.6% (Section 3.4) and 11.11%, respectively, their compounding uncertainty is 12.92% ($=\sqrt{r_H^2 + r_{84}^2}$. Therefore, the overall uncertainty of friction factor should vary between 4.31% and 25.85%, which agrees with the average friction factor uncertainty of 15.65% as mentioned in Section 3.5. The uncertainty in interstitial velocity magnitude is simpler because it only depends on the uncertainties of three input parameters: velocity, grain size, and depth. In this work, as the velocity uncertainty is not available, we assume an uncertainty level of 10% based on previous work on velocity measurements with ADCPs (Mueller et al., 2007). As the overall uncertainty in grain size D50 and depth are 7.33% and 6.6%, the overall compounding uncertainty from the three input parameters is around 23.81% (computed from Equation 8) which is close to the average uncertainty (24.06%) calculated from Figure 8c (see Section 3.5).

The uncertainty in nitrate uptake velocity is much more complex because it depends on the uncertainty in velocity, nitrate, and the friction factor that further depends on the values and uncertainties in depth and grain sizes. Such complexity can be verified by Figure 8d where large changes in uptake velocity uncertainty (yellow line) are observed. As the mean uncertainty in friction factor can be estimated by $r_f^m = c_0\sqrt{r_H^2 + r_{84}^2}$ with $c_0$ in the range 1/3 – 2, the mean uncertainty in uptake velocity ($r_{uf}^m$) can be estimated by Equation 10. As the measured nitrate uptake uncertainty is not available, a 10% uncertainty is assumed based on previous work on nitrate measurement uncertainty (Jiang

et al., 2014). With the overall uncertainty for velocity (10%), depth (6.6%), pixel D84 (10.65%), photo resolution (2.32%), and nitrate (10%), the overall uptake velocity uncertainty should fall in the range of $r_{uf}^{md}$ and $r_{uf}^{ms}$ with $r_{uf}^{md}$ and $r_{uf}^{ms}$ representing the mean characteristic uncertainty in deep and shallow rivers. Here the two terms are calculated by $r_{uf}^{md} = r_{uf}^{m}(c_0 = 1/3)$ and $r_{uf}^{ms} = r_{uf}^{m}(c_0 = 2)$ and their values are 11.34% and 16.92%, respectively. As mentioned in Section 3.5, the average uncertainty in uptake velocity calculated from Figure 8d (yellow line) is 13.88%, which falls in the range of characteristic uncertainty. Therefore, the Equation 10 can be used as a fast estimate of the uncertainty in uptake velocity if the uncertainty of 5 inputs are available.

Equation 10 also suggests that the final uncertainty depends on whether the constant $c_0$ leans to the upper bound (2) or the lower bound (1/3), which is mainly determined by the ratio of water depth to grain size D84. In shallow water ($c_0 = 2$) condition, the dominant sources of uncertainties will be velocity, depth, and YOLO-accuracy for D84 because $c_0^2/4 = 1$ and $c_7^2 \approx 0.24$. In deep water ($c_0 = 1/3$), the main sources will be velocity and nitrate concentration because $c_0^2/4 \approx 0.03$. Another important aspect of such an equation is that the uncertainties in velocity, depth, and nitrate concentration represent clear physical meaning, while the uncertainties in pixel D84 and photo resolution are instead associated with AI model and photo induced uncertainties. With further improvements of AI training and photo resolution estimation, these nonphysical uncertainties can likely be reduced to a negligible level (see details in Sections 4.3 - 4.5), and Equation 10 can be reduced to Equation 11 that represents physics-driven uncertainty for uptake velocity. Furthermore, in very dynamic unsteady processes, the uncertainty terms, $r_U$, $r_H$, and $r_N$, more represent the deviation of the actual physical processes away from their time average values, therefore, the compounding uncertainty in Equation 11 can be treated as a metric to quantify the magnitude of the dynamics in nitrate uptake processes.

$$r_{uf}^{m} = \sqrt{r_U^2 + \frac{c_0^2}{4}(r_H^2 + r_{84p}^2 + r_{SC}^2) + c_7^2 r_N^2} \tag{10}$$

$$r_{uf}^{mp} = \sqrt{r_U^2 + \frac{c_0^2}{4} r_H^2 + c_7^2 r_N^2} \tag{11}$$

### 4.3 Effects of photo number

To minimize non-physical uncertainties from the AI model, one way is to increase the number of training photos and labels. Figure 9 shows the effects of photo number on AI-training convergence and accuracy in predicting grain size distribution and characteristic size such as D50. Here the M0a, M0b, and M0c represent three models trained with 11 (5,272 labels), 21 (10,154 labels), and 36 (11,977 labels) photos (see photo locations in Figure 1a and label preparation in Section 2.2). The results show that increasing the number of photos improves the accuracy of the YOLO model, with mAP@50 increasing from 0.54 to 0.64 and mAP@50-95 increasing from 0.28 to 0.34.

Though the model metrics are improved, their accuracy improvements in predicting grain size distributions and D50 depend on the complexity of the streambed. For the dry bed with large grain size ratio (Figure 9b), all three models provide accurate prediction of the GSD though the M0c model (blue line) performs better in capturing smaller grains (<50% percentage finer) and M0a model (black line) performs better in capturing larger grains (>50% percentage finer) when compared to manual measurements (dashed red line). For submerged bed with static water (Figure 9c), the M0c model outperforms M0a and M0b for most of the sizes (<80% percentage finer).

A systematic evaluation of the model accuracy is illustrated in Figure 9d-e in terms of the 1:1 plot between the model-predicted and measured D50 as well as the relative error of predicted D50. The result shows that the M0c model outperforms M0a and M0b in terms of higher R2 (0.98 vs 0.92) and closer alignment with the 1:1 line for all the points (Figure 9d). The closer alignment of model M0c can also be verified in Figure 9e where we can observe 18 points (black circle dots) in the range ± 10% for M0c while those for M0a and M0b are 13 points despite including the points outside but close to the range. The mean absolute relative error for M0a, M0b, and M0c, with values of 11.88%, 11.20%, and 6.65%, also point to the much better performance in M0c.

With available manual labels, it is straightforward to evaluate the model's accuracy. However, it is impractical to manually draw grain sizes for all 1,999 photos used in groups 3 and 4 for prediction purpose (see Section 2.1). Nevertheless, we can evaluate the differences in predicted D50 between the higher accuracy model M0c and the lower quality models as shown in Figure 9f. Statistically, the bias and root-mean-square between M0a and M0c are -0.26 and 2.85 cm; and that between M0b and M0c are -1.22

and 3.14 cm, respectively. As the most likely D50 is 6.63 cm (obtained from M0c model; Section 3.2) and 47% (821 out of 1743 points) of the grain sizes are less than such a value, the uncertainty induced by lower quality models is likely important. Therefore, it is critical to train the YOLO with sufficient data in order to avoid systematic impacts on grain size quantification and subsequent HBGC estimation. In the context of grain size prediction, the number of sufficient data may be determined by checking if the mean absolute relative error between the model prediction and testing labels becomes smaller or comparable to typical uncertainties in field observations or other manual approaches.

### 4.4 Effects of YOLO probability threshold

Another factor that affects the YOLO accuracy is the selection of the probability threshold built in YOLO. A probability threshold is required because the YOLO uses a probability, in the range $0 - 1$, to determine whether an object (grain, grass, water, etc.) in a photo is the target object (e.g., grain in this work). Under-estimation (small value) of the threshold will select too many objects that are not the target, but over-estimation (high value) will ignore objects that are desired. To identify a proper way of selecting the threshold, Figure 10 shows the variation of R2, mean error (ME), mean absolute error (MAE), and the average detected grain number per photo between the prediction (from model M0c) and manual labels, with respect to probability threshold. The best probability threshold should maximize R2, minimize ME and MAE, and identify the number of grain sizes closest to manual measurements. Following these rules, 0.35 is selected as the final probability threshold because R2 reaches maximum (Figure 10a), ME is nearest 0, MAE is at its minimum (Figure 10b), and the number of grains per photo is closest to the manually measured number (Figure 10c). Grains with a YOLO probability less than 0.35 are excluded from the grain size quantification. It is worth mentioning that selecting the probability threshold is a well constrained problem because simultaneously minimizing the ME and identifying the closest number of grains will likely lead to a unique value.

### 4.5 Estimation of photo resolution

How to properly estimate the photo resolution affects not only the accuracy of grain sizes, HBGC parameters, and their compounding uncertainties, but also the efficiency of data collection and post-processing. In general, photo resolution could be estimated

manually or automatically. The manual approach is easy for field implementation, but prone to human error and high data processing costs. In this work, we brought full quadrats and white boards with known sizes into the field, placed them on top of the grains, took photos, manually measured the pixel length of the known scales, and finally obtained the photo resolution, represented by millimeter per pixel (Figure 11a). The manual scale measurement process for 2,121 photos involves 8 person and costs around 200 hours of human labor. Large errors occur due to the unevenness of the quadrats/boards, inaccurate recording of the pixel coordinates from the computer screen, and matching the coordinates to incorrect photo names. To mitigate such errors and reduce costs, an automated scaling approach is desired. Figure 11 illustrates how an automated scaling could be implemented and whether such approaches could be comparable to the manual approach in terms of the resolution and minimum detectable sizes.

It is observed from Figure 11a that the photo resolution clusters at two ranges, i.e., 0.066 – 0.15 and 0.3 – 0.7 mm/pixel (see scale for each photo on Figure 7a and discussion in Section 3.4) and the detectable minimum grain sizes from all photos in groups 3 and 4 vary between 0.82 mm and 21 mm. The typical reference scales for the higher (red star) and lower (blue diamond) photo resolution are visualized in Appendix Figure 13(a,b), respectively. From these figures, we can see that the pixel lengths of the quadrat (white pipes) and strings (red lines) are skewed, which brings errors to resolution estimation and difficulties in manual measurements.

To expedite the photo resolution estimation, a potential way is to train a scale AI model, e.g., model Msc (see details in Sections 2.1 – 2.2), and then use it to measure the pixel sizes of the reference scales automatically. The trained Msc model can detect 10 different scales as mentioned in Section 2.2. However, the accuracy is low for all non-circular shaped reference scales because the YOLO can only use horizontally-placed rectangular boxes (see green line bounding boxes in Appendix Figure 13a,b) to capture the reference scales which could be non-horizontally placed and non-rectangular shape. Interestingly, all the scales with circular shape (e.g., green and blue caps) are accurately detected by the trained scale model at both submerged and dry conditions (Figure 13c,d). For those photos in group 9 (used for scale AI validation) with green/blue caps, we manually measured the photo resolution and then compared their values with those predicted by the scale AI model as shown in Figure 11b. The result verifies the visual observation in Figure 13c,d and provides an accuracy estimation of such an automated approach. For

the blue caps (3.7 cm diameter): the mean error (ME), mean absolute error (MAE), mean absolute relative error (MARE), minimum relative error, and maximum relative error are 0.0039 mm, 0.0065 mm, 3.2%, -5.3%, and 7.5%, respectively. For the green caps (2.5 cm diameter), their values are -0.0006 mm, 0.002 mm, 1.1%, -2.1%, and 1.4%, respectively. Note that the photo resolution uncertainty from manual estimation varies between ±10% and has an overall MARE of 2.32%. This means the cap-based automated scaling approach has a better overall accuracy and a much smaller uncertainty range than the manual approach. Meanwhile, the automated scaling can provide photo resolution of 0.12 – 0.35 mm/pixel, which is also better than the range obtained in the manual approach. Overall, the cap-based automated scaling approach is an efficient alternative to the manual approach in terms of accuracy and resolution.

Both the manual and automated approaches mentioned above are limited for locations we have site accessibility and working permits where we are able to deploy reference scales and use hand-held cameras. These limitations restrict the spatial scale we can observe. Overcoming such limitations necessitates the use of fast remote sensing techniques, such as drones, and requires an approach to reliably estimate the photo resolution captured by the drone cameras. Here we show that the photo resolution can be estimated based on camera height and camera-specific resolution-height relationships. Figure 11c shows the variation of photo resolution (from manual measurements) with respect to height for 3 smartphones, i.e., iPhone 12, 13, and 14 Pro (see photo taken locations in Section 2.1). These relationships provide an additional way to estimate photo resolution for both hand-held and unmanned devices if height information is available.

### 4.6 Limitations

Despite the promise of the proposed approach, limitations exist in photo collection, training data preparation, and HBGC empirical formulas. First of all, by using hand-held devices (e.g., smartphones, tablets, and cameras), the maximum spatial scale and the highest photo resolution are limited. In this work, the actual photo area is limited to be 2.81 $m^2$ (minimum 0.03 and mean 0.26 $m^2$; see details in data package (Y. Chen et al., 2023)). Such a limitation is mainly caused by how high a user can hold a camera. Also, the highest photo resolution is 0.05 mm/pixel and the minimum detectable grain size by YOLO is 0.45 mm. This means that sediments smaller than medium (0.25 – 0.5 mm) or coarse (0.5 – 1 mm) sands may not be reliably detected. Due to these lim-

itations, a much large number of photos are required in order to fully characterize the stream grain sizes and HBGC at watershed scales. The second limitation is the high labor costs required to prepare the training data. Due to the diversity of natural streams, a large number of labels with high quality are needed for reliable prediction of grain sizes (see effects of insufficient training data in Section 4.3). In this work, we spent around 200 hours to label around 17,000 grains to represent most of the stream conditions. Despite such effort, the trained AI still has 20% – 25% relative error for 2 photos (Figure 4d; Figure 9e; Figure 12(f,r)). More data and improved YOLO algorithms may be needed to better capture very large grains at the boundary of the photos.

Additionally, there are limitations in the empirical formulas for HBGC estimations. Due to the low uncertainty and good agreement with calibrated values (Sections 4.1 – 4.2), the Equations 1 and 6 are likely reliable for estimating Manning coefficient and its uncertainty. For friction factor, though it demonstrates large variations and uncertainty (Sections 3.3 – 3.5), the accuracy of Equation 2 has been comprehensively studied and was recognized as the second best formula for resistance estimation with depth and grain size as inputs (Powell, 2014). The Equation 3 for estimating streambed interstitial velocity magnitude is derived from 17 high-resolution CFD simulations driven by structure-from-motion reconstructed streambeds (Y. Chen et al., 2019, 2021). Though it successfully estimates the most likely magnitude of interstitial velocity (Section 4.1), further simulations or experiments with more streambed conditions may be needed to further evaluate its applicability for diverse streambed conditions, especially the relationship between subsurface permeability and the 5th percentile grain size distribution. For uptake velocity, the hydrogeology-biochemistry interaction efficiency term ($\phi$ in Equation 4) is fitted based on field measured data and thus its applicability in diverse streambed conditions also requires further evaluation.

### 4.7 Future directions

As discussed in Section 4.6, the scale and resolution are limited by hand-held approaches. A natural solution is to replace hand-held devices with drones. By using drones it is possible to increase the number of photos and videos with much higher temporal resolution (e.g., 4K and 5.4K videos) and also increase spatial scales. This is primarily due to their high speed (e.g., Skydio 2 and DJI could fly upto 15 – 27 m/s). With available high-resolution streambed data from drones and hand-held devices, an important

future direction is to directly integrate photo-derived high-resolution streambed data with pore-resolved surface-subsurface coupled models and use the simulated pressure, exchange velocity, and turbulence data to improve the empirical formulas for HBGC estimations. With both the improved formulas and high-resolution data, a further step is to integrate the photo-derived streambed grain sizes and HBGC parameters into watershed-scale models aimed at predicting hydro-biogeochemical dynamics.

## 5 Conclusions

This work presents a workflow to extract the quantities, distributions, and uncertainties of streambed grain sizes and hydro-biogeochemistry from photos using YOLO and empirical formulas. The YOLO, an object detection AI model, is firstly trained with 11,977 grain labels from 36 photos representing 9 stream environments, and demonstrates an accuracy of 0.98, 0.98, and 6.65% in terms of the coefficient of determination, the Nash–Sutcliffe efficiency, and mean absolute relative error in predicting the median grain size D50. The model is then used to predict the grain size distributions (GSDs) for 1,999 photos collected at 66 sites in the Yakima River Basin. Three characteristic grain sizes, including the 5th, 50th, and 84th percentiles of GSDs, are subsequently calculated and used to estimate key hydro-biogeochemical parameters, including Manning coefficient, Darcy-Weisbach friction factor, interstitial velocity magnitude, and nitrate uptake velocity.

From the data, the characteristic grain sizes, Manning coefficient, friction factor, interstitial velocity magnitude, and uptake velocity are found to follow log-normal, normal, positively skewed, near log-normal, and negatively skewed distributions, respectively. Their most likely values, i.e., the mode of the distributions, are 6.63 cm (for D50), 0.0339 $s{\cdot}m^{-1/3}$, 0.18, 0.07 m/day, and 1.2 m/day, respectively. And their average uncertainty or variability are reported as 7.33% (for D50), 1.85%, 15.65%, 24.06%, and 13.88%, respectively. The major sources of uncertainties in grain sizes and hydro-biogeochemical parameters are also identified. Specifically, the accuracy of YOLO is the main factor controlling grain size uncertainty. Both YOLO accuracy and stream depth control friction factor uncertainty. The interstitial velocity magnitude uncertainty is determined by both velocity uncertainty and YOLO accuracy. For the uptake velocity uncertainty, it is controlled by uncertainties in velocity, depth, and YOLO accuracy in shallow streams, while controlled by velocity and nitrate concentration uncertainties in deep rivers.

Further analyses of the effects of training data size on YOLO accuracy show that training data with an insufficient number of photos and stream environment types can cause considerable errors in extracting grain size distributions and the statistics of characteristic grain sizes. The selection of a proper class probability threshold is important for avoiding missing or incorrectly selecting individual grains as desired. The photo resolution analyses demonstrate that the integration of circular caps with an AI model can provide an automated scaling approach better than the manual approach in terms of the accuracy and resolution. We also identified the limitations in photo resolution and spatial scale using hand-held cameras, the high labor costs in training data preparation, and the necessity to further improve the empirical formulas for hydro-biogeochemistry estimations. These limitations may be addressed in future research by integrating drone-derived high-resolution streambed data with pore-scale models, and incorporating photo-derived grain sizes and hydro-biogeochemistry parameters to watershed-scale models.

**Acknowledgments**

This research was supported by the United States Department of Energy (DOE) Office of Biological and Environmental Research (BER), Environmental System Science program, through the PNNL River Corridor Science Focus Area (SFA) project (http://www.pnnl.gov/projects/river-corridor). PNNL is operated for the DOE by Battelle Memorial Institute under Contract No. DE-AC05-76RL01830.

## Data availability

All data are available at the ESS-DIVE repository with DOI: 10.15485/1999774 (Y. Chen et al., 2023).

# Appendix

## Derivations of uncertainty propagation equations

As mentioned in Section 2.5, uncertainties occur in YOLO-predicted pixel grain sizes ($D_{xp}$), photo scale measurement ($SC$), and measurements for water depth ($H$), velocity ($U$), and nitrate concentration ($[NO_3^-]$). These uncertainties can further propagate to real grain sizes ($D_x$) and HBGC parameters such as Manning coefficient ($n$), friction factor ($f$), interstitial velocity magnitude ($\sigma_w$), and nitrate uptake velocity ($u_f$). All these uncertainty can be quantified by the ratio of the absolute uncertainty of these quantities to their representative values, for example, manually measured grain sizes and scales, spatial and/or temporal average of depth and velocity, and direct measurement of nitrate concentrations. If denoting the input parameters and subsequently derived grain sizes/HBGC parameters as $x_i$ (i = 1,2,...) and $y_j$ (j = 1,2,...), then the absolute uncertainty can be quantified by $\delta x_i$ and $\delta y_j$ and the relative uncertainty can be calculated as $r_{x_i} = |\delta x_i|/x_i$ and $r_{y_j} = |\delta y_j|/y_j$, respectively. Statistically, such relative uncertainty can be mean absolute relative error (MARE), root-mean-square of the relative error (RMSRE), and the standard deviation of the relative error (STDRE). Here we choose MARE as the reporting metric, however, it can be easily replaced by RMSRE and STDRE. In general, the target $y_j$ is a function of the input parameters $x_i$, which has the form of $y_j = F_j(x_1, ..., x_i, ..., x_n)$. Based on the multi-variable chain rule and the error propagation law (Ku, 1966), the uncertainty of $y_j$ can be computed through Equation 12.

$$r_{y_j}^2 = \frac{(\delta y_j)^2}{y_j^2} = y_j^{-2}\Big[\sum_{i=1}^{n}(\frac{\partial F_j}{\partial x_i})^2(\delta x_i)^2 + \sum_{i=1}^{n}\sum_{k=1,k\neq i}^{n}\frac{\partial F_j}{\partial x_i}\frac{\partial F_j}{\partial x_k}\delta x_i \delta x_k\Big] \tag{12}$$

The last term in Equation 12 represents the correlation among input variable uncertainty and could be assumed as 0 if the uncertainty of input variables are independent to each other. With such an assumption, Equation 12 can be rewritten as Equation 13.

$$r_{y_j} = \sqrt{y_j^{-2}\sum_{i=1}^{n}(\frac{\partial F_j}{\partial x_i})^2(\delta x_i)^2} = \sqrt{\sum_{i=1}^{n}(\frac{\partial F_j}{\partial x_i})^2\frac{(\delta x_i)^2}{x_i^2}\frac{x_i^2}{y_j^2}} = \sqrt{\sum_{i=1}^{n}\Big(\frac{\partial F_j}{\partial x_i}\frac{x_i}{y_j}\Big)^2 r_{x_i}^2} = \sqrt{\sum_{i=1}^{n} s_{x_i}^2 r_{x_i}^2} \tag{13}$$

where $\frac{\partial F_j}{\partial x_i}\frac{x_i}{y_j}$ is the uncertainty propagation scale of $y_j$ to input $x_i$, and is denoted by $s_{x_i}$ for convenience. With such a general form of uncertainty propagation equation, we apply it for real grain size $D_x$ and the four HBGC parameters in Equations 1 – 4.

For $D_x$, it depends on two independent variables $D_{xp}$ and $SC$. Its uncertainty propagation scales are both 1 for $D_{xp}$ and $SC$, which results in Equation 5. For Manning co-

efficient, it depends on only one variable and its propagation scale is 1/6. For friction factor, if denoting $H/D_{84}$ by $H_{D84}$, then Equation 2 becomes a single variable function of $H_{D84}$. Its uncertainty can be calculated by $r_f = |s_{H_{D84}}| r_{H_{D84}}$ with $|s_{H_{D84}}|$ represented by Equation 14.

$$|s_{H_{D84}}| = |\frac{\partial f}{\partial H_{D84}} \frac{H_{D84}}{f}| = \frac{6c_1^2 + c_2^2 H_{D84}^{5/3}}{3c_1^2 + 3c_2^2 H_{D84}^{5/3}} = 2 - \frac{5}{3} \frac{c_2^2 H_{D84}^{5/3}}{c_1^2 + c_2^2 H_{D84}^{5/3}} = 2 - \frac{5}{3} \frac{1}{c_1^2/c_2^2 H_{D84}^{-5/3} + 1}$$
$$= 2\Big\{1 - \frac{5}{6}\big[1 + \frac{c_1^2}{c_2^2} H_{D84}^{-5/3}\big]^{-1}\Big\} = 2\Big\{1 - \frac{5}{6}\big[1 + \frac{c_1^2}{c_2^2}(\frac{H}{D_{84}})^{-5/3}\big]^{-1}\Big\} \tag{14}$$

For the uncertainty term $r_{H_{D84}}$, because $H_{D84} = H/D_{84}$, its uncertainty propagation scales for $H$ and $D_{84}$ are both 1, therefore, $r_{H_{D84}} = \sqrt{r_H^2 + r_{84}^2}$. Such an equation together with Equation 14 leads to Equation 7.

For interstitial velocity magnitude (Equation 3), both $D_5$ and $D_{50}$ are used as inputs. However, these two variables are not independent. To avoid using both sizes as inputs, we use a simplified $D_5$ relationship, $D_5 = 0.23 D_{50}$ (fitted from data; see Section 3.2 and Figure 5d), to replace the YOLO-derived $D_5$ for uncertainty quantification purpose. With such an simplification, Equation 3 is converted to Equation 15.

$$\sigma_w = \frac{0.23^2 c_3 c_5}{2\nu} U^2 D_{50}^{1-c_4} H^{c_4} \tag{15}$$

The uncertainty propagation scales of Equation 15 with respect to inputs $U$, $D_{50}$, and $H$ were computed as 2, (1-$c_4$), and $c_4$, respectively. Combining these scales and the uncertainty of input parameters will lead to Equation 8.

For nitrate uptake velocity, we rewrite Equation 4 in the form of Equation 16 to utilizing the uncertainty equation for friction factor. If we assume no correlation among the three inputs, then the uncertainty propagation scales of $u_f$ with respect to $U$, $f$, and $[NO_3^-]$ are 1, 1/2, and $c_7$, respectively. Combining these scales and the uncertainty of input parameters leads to Equation 9.

$$u_f = \frac{0.17 Sc^{-2/3} c_6}{\sqrt{8}} U f^{1/2} [NO_3^-]^{c_7} \tag{16}$$

901 **Grain size distribution of 20 test photos**

## References


Abeshu, G. W., Li, H. Y., Zhu, Z., Tan, Z., Leung, L. R., & Li, H. Y. (2022). Median bed-material sediment particle size across rivers in the contiguous US. *Earth System Science Data*, *14*(2), 929–942. doi: 10.5194/essd-14-929-2022

Adams, J. (1979). Gravel Size Analysis from Photographs. *Journal of the Hydraulics Division*, *105*(10), 1247–1255. doi: 10.1061/JYCEAJ.0005283

Anastasi, G. (1984). *Geschiebeanalysen im felde unter beruecksichtigung Von grobkomponenten [Field grain-size analyses with special attention to coarse material]* (Tech. Rep. No. 70).

Black, M., Carbonneau, P., Church, M., & Warburton, J. (2014). Mapping sub-pixel fluvial grain sizes with hyperspatial imagery. *Sedimentology*, *61*, 691–711. doi: 10.1111/sed.12072

Bunte, K., & Abt, S. R. (2001). *Sampling surface and subsurface particle-size distributions in wadable gravel- and cobble-bed streams for analyses in sediment transport, hydraulics, and streambed monitoring* (Tech. Rep.). Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station.

Buscombe, D. (2013). Transferable wavelet method for grain-size distribution from images of sediment surfaces and thin sections, and other natural granular patterns. *Sedimentology*, *60*, 1709–1732. doi: 10.1111/sed.12049

Buscombe, D., & Masselink, G. (2009). Grain-size information from the statistical properties of digital images of sediment. *Sedimentology*, *56*, 421–438. doi: 10.1111/j.1365-3091.2008.00977.x

Buscombe, D., & Rubin, D. M. (2012). Advances in the simulation and automated measurement of well-sorted granular material: 2. Direct measures of particle properties. *Journal of Geophysical Research: Earth Surface*, *117*, F02002. doi: 10.1029/2011JF001975

Buscombe, D., Rubin, D. M., & Warrick, J. A. (2010). A universal approximation of grain size from images of noncohesive sediment. *Journal of Geophysical Research: Earth Surface*, *115*, F02015. doi: 10.1029/2009jf001477

Butler, J. B., Lane, S. N., & Chandler, J. H. (2001). Automated extraction of grain-size data from gravel surfaces using digital image processing. *Journal of Hydraulic Research*, *39*(5), 519–529. doi: 10.1080/00221686.2001.9628276

Carbonneau, P. E., Bergeron, N., & Lane, S. N. (2005a). Automated grain size measurements from airborne remote sensing for long profile measurements of fluvial grain sizes. *Water Resources Research*, *41*, W11426. doi: 10.1029/2005WR003994

Carbonneau, P. E., Bergeron, N. E., & Lane, S. N. (2005b). Texture-based image segmentation applied to the quantification of superficial sand in salmonid river gravels. *Earth Surface Processes and Landforms*, *30*, 121–127. doi: 10.1002/esp.1140

Carbonneau, P. E., Lane, S. N., & Bergeron, N. E. (2004). Catchment-scale mapping of surface grain size in gravel bed rivers using airborne digital imagery. *Water Resources Research*, *40*, W07202. doi: 10.1029/2003WR002759

Chaudhry, M. H. H. (2008). *Open-channel flow* (2nd ed.). New York, NY, USA: Springer US. doi: 10.1007/978-0-387-68648-6

Chen, K., Chen, X., Stegen, J. C., Villa, J. A., Bohrer, G., Song, X., ... Zheng, C. (2023). Vertical hydrologic exchange flows control methane emissions from riverbed sediments. *Environmental Science and Technology*, *57*(9), 4014–4026. doi: 10.1021/acs.est.2c07676

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. *Lecture Notes in Computer Science*, *11211*, 833–851. doi: 10.1007/978-3-030-01234-2_49

Chen, Y., Bao, J., Chen, Y., Yang, Y., Stegen, J., Renteria, L., ... Scheibe, T. (2023). *Artificial intelligence models, photos, and data associated with the manuscript "Quantifying streambed grain sizes and hydro-biogeochemistry using YOLO and photos".* Richland, WA, USA: River Corridor and Watershed Biogeochemistry SFA, ESS-DIVE repository. doi: 10.15485/1999774

Chen, Y., Bao, J., Li, B., Liu, X., DiBiase, R., & Scheibe, T. (2021). Estimating hydrological exchange flows from streambed images. In *Agu fall meeting* (pp. B25D–1480). New Orleans, LA.

Chen, Y., DiBiase, R. A. R. A., McCarroll, N., & Liu, X. (2019). Quantifying flow resistance in mountain streams using computational fluid dynamics modeling over structure-from-motion photogrammetry-derived microtopography. *Earth Surface Processes and Landforms*, *44*(10), 1973–1987. doi: 10.1002/esp.4624

Chen, Z., Scott, T. R., Bearman, S., Anand, H., Keating, D., Scott, C., ... Das, J. (2020). Geomorphological analysis using unpiloted aircraft systems, structure from motion, and deep learning. In *2020 ieee/rsj international conference on intelligent robots and systems (iros)* (pp. 1276–1283). IEEE. doi: 10.1109/IROS45743.2020.9341354

Delgado, D., Barnes, M., Boehnke, B., Chen, X., Cornwell, K., Forbes, B., ... Stegen, J. (2023). *Spatial study 2022: surface water samples, cotton strip degradation, and hydrologic sensor data across the Yakima River Basin, Washington, USA (v2)* (Tech. Rep.). Richland, WA, USA: Pacific Northwest National Laboratory. doi: 10.15485/1969566

Detert, M., & Weitbrecht, V. (2012). Automatic object detection to analyze the geometry of gravel grains - a free stand-alone tool. In *River flow 2012 - proceedings of the international conference on fluvial hydraulics* (pp. 595–600). San Jose, Costa Rica: Taylor and Francis.

Elliott, A. H., & Brooks, N. H. (1997). Transfer of nonsorbing solutes to a streambed with bed forms: Theory. *Water Resources Research*, *33*(1), 123–136. doi: 10.1029/96WR02783

Ermilov, A. A., Benkő, G., & Baranya, S. (2022). Automated riverbed material analysis using Deep Learning on underwater images. *Earth Surf. Dynam. Discuss*(November), 1–34. doi: 10.5194/esurf-2022-56

Fehr, R. (1987). Geschiebeanalysen in Gebirgsflüssen [Bed material analyses in mountain streams. Transformation and comparison of various analyses procedures]. *Mitteilungen der Versuchsanstalt für Wasserbau, Hydrologie und Glaziologie*(92), 139.

Ferguson, R. (2007). Flow resistance equations for gravel- and boulder-bed streams. *Water Resources Research*, *43*, W05427. doi: 10.1029/2006WR005422

Ferguson, R. (2010). Time to abandon the Manning equation? *Earth Surface Processes and Landforms*, *35*(15), 1873–1876. doi: 10.1002/esp.2091

Ferguson, R. (2022). Reach-scale flow resistance. *Treatise on Geomorphology*, *6.1*, 110–132. doi: 10.1016/B978-0-12-409548-9.09386-6

Fripp, J. B., & Diplas, P. (1993, apr). Surface sampling in gravel streams. *Journal of Hydraulic Engineering*, *119*, 473–490. doi: 10.1061/(ASCE)0733-9429(1993)119:4(473)

Gomez-Velez, J. D., Harvey, J. W., Cardenas, M. B., & Kiel, B. (2015). Denitrification in the Mississippi River network controlled by flow through river bedforms. *Nature Geoscience*, *8*(12), 941–945. doi: 10.1038/ngeo2567

Graham, D. J., Reid, I., & Rice, S. P. (2005, jan). Automated sizing of coarse-grained sediments: image-processing procedures. *Mathematical Geology*, *37*(1), 1–28. doi: 10.1007/s11004-005-8745-x

Grant, S. B., Azizian, M., Cook, P., Boano, F., & Rippy, M. A. (2018, mar). Factoring stream turbulence into global assessments of nitrogen pollution. *Science*, *359*(6381), 1266–1269. doi: 10.1126/science.aap8074

Grieger, S., Barnes, M., Borton, M., Chen, X., Chu, R., Farris, Y., ... Stegen, J. (2022). *Spatial study 2021: sample-based surface water chemistry and organic matter characterization across watersheds in the Yakima River Basin, Washington, USA (v2)* (Tech. Rep.). Richland, WA, USA: Pacific Northwest National Laboratory. doi: 10.15485/1898914

He, K., Gkioxari, G., Dollar, P., & Girshick, R. (2017, oct). Mask R-CNN. In *2017 ieee international conference on computer vision (iccv)* (pp. 2980–2988). IEEE. doi: 10.1109/ICCV.2017.322

He, K., Zhang, X., Ren, S., & Sun, J. (2014). Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. In *European conference on computer vision* (pp. 346–361). Retrieved from `http://link.springer.com/10.1007/978-3-319-10578-9_23` doi: 10.1007/978-3-319-10578-9_23

Ibbeken, H., & Schleyer, R. (1986). Photo-sieving: A method for grain-size analysis of coarse-grained, unconsolidated bedding surfaces. *Earth Surface Processes and Landforms*, *11*(1), 59–77. doi: 10.1002/esp.3290110108

Jiang, Y., Frankenberger, J. R., Bowling, L. C., & Sun, Z. (2014). Quantification of uncertainty in estimated nitrate-N loads in agricultural watersheds. *Journal of Hydrology*, *519*(PA), 106–116. doi: 10.1016/j.jhydrol.2014.06.027

Kaufman, M., Delgado, D., Barnes, M., Boehnke, B., Chen, X., Cornwell, K., ... Stegen, J. (2023). *Spatial study 2022: water column, sediment, and total ecosystem respiration rates across the Yakima River Basin, Washington, USA* (Tech. Rep.). Richland, WA, USA: Pacific Northwest National Laboratory. doi: 10.15485/1987520

Kellerhals, R., & Bray, D. I. (1971, aug). Sampling procedures for coarse fluvial sed-

iments. *Journal of the Hydraulics Division*, *97*(8), 1165–1180. doi: 10.1061/JYCEAJ.0003044

Kenney, T. C., Lau, D., & Ofoegbu, G. I. (1984, nov). Permeability of compacted granular materials. *Canadian Geotechnical Journal*, *21*, 726–729. doi: 10.1139/t84-080

Ku, H. H. (1966). Notes on the use of propagation of error formulas. *Journal of Research of the National Bureau of Standards, Section C: Engineering and Instrumentation*, *70C*(4), 263–273. doi: 10.6028/jres.070C.025

Lang, N., Irniger, A., Rozniak, A., Hunziker, R., Wegner, J. D., & Schindler, K. (2021, may). GRAINet: mapping grain size distributions in river beds from UAV images with convolutional neural networks. *Hydrology and Earth System Sciences*, *25*, 2567–2597. doi: 10.5194/hess-25-2567-2021

Lang, S., Ladson, T., & Anderson, B. (2004). A review of empirical equations for estimating stream roughness and their application to four streams in Victoria. *Australasian Journal of Water Resources*, *8*(1), 69–82. doi: 10.1080/13241583.2004.11465245

Leopold, L. B. (1970). An improved method for size distribution of stream bed gravel. *Water Resources Research*, *6*(5), 1357–1366. doi: 10.1029/WR006i005p01357

Mueller, D. S., Abad, J. D., García, C. M., Gartner, J. W., García, M. H., & Oberg, K. A. (2007). Errors in acoustic doppler profiler velocity measurements caused by flow disturbance. *Journal of Hydraulic Engineering*, *133*(12), 1411–1420. doi: 10.1061/(asce)0733-9429(2007)133:12(1411)

Mulholland, P. J., Hall, R. O., Sobota, D. J., Dodds, W. K., Findlay, S. E. G., Grimm, N. B., ... Thomasn, S. M. (2009, may). Nitrate removal in stream ecosystems measured by 15N addition experiments: Denitrification. *Limnology and Oceanography*, *54*(3), 666–680. doi: 10.4319/lo.2009.54.3.0666

Mulholland, P. J., Helton, A. M., Poole, G. C., Hall, R. O., Hamilton, S. K., Peterson, B. J., ... Thomas, S. M. (2008). Stream denitrification across biomes and its response to anthropogenic nitrate loading. *Nature*, *452*(7184), 202–205. doi: 10.1038/nature06686

Niehus, S., Perkins, W., & Richmond, M. (2014). *Simulation of Columbia River Hydrodynamics and Water Temperature from 1917 through 2011 in the Han-*

*ford Reach* (Tech. Rep. No. PNWD-3278). Richland, WA: Battelle-Pacific Northwest Division. doi: 10.13140/RG.2.1.5146.8409

NOAA. (2023). *The National Water Model.* Retrieved from `https://water.noaa.gov/about/nwm`

O'Connor, B. L., & Hondzo, M. (2008). Dissolved oxygen transfer to sediments by sweep and eject motions in aquatic environments. *Limnology and Oceanography*, *53*(2), 566–578. doi: 10.4319/lo.2008.53.2.0566

Picioreanu, C., van Loosdrecht, M. C. M., & Heijnen, J. J. (1997, jul). Modelling the effect of oxygen concentration on nitrite accumulation in a biofilm airlift suspension reactor. *Water Science and Technology*, *36*(1), 147–156. doi: 10.2166/wst.1997.0034

Powell, D. M. (2014). Flow resistance in gravel-bed rivers: progress in research. *Earth-Science Reviews*, *136*, 301–338. doi: 10.1016/j.earscirev.2014.06.001

Purinton, B., & Bookhagen, B. (2019). Introducing PebbleCounts: a grain-sizing tool for photo surveys of dynamic gravel-bed rivers. *Earth Surface Dynamics*, *7*, 859–877. doi: 10.5194/esurf-7-859-2019

PyTorch. (2022). *PyTorch.* Retrieved from `https://pytorch.org/`

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: unified, real-time object detection. In *2016 ieee conference on computer vision and pattern recognition (cvpr)* (pp. 779–788). IEEE. doi: 10.1109/CVPR.2016.91

Ren, H., Hou, Z., Duan, Z., Song, X., Perkins, W. A., Richmond, M. C., ... Scheibe, T. D. (2020). Spatial mapping of riverbed grain-size distribution using machine learning. *Frontiers in Water*, *2*, 551627. doi: 10.3389/frwa.2020.551627

Rickenmann, D., & Recking, A. (2011). Evaluation of flow resistance in gravel-bed rivers through a large field data set. *Water Resources Research*, *47*, W07538. doi: 10.1029/2010WR009793

Rubin, D. M. (2004). A Simple Autocorrelation Algorithm for Determining Grain Size from Digital Images of Sediment. *Journal of Sedimentary Research*, *74*(1), 160–165. doi: 10.1306/052203740160

Sadayappan, K., Kerins, D., Shen, C., & Li, L. (2022, nov). Nitrate concentrations predominantly driven by human, climate, and soil properties in US rivers. *Water Research*, *226*, 119295. doi: 10.1016/j.watres.2022.119295

Shepherd, R. G. (1989). Correlations of permeability and grain Size. *Ground Water*, *27*(5), 633–638. doi: 10.1111/j.1745-6584.1989.tb00476.x

Soloy, A., Turki, I., Fournier, M., Costa, S., Peuziat, B., & Lecoq, N. (2020). A Deep Learning-Based Method for Quantifying and Mapping the Grain Size on Pebble Beaches. *Remote Sensing*, *12*, 3659. doi: 10.3390/rs12213659

Son, K., Fang, Y., Gomez-Velez, J. D., Byun, K., & Chen, X. (2022, dec). Combined effects of stream hydrology and land use on basin-scale hyporheic zone denitrification in the Columbia River Basin. *Water Resources Research*, *58*, e2021WR031131. doi: 10.1029/2021WR031131

Steer, P., Guerit, L., Lague, D., Crave, A., & Gourdon, A. (2022, dec). Size, shape and orientation matter: fast and semi-automatic measurement of grain geometries from 3D point clouds. *Earth Surface Dynamics*, *10*(6), 1211–1232. doi: 10.5194/esurf-10-1211-2022

Strickler, A. (1923). *Contributions to the question of a velocity formula and roughness data for streams, channels and closed pipelines* (Tech. Rep.). Bern, Switzerland: Bureau of Water Affairs, Swiss Department of the Interior.

Ultralytics. (2020). *YOLO V5.* Retrieved from `https://github.com/ultralytics/yolov5`

Vázquez-Tarrío, D., Borgniet, L., Liébault, F., & Recking, A. (2017). Using UAS optical imagery and SfM photogrammetry to characterize the surface grain size of gravel bars in a braided river (Vénéon River, French Alps). *Geomorphology*, *285*, 94–105. doi: 10.1016/j.geomorph.2017.01.039

Verdú, J. M., Batalla, R. J., & Martínez-Casasnovas, J. A. (2005, dec). High-resolution grain-size characterisation of gravel bars using imagery analysis and geo-statistics. *Geomorphology*, *72*, 73–93. doi: 10.1016/j.geomorph.2005.04.015

Wang, C.-Y., Mark Liao, H.-Y., Wu, Y.-H., Chen, P.-Y., Hsieh, J.-W., & Yeh, I.-H. (2020). CSPNet: a new backbone that can enhance learning capability of CNN. In *2020 ieee/cvf conference on computer vision and pattern recognition workshops (cvprw)* (pp. 1571–1580). IEEE. doi: 10.1109/CVPRW50498.2020.00203

Wohl, E. E., Anthony, D. J., Madsen, S. W., & Thompson, D. M. (1996). A comparison of surface sampling methods for coarse fluvial sediments. *Water Resources*

*Research*, *32*(10), 3219–3226. doi: 10.1029/96WR01527

Wolman, M. G. (1954). A method of sampling coarse river-bed material. *Transactions, American Geophysical Union*, *35*(6), 951. doi: 10.1029/TR035i006p00951

Zhang, W., Itoh, K., Tanida, J., & Ichioka, Y. (1990). Parallel distributed processing model with local space-invariant interconnections and its optical architecture. *Applied Optics*, *29*(32), 4790. doi: 10.1364/AO.29.004790

Zhang, X., Zhang, Y., Shi, P., Bi, Z., Shan, Z., & Ren, L. (2021). The deep challenge of nitrate pollution in river water of China. *Science of the Total Environment*, *770*, 144674. doi: 10.1016/j.scitotenv.2020.144674

1143 ## Figure captions

**Figure 1.** The locations, site-average median grain sizes, and labels of photos used for AI training/validation/testing (a), prediction (b), scaling sensitivity and accuracy purposes (c), as well as the number of photos at each site (d). The site locations of group 1 (green circles) are invisible due to too close to group 0 and 2. Their locations are described with a character "V" following the site names in (a).

**Figure 2.** The labels of individual grains (a – i) and scales (j – o) in representative river corridor environments.

**Figure 3.** The sketch of the YOLO version 5 network. Modified from Ultralytics (2020).

**Figure 4.** The convergence history of YOLO training (a) and the accuracy of YOLO predicted grain size distribution (b), median grain size D50 (c) as well as the relative error of D50 prediction (d). NSE in (a) is Nash–Sutcliffe efficiency.

**Figure 5.** The probability density distributions of D50 (a), D5 (b), D84 (c), and the relationship between D5/D84 and D50 (d).

**Figure 6.** The probability density distribution of Manning coefficient (a), Darcy-weisbach friction factor (b), fluctuation magnitude of vertical exchange flux (c), and total nitrate uptake velocity attributed to microbes and turbulence mass transfer (d).

**Figure 7.** The in-site variations (blue cross dots), site average values (horizontal black lines), and estimated relative variations (yellow dot lines) of photo resolution (a), log2-transformed D50 (b), log2-transformed D84 (c), log2-transformed D5 (d), and water depth (e) for 32 sites. The site name is reordered in an alphabetical order for convenience. The nearest region to the right of site name represents the data within the site. The site-average value in (b), (c), and (d) are first averaged over the actual data and then log2-transformed.

**Figure 8.** The in-site variations (blue cross dots), site average values (horizontal black lines), and estimated relative variations (red dot lines) for Manning's coefficient (a), log10-transformed friction factor (b), log10-transformed streambed interstitial velocity magnitude (c), and streambed nitrate uptake velocity (d). The site-average value in (b) and (c) are first averaged over the actual data and then log10-transformed.

**Figure 9.** The effects on training photo number on YOLO precision (a), individual grain size distributions (b,c), median grain size (d) and relative error (e) of testing photos, as well as the prediction of median grain size of prediction photos (f). M0a, M0b, and M0c represent models trained with 11, 21, and 36 photos.

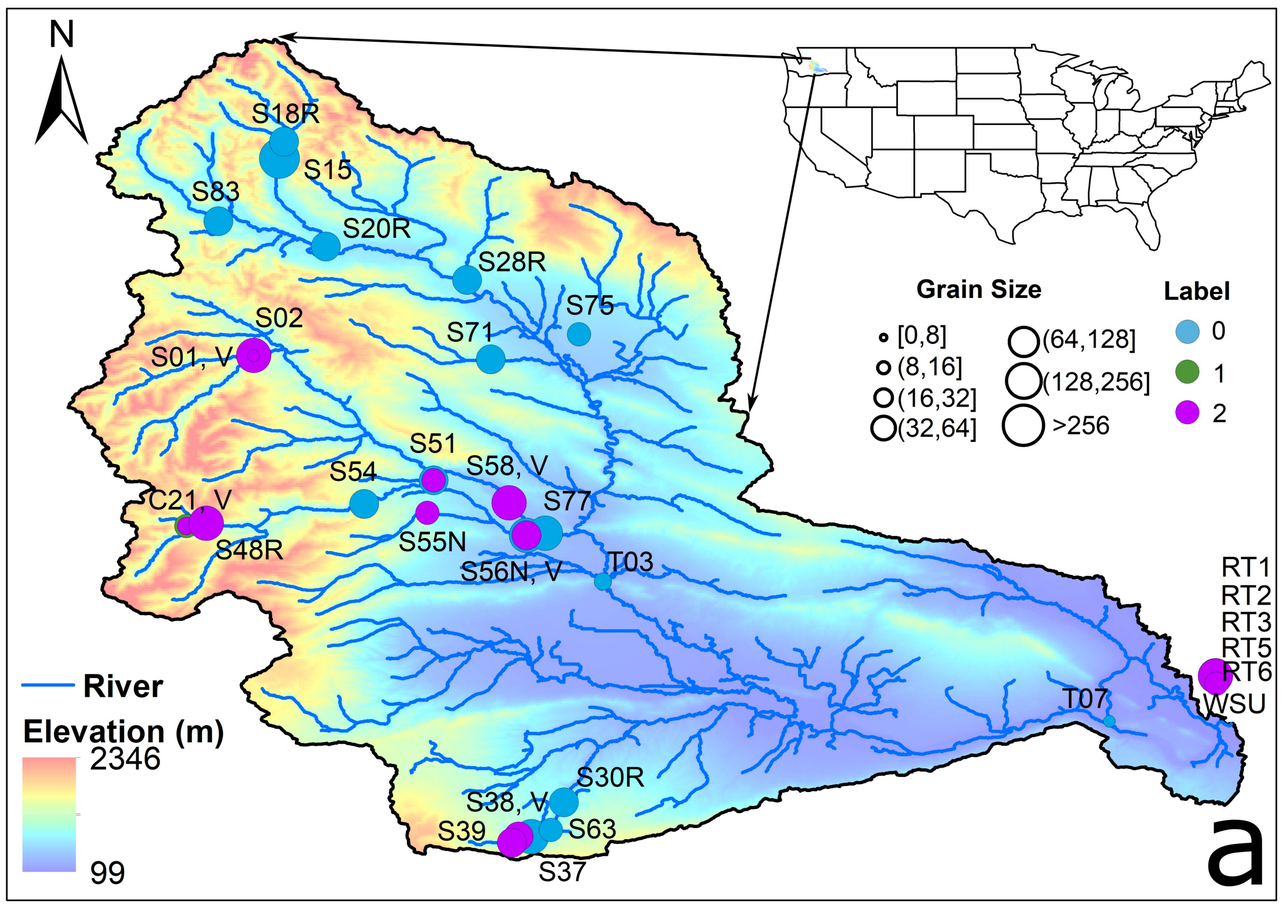
**Figure 10.** The effects of probability threshold on model performance metrics R2 (a), mean and mean absolute error (b), and the average number of grains detected by the model (c.)

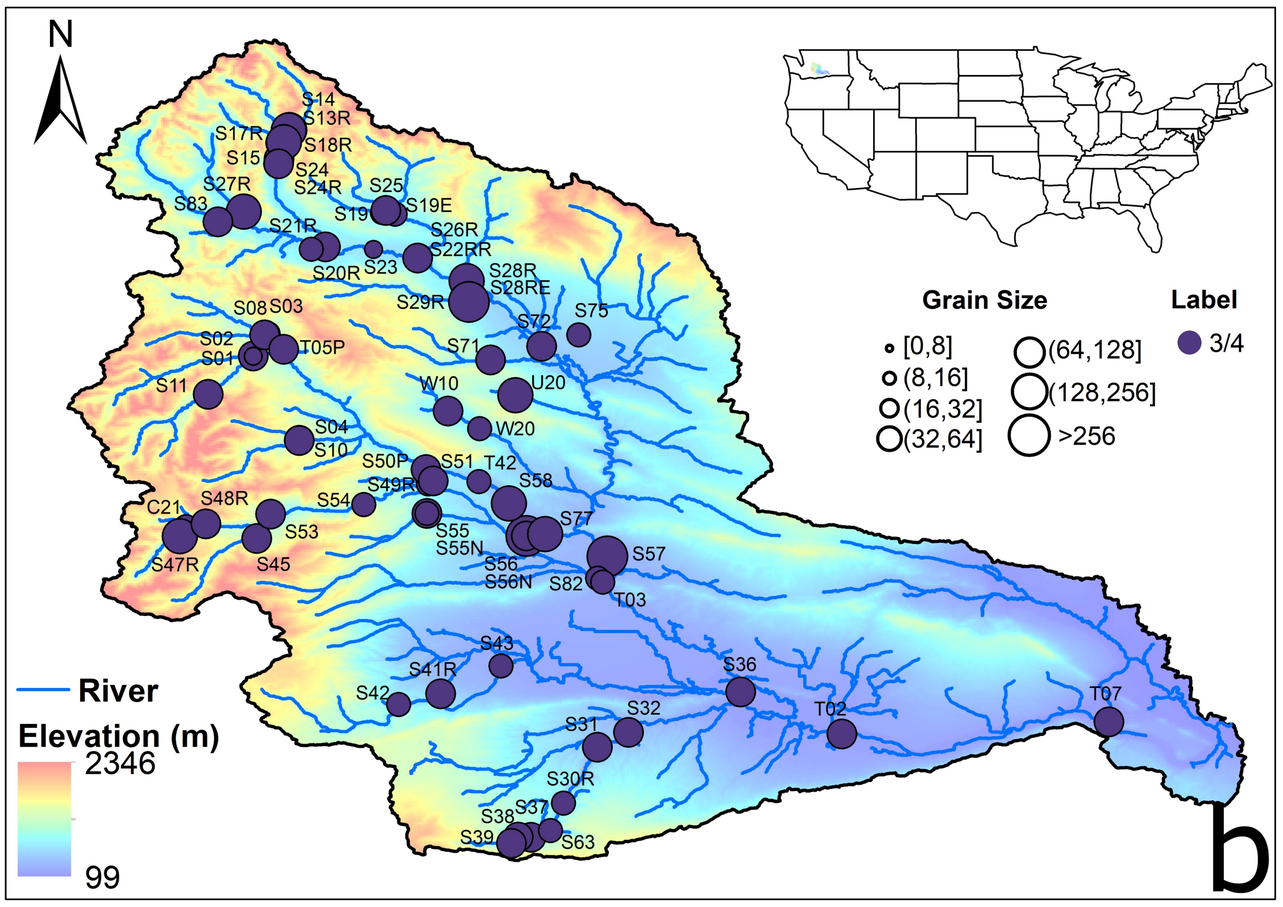
**Figure 11.** The values of photo resolution and associated detected minimum grain sizes using square quadrats and manual measurements of resolution (a), the comparison of automatically predicted photo resolution to the manually measured values using circular caps (b), and the relationship between photo resolution and camera height (c).

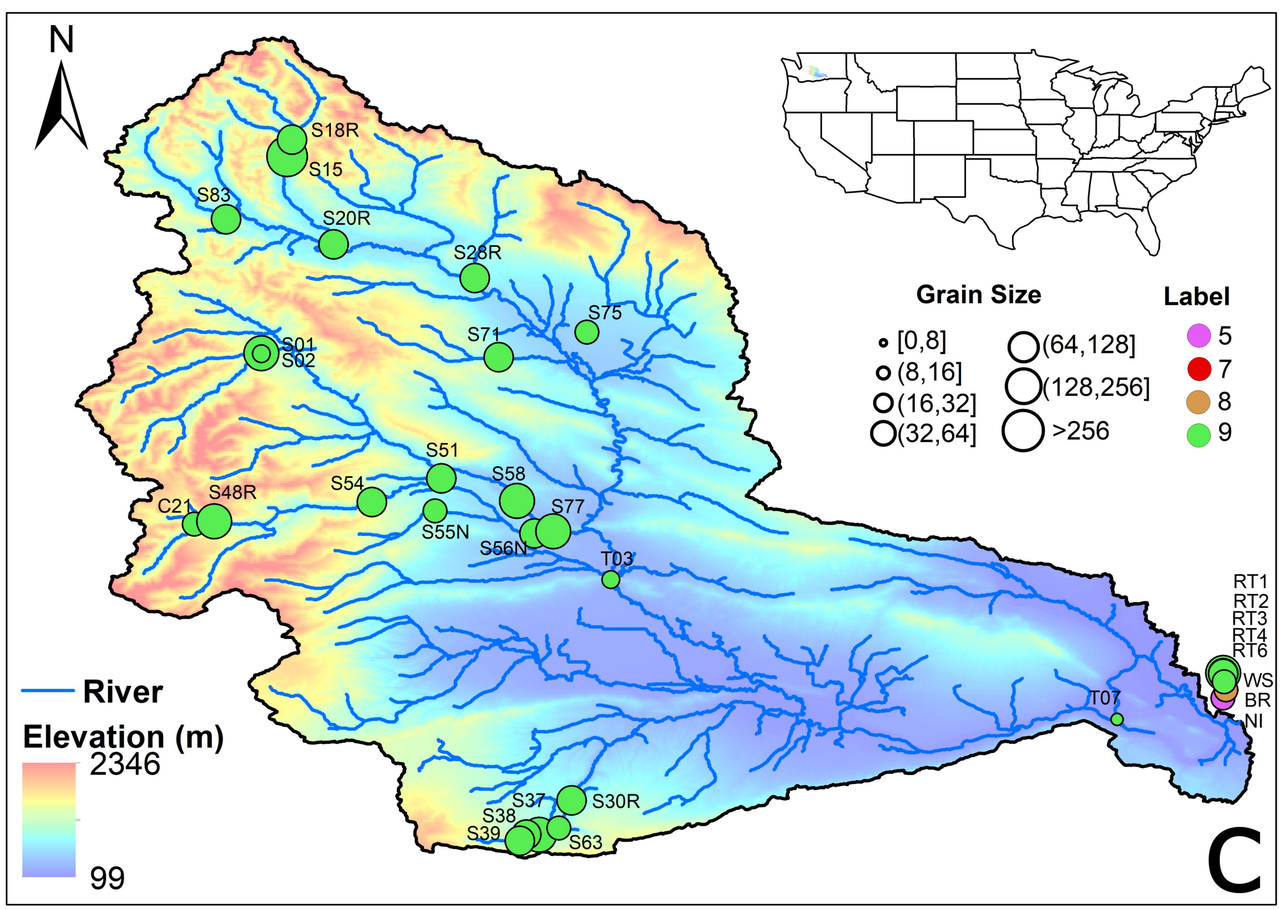
**Figure 12.** The comparison of grain size distribution between YOLO (M0c) prediction and manual measurements for 20 testing photos.

**Figure 13.** The typical scales and YOLO (Msc) predicted scales for the full quadrat (a), 1/4 of the quadart (b), green and blue caps in flowing water (c), and blue cap in dry bed (d).

**Figure 1.**


N
S18R
S15
S83
S20R
S28R
S75
S71
S02
S01, V
S51
S54
S58, V
S77
C21, V
S48R
S55N
S56N, V
T03
T07
S30R
S38, V
S39
S63
S37
RT1
RT2
RT3
RT5
RT6
WSU
River
Elevation (m)
2346
99
Grain Size
[0,8]
(8,16]
(16,32]
(32,64]
(64,128]
(128,256]
>256
Label
0
1
2
a


N
S14
S13R
S17R
S18R
S15
S24
S24R
S27R
S25
S83
S19
S19E
S21R
S26R
S22RR
S20R
S23
S28R
S28RE
S29R
S08
S03
S75
S72
S02
S01
T05P
S71
S11
W10
U20
W20
S04
S10
S50P
S51
T42
S58
S49R
S48R
C21
S54
S53
S55
S77
S55N
S56
S57
S47R
S45
S56N
S82
T03
S43
S41R
S36
S42
S32
T02
T07
S31
S30R
S38
S37
S39
S63
River
Elevation (m)
2346
99
Grain Size
[0,8]
(8,16]
(16,32]
(32,64]
(64,128]
(128,256]
>256
Label
3/4
b


N
S18R
S15
S83
S20R
S28R
S75
S71
S01
S02
S51
S54
S58
S77
C21
S48R
S55N
S56N
T03
T07
S37
S30R
S38
S39
S63
RT1
RT2
RT3
RT4
RT6
WS
BR
NI
River
Elevation (m)
2346
99
Grain Size
[0,8]
(8,16]
(16,32]
(32,64]
(64,128]
(128,256]
>256
Label
5
7
8
9
C

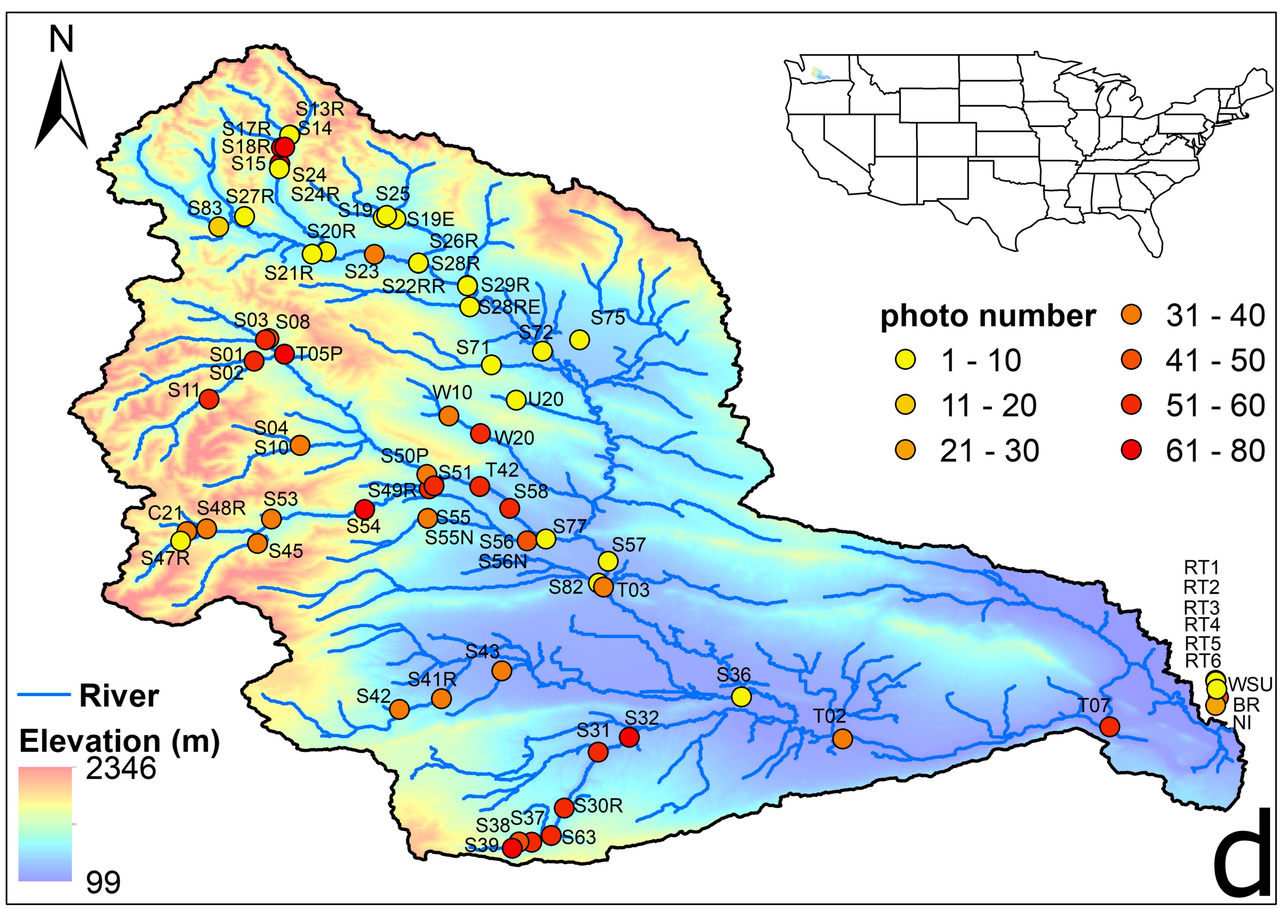
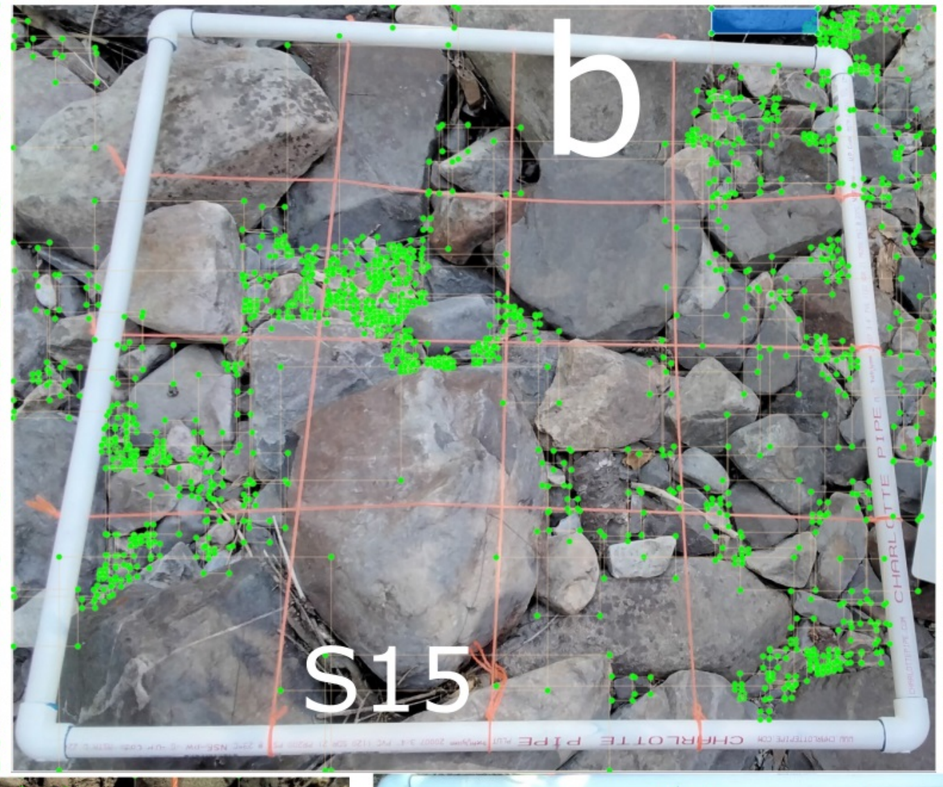
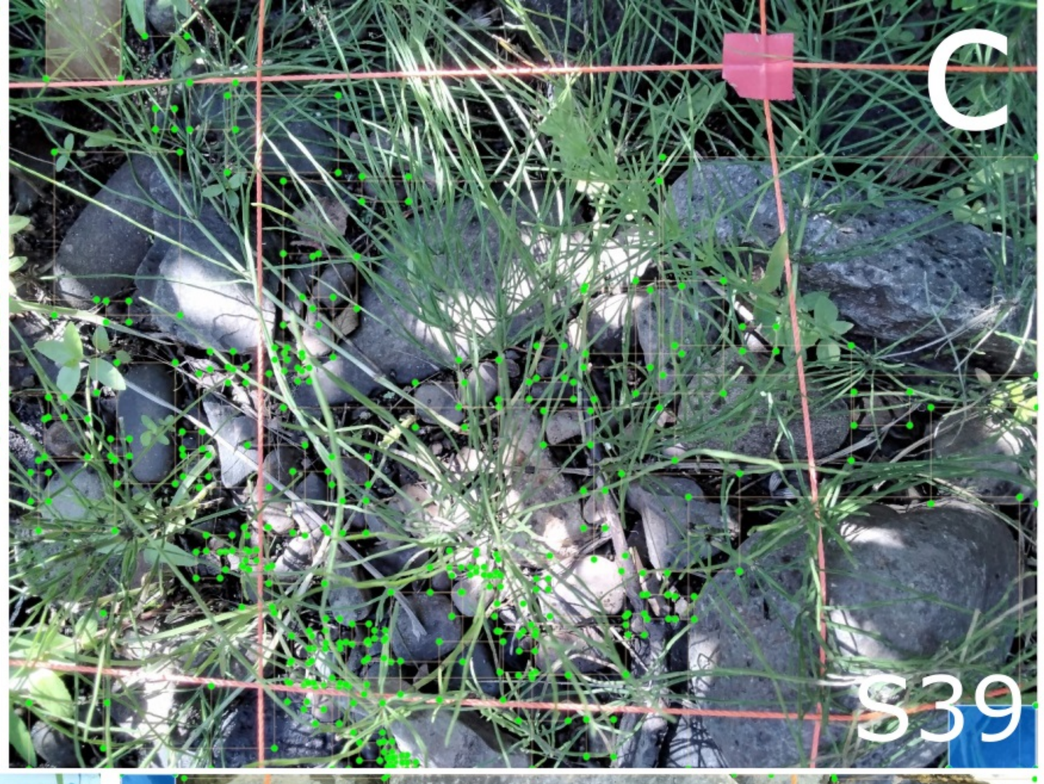
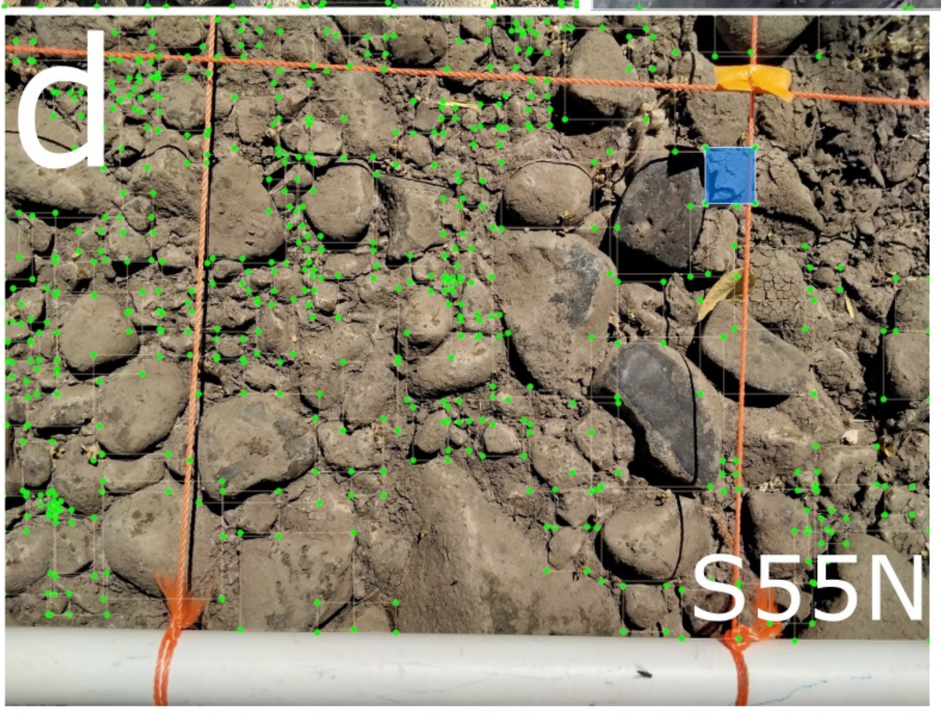

N
S13R
S14
S17R
S18R
S15
S24
S24R
S27R
S25
S83
S19
S19E
S20R
S26R
S21R
S23
S28R
S22RR
S29R
S28RE
S03
S08
S75
S72
S01
T05P
S71
S02
S11
W10
U20
W20
S04
S10
S50P
S51
T42
S58
S49R
S48R
S53
C21
S54
S55
S77
S55N
S56
S57
S47R
S45
S56N
S82
T03
S43
S41R
S36
S42
S32
T02
T07
S31
S30R
S38
S37
S39
S63
RT1
RT2
RT3
RT4
RT5
RT6
WSU
BR
NI
River
Elevation (m)
2346
99
photo number
1 - 10
11 - 20
21 - 30
31 - 40
41 - 50
51 - 60
61 - 80
d

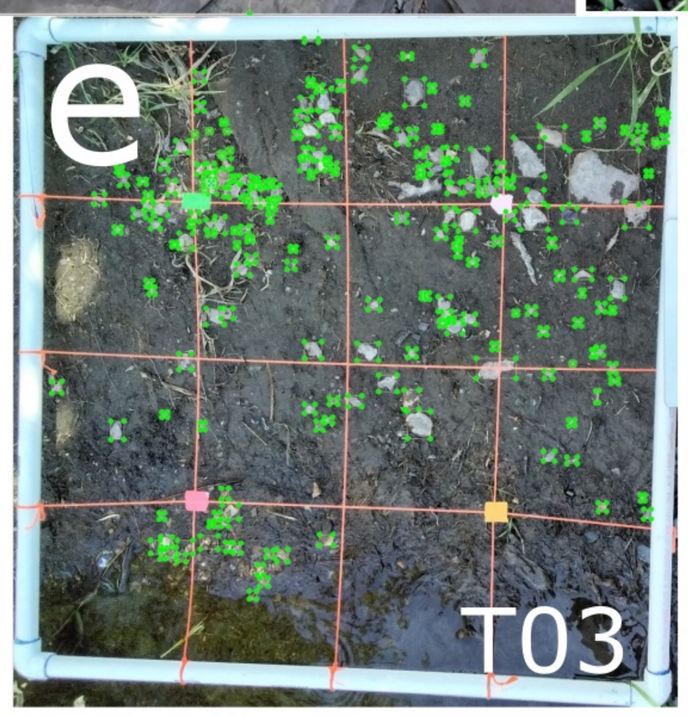
**Figure 2.**


a
RT3

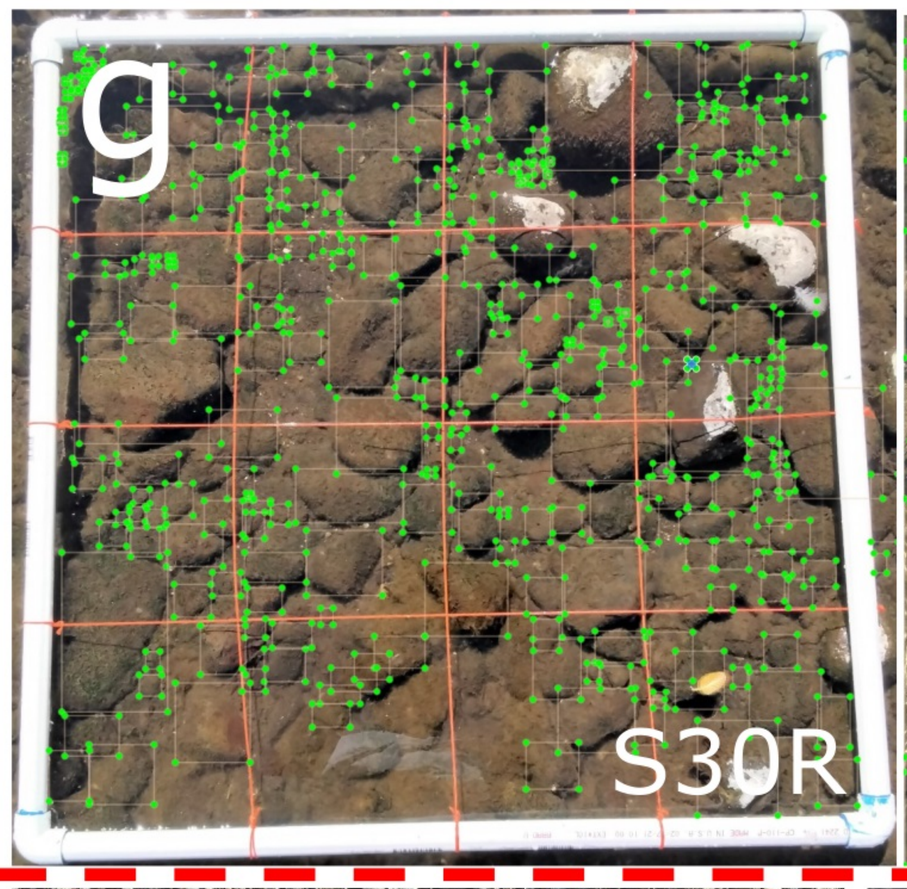

b
S15

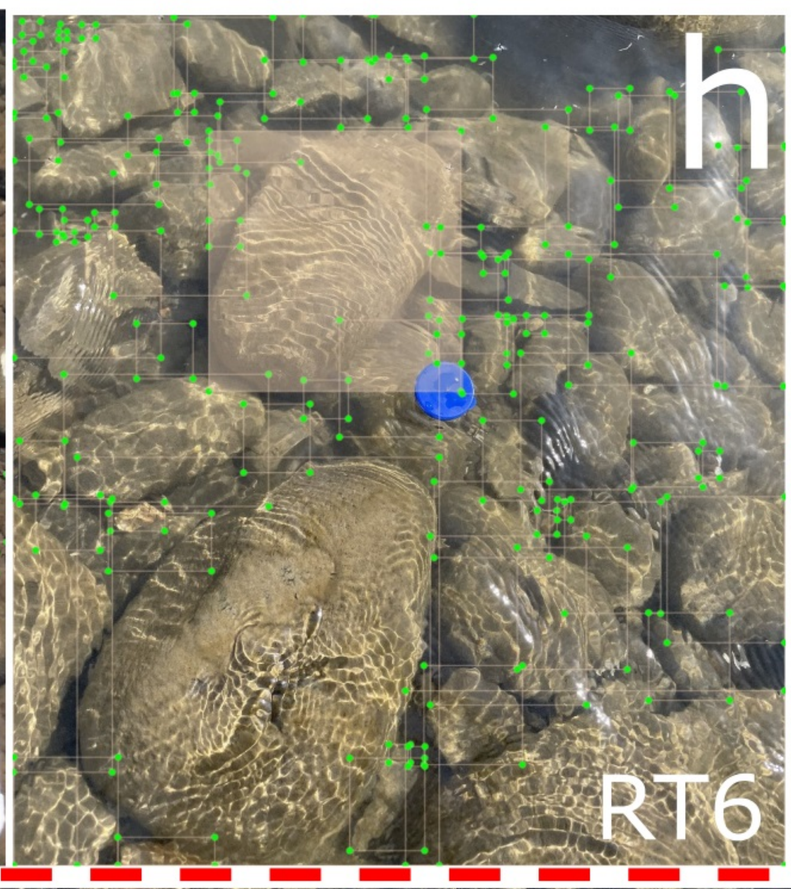

c
S39

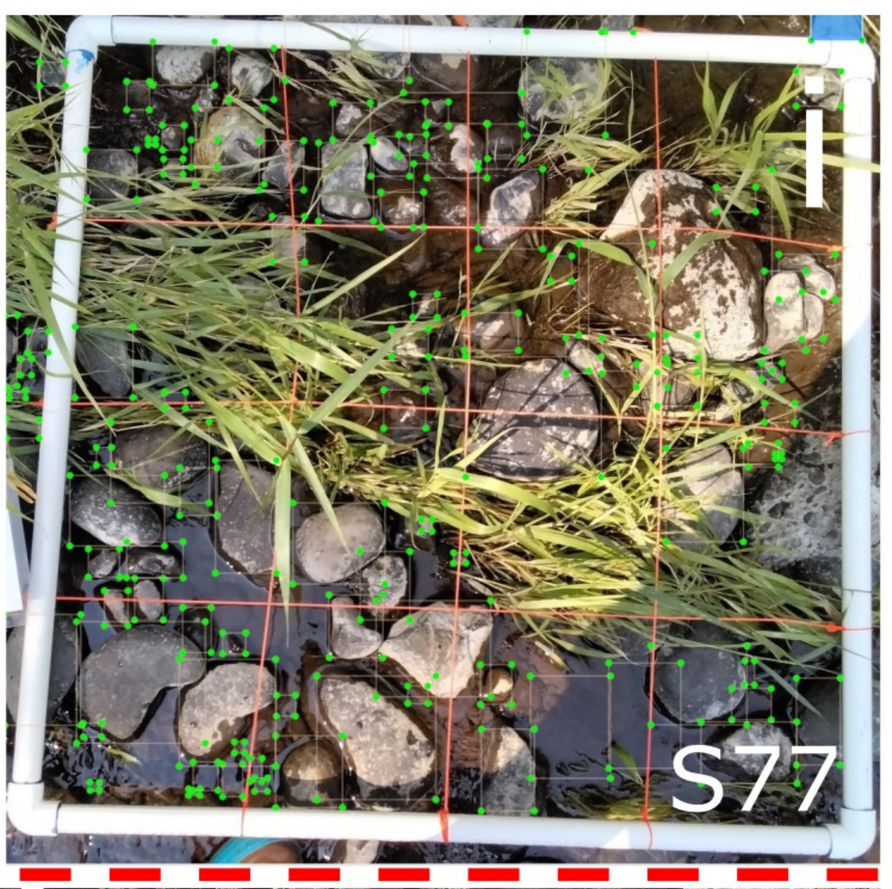

d
S55N


e
T03

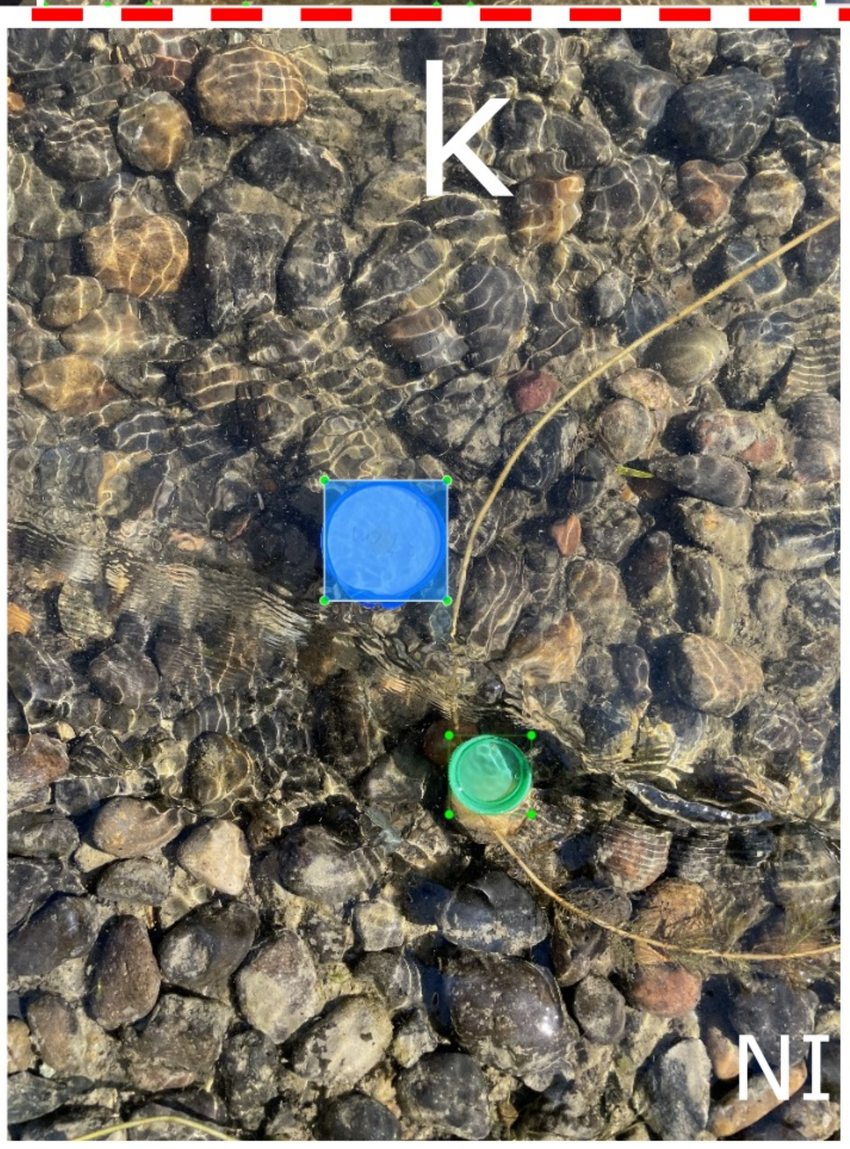

f
S56N


g
S30R


h
RT6


i
S77


j
NI


k
NI


l
NI


m
RT5


n
S56
7/26/22
S56N

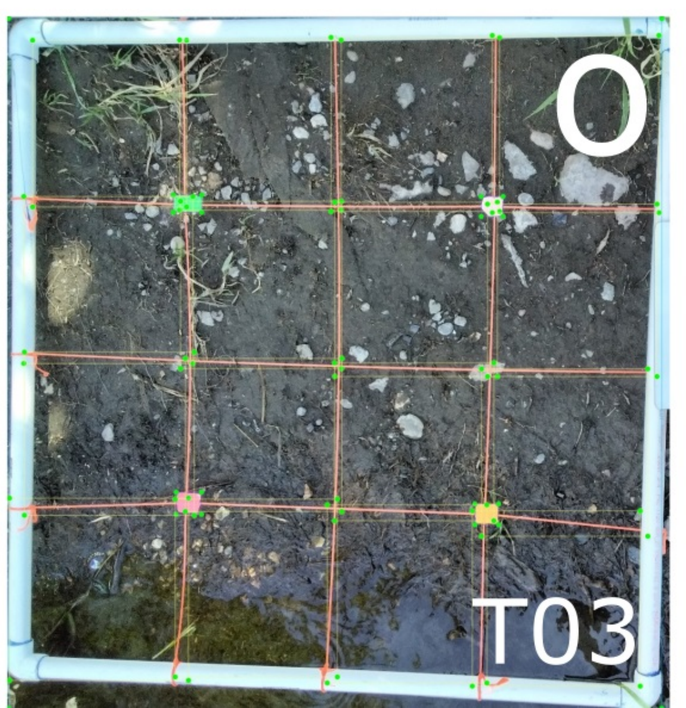

o
T03

**Figure 3.**

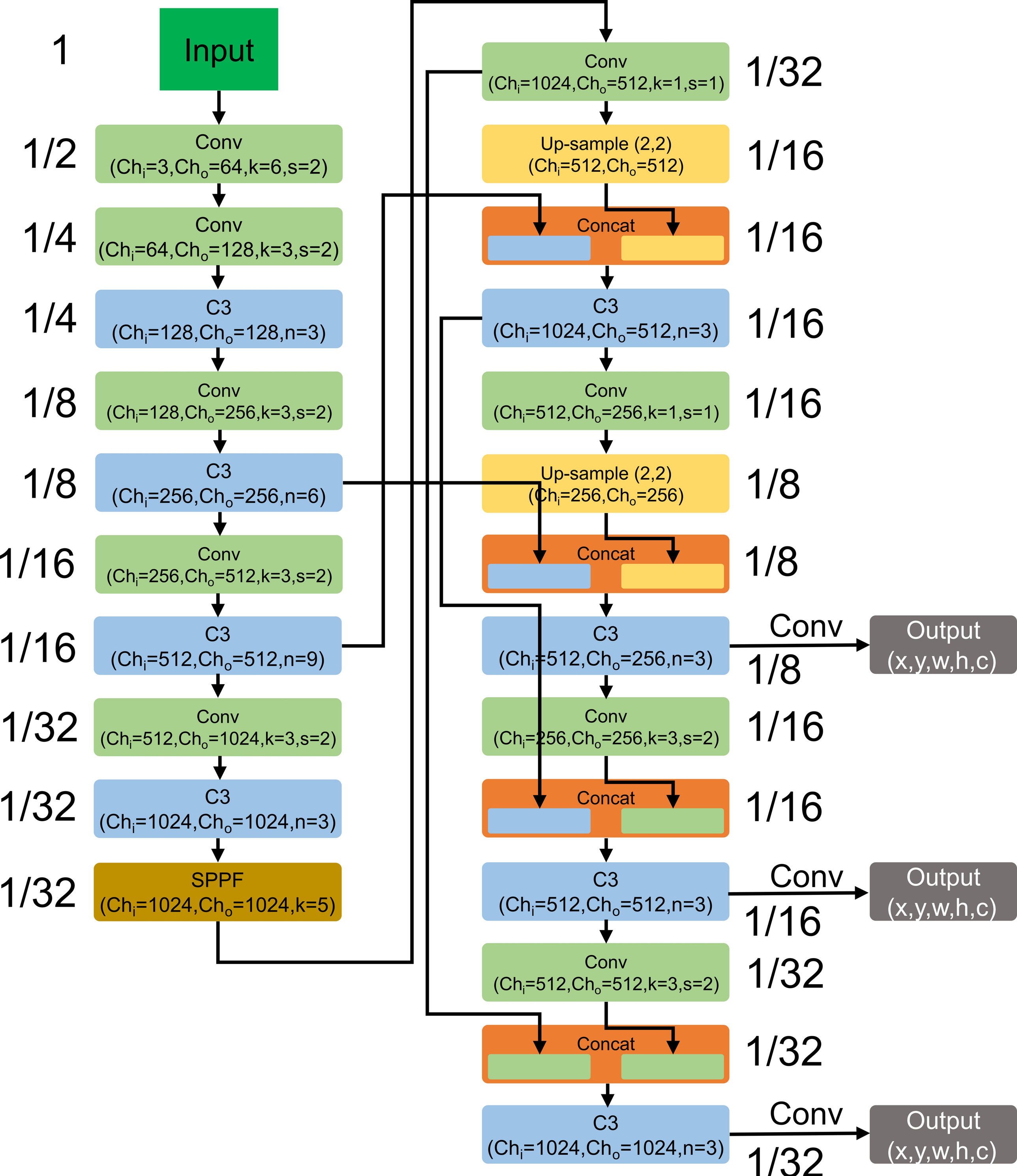
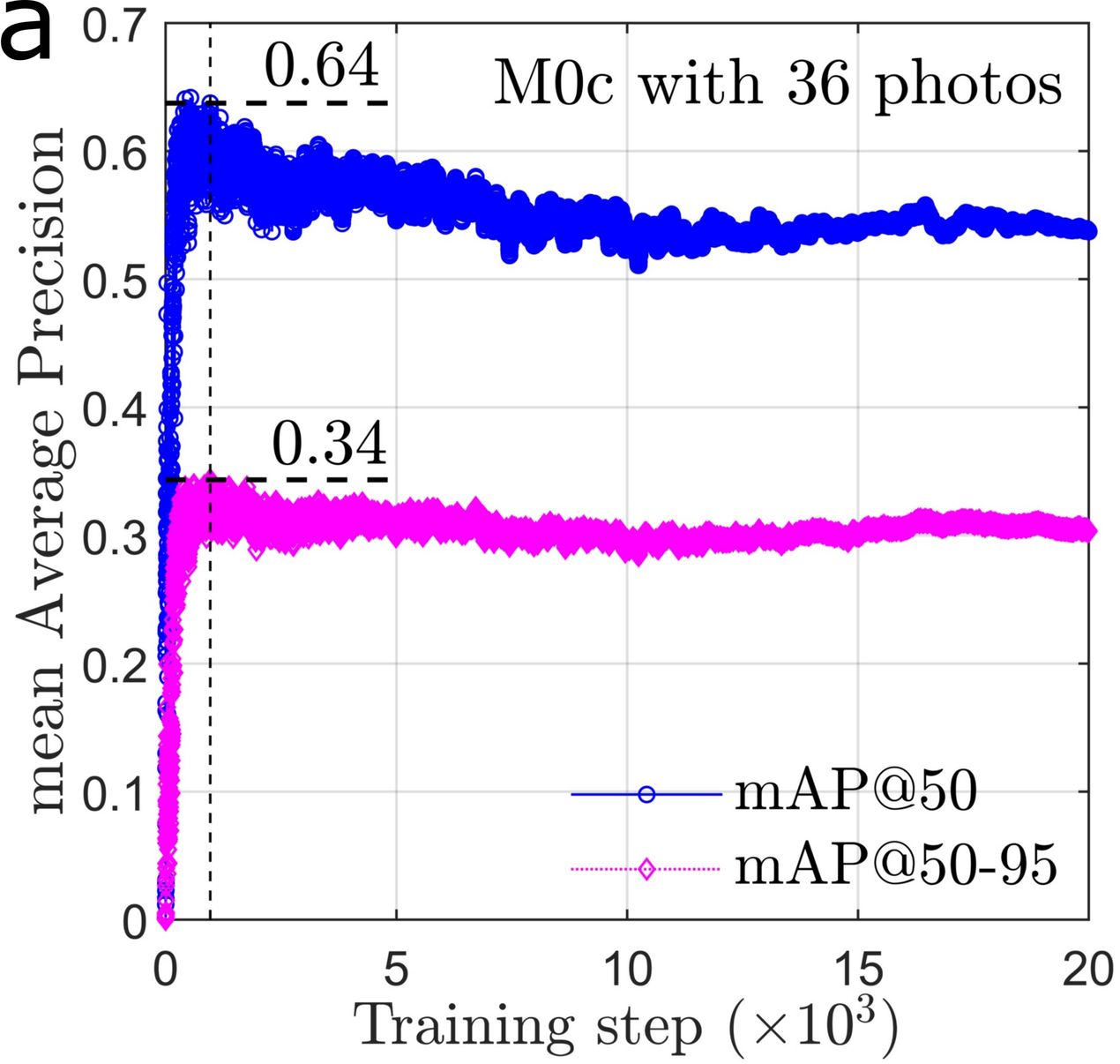
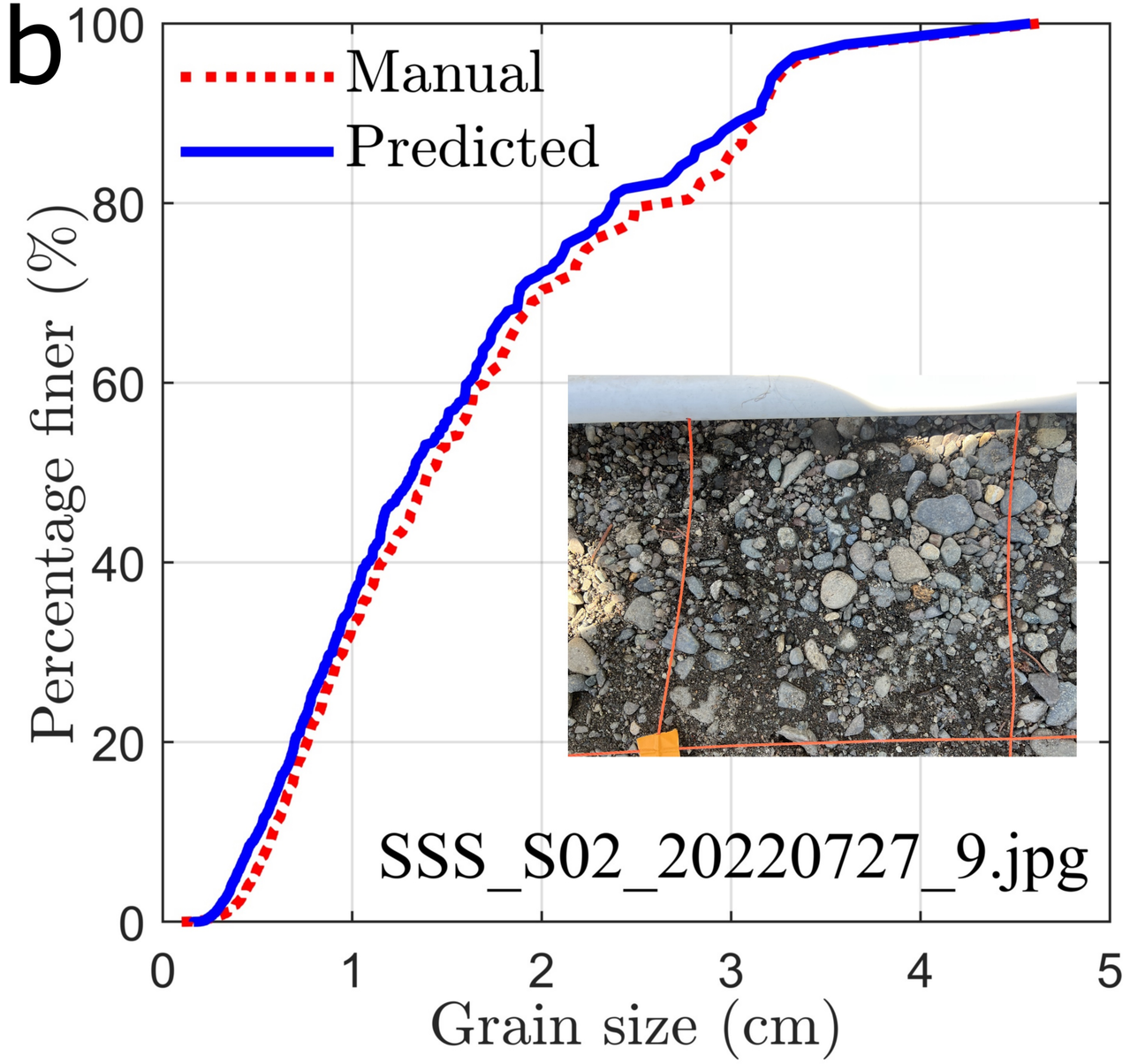

1
Input
1/2
Conv
(Ch$_i$=3,Ch$_o$=64,k=6,s=2)
1/4
Conv
(Ch$_i$=64,Ch$_o$=128,k=3,s=2)
1/4
C3
(Ch$_i$=128,Ch$_o$=128,n=3)
1/8
Conv
(Ch$_i$=128,Ch$_o$=256,k=3,s=2)
1/8
C3
(Ch$_i$=256,Ch$_o$=256,n=6)
1/16
Conv
(Ch$_i$=256,Ch$_o$=512,k=3,s=2)
1/16
C3
(Ch$_i$=512,Ch$_o$=512,n=9)
1/32
Conv
(Ch$_i$=512,Ch$_o$=1024,k=3,s=2)
1/32
C3
(Ch$_i$=1024,Ch$_o$=1024,n=3)
1/32
SPPF
(Ch$_i$=1024,Ch$_o$=1024,k=5)
Conv
(Ch$_i$=1024,Ch$_o$=512,k=1,s=1)
1/32
Up-sample (2,2)
(Ch$_i$=512,Ch$_o$=512)
1/16
Concat
1/16
C3
(Ch$_i$=1024,Ch$_o$=512,n=3)
1/16
Conv
(Ch$_i$=512,Ch$_o$=256,k=1,s=1)
1/16
Up-sample (2,2)
(Ch$_i$=256,Ch$_o$=256)
1/8
Concat
1/8
C3
(Ch$_i$=512,Ch$_o$=256,n=3)
Conv
1/8
Output
(x,y,w,h,c)
Conv
(Ch$_i$=256,Ch$_o$=256,k=3,s=2)
1/16
Concat
1/16
C3
(Ch$_i$=512,Ch$_o$=512,n=3)
Conv
1/16
Output
(x,y,w,h,c)
Conv
(Ch$_i$=512,Ch$_o$=512,k=3,s=2)
1/32
Concat
1/32
C3
(Ch$_i$=1024,Ch$_o$=1024,n=3)
Conv
1/32
Output
(x,y,w,h,c)

**Figure 4.**

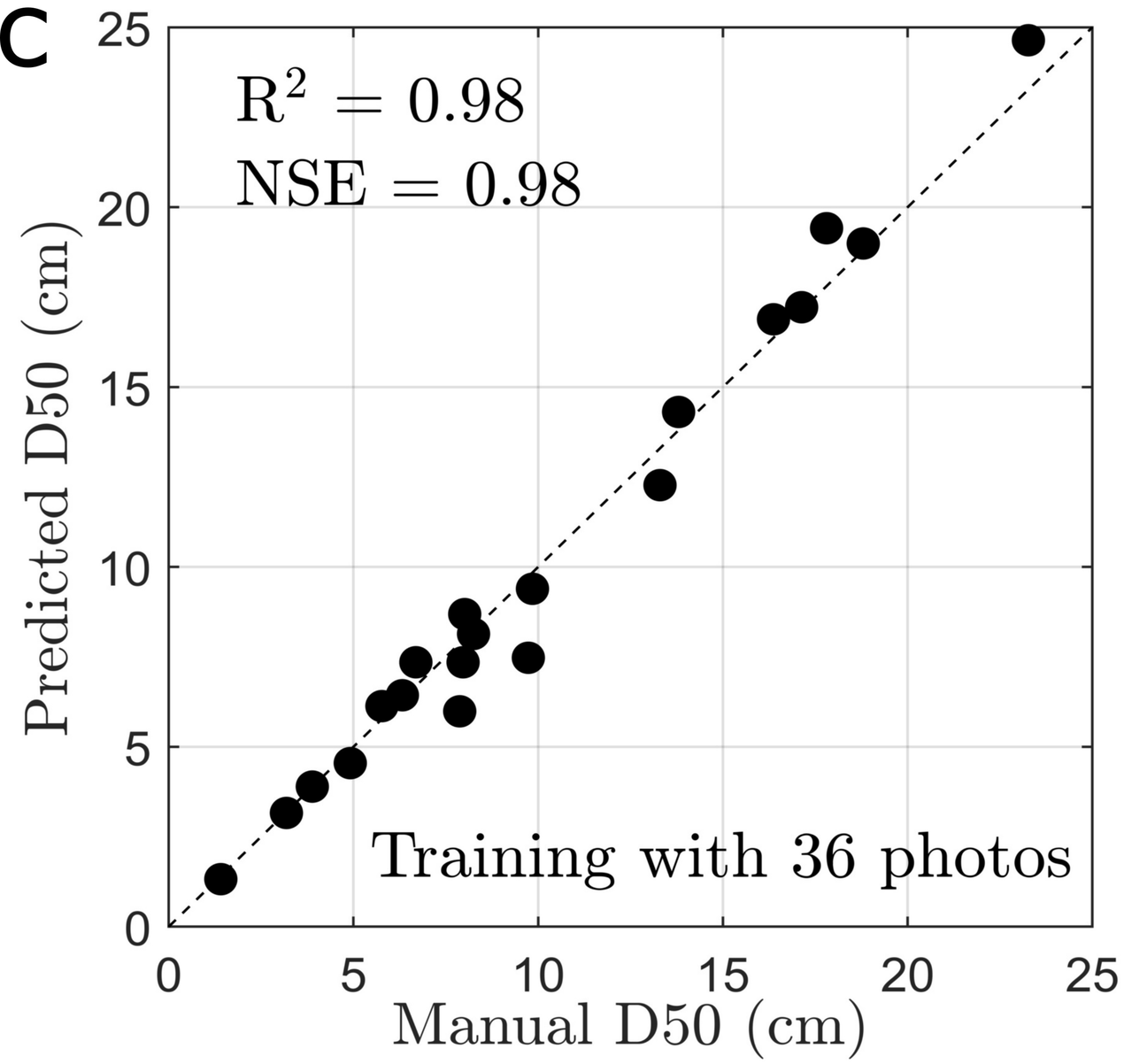
a


0.64
M0c with 36 photos
0.34
mAP@50
mAP@50-95
mean Average Precision
Training step (×10^3)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0
5
10
15
20

b


Manual
Predicted
Percentage finer (%)
Grain size (cm)
SSS_S02_20220727_9.jpg
0
20
40
60
80
100
0
1
2
3
4
5

c


R^2 = 0.98
NSE = 0.98
Training with 36 photos
Predicted D50 (cm)
Manual D50 (cm)
0
5
10
15
20
25

d

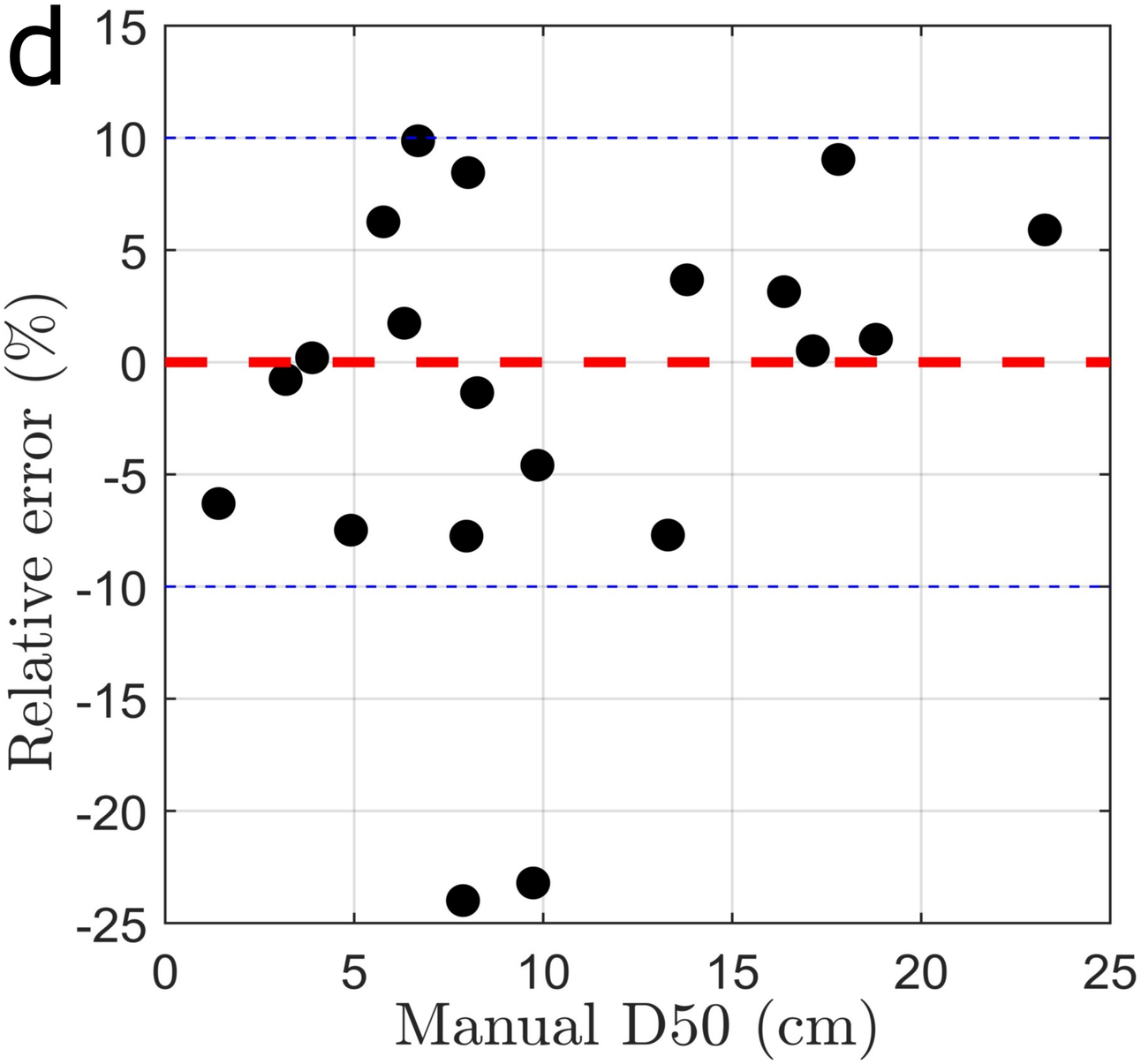
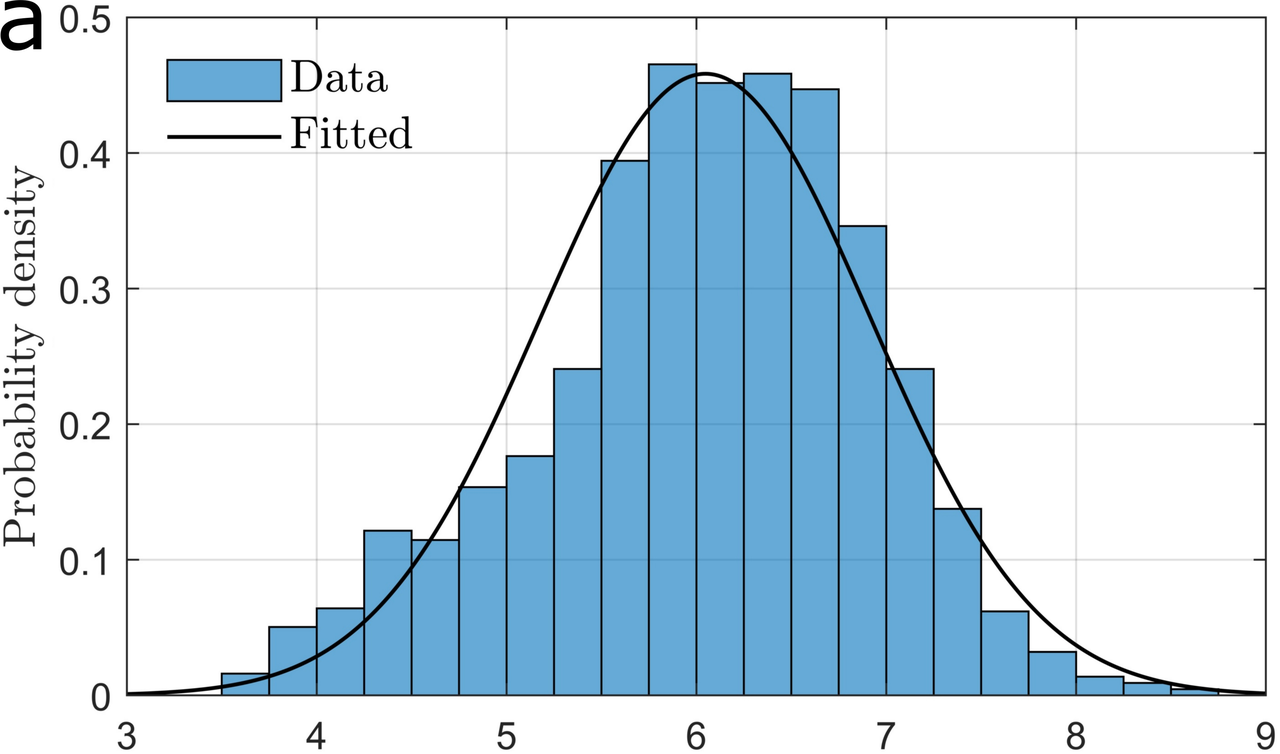
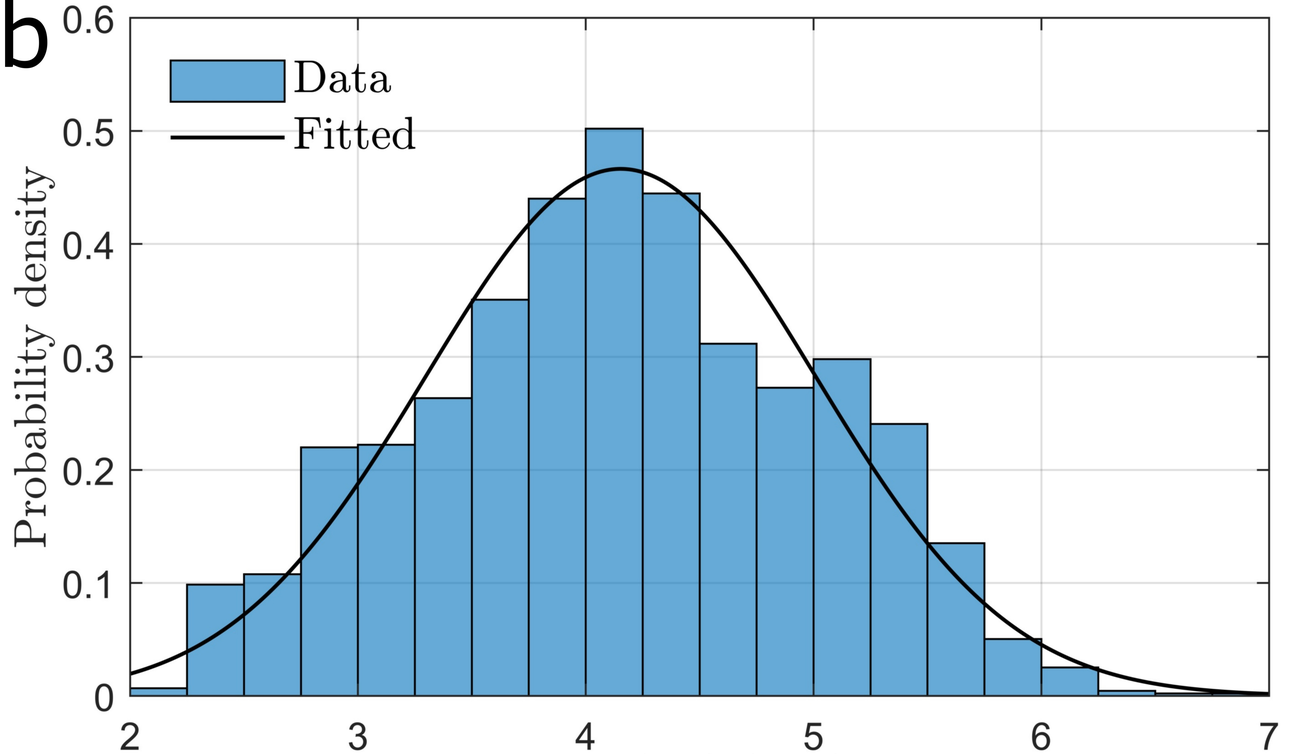

Relative error (%)
Manual D50 (cm)
-25
-20
-15
-10
-5
0
5
10
15
0
5
10
15
20
25

**Figure 5.**

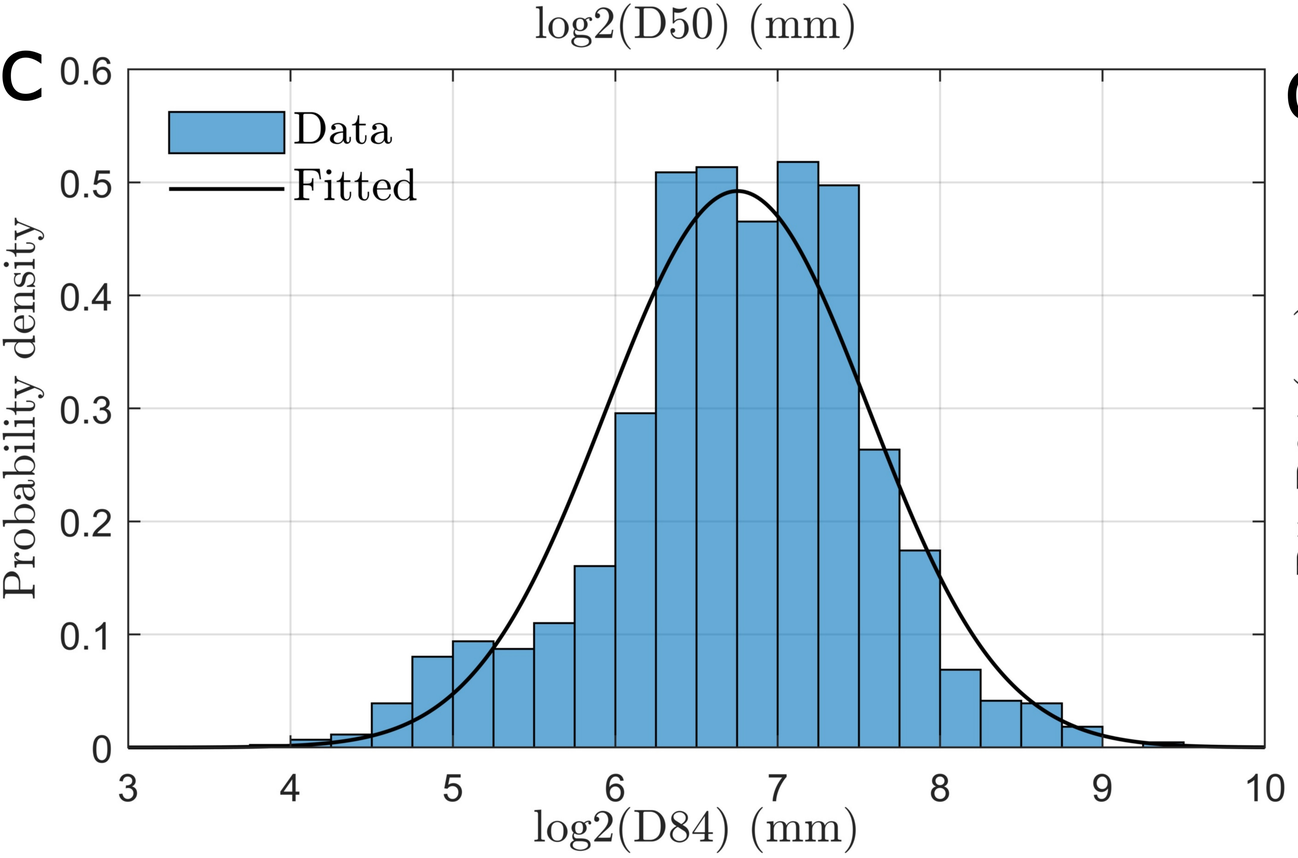
a


Data
Fitted
Probability density
0
0.1
0.2
0.3
0.4
0.5
3
4
5
6
7
8
9

b


Data
Fitted
Probability density
0
0.1
0.2
0.3
0.4
0.5
0.6
2
3
4
5
6
7
log2(D5) (mm)

c


Data
Fitted
Probability density
0
0.1
0.2
0.3
0.4
0.5
0.6
3
4
5
6
7
8
9
10
log2(D50) (mm)
log2(D84) (mm)

d

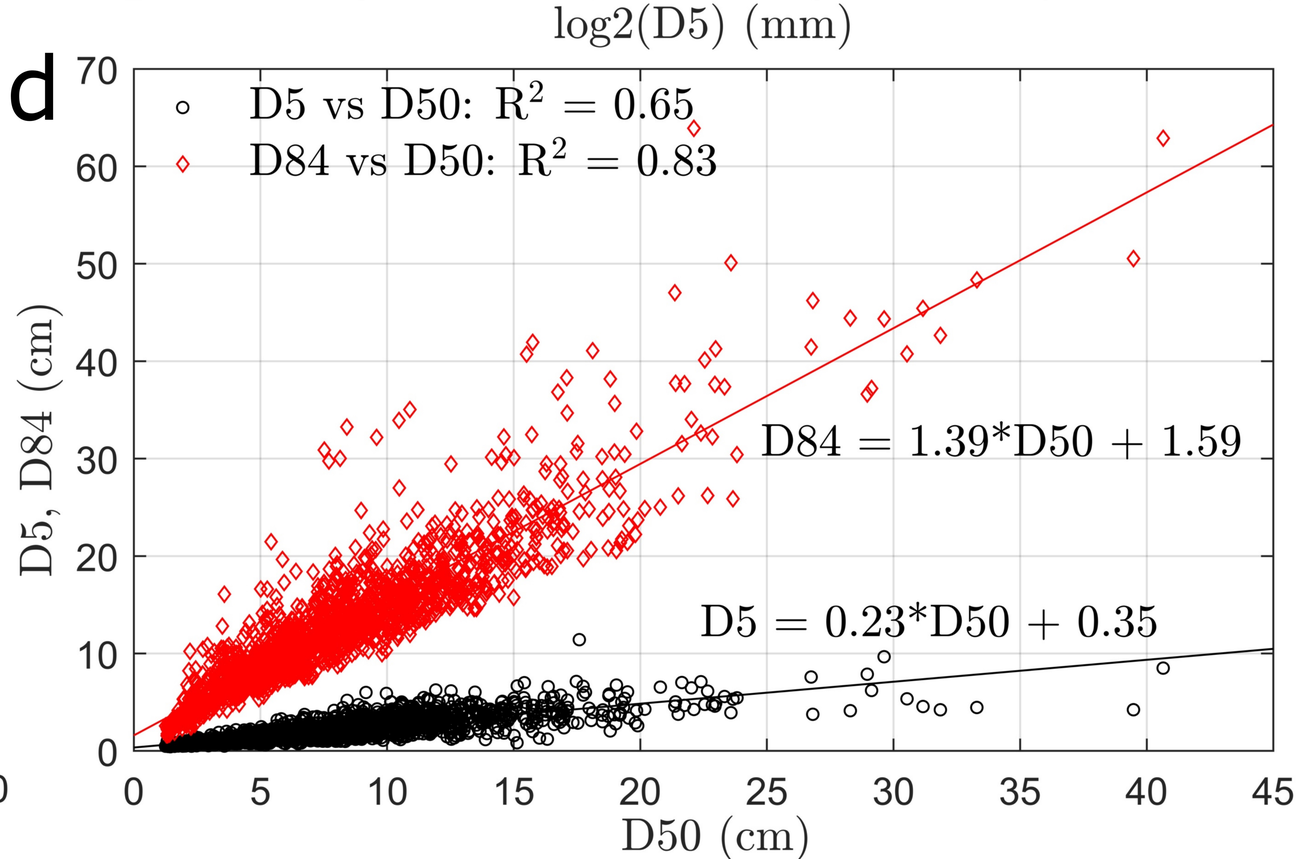
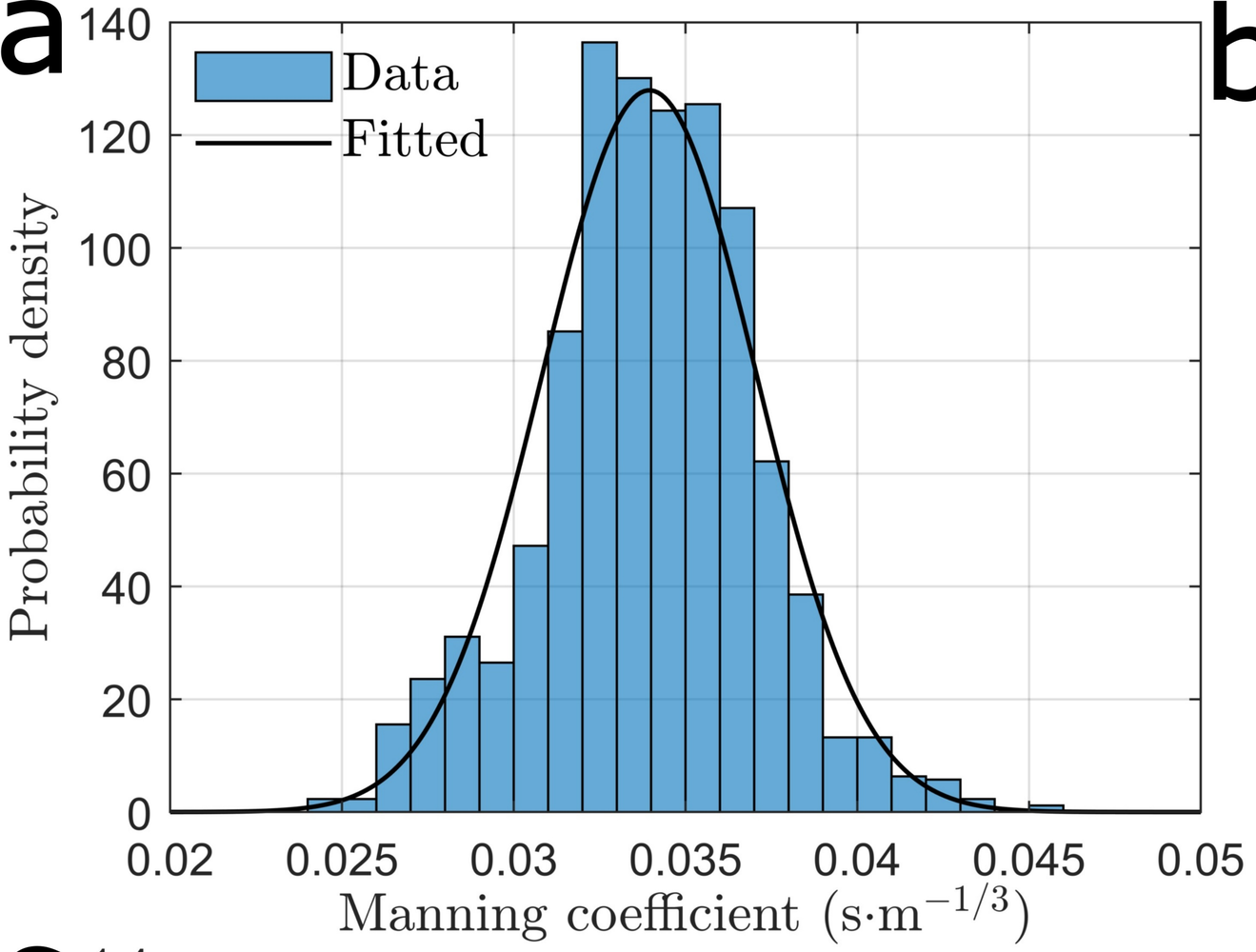
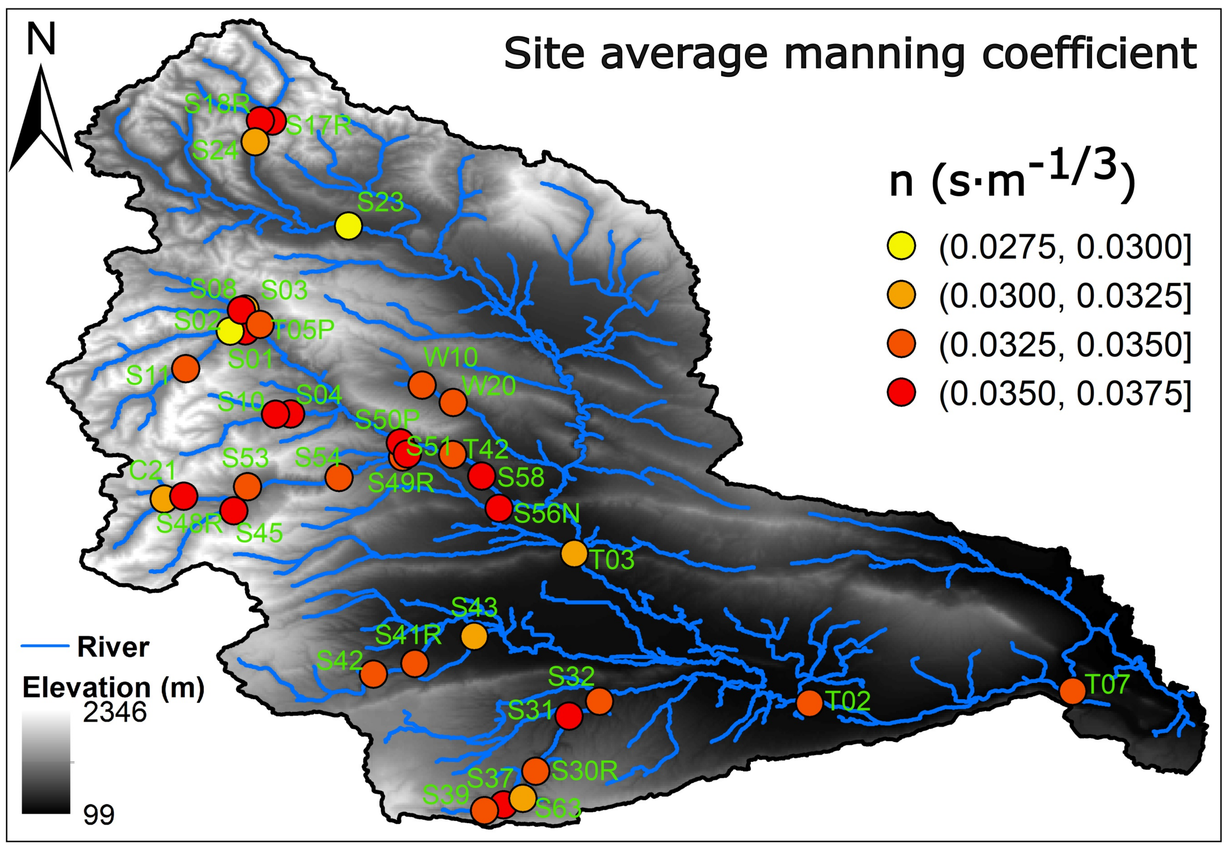

D5 vs D50: R² = 0.65
D84 vs D50: R² = 0.83
D84 = 1.39*D50 + 1.59
D5 = 0.23*D50 + 0.35
D5, D84 (cm)
0
10
20
30
40
50
60
70
0
5
10
15
20
25
30
35
40
45
D50 (cm)

**Figure 6.**

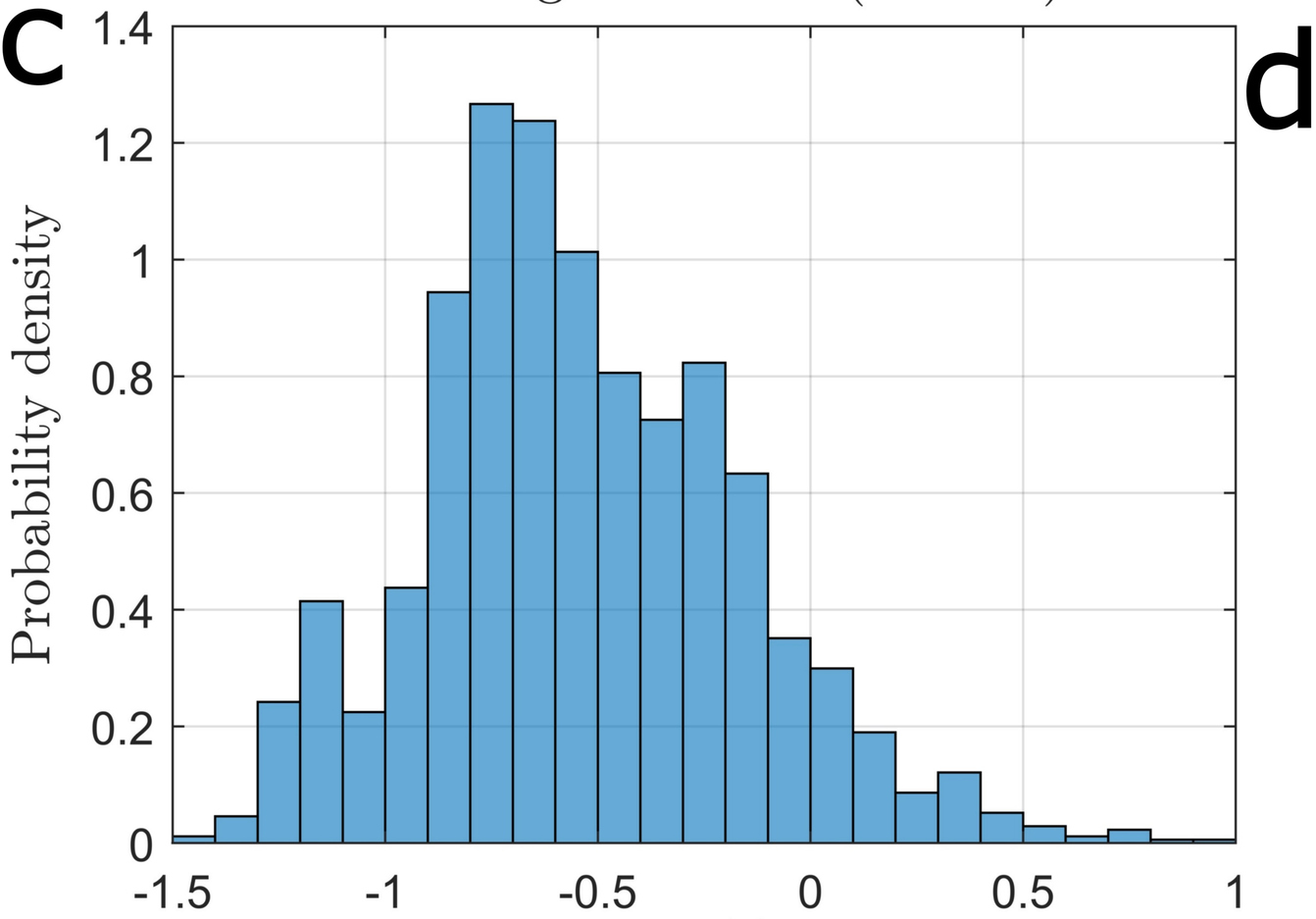
a


Data
Fitted
Probability density
Manning coefficient (s·m$^{-1/3}$)
0
20
40
60
80
100
120
140
0.02
0.025
0.03
0.035
0.04
0.045
0.05

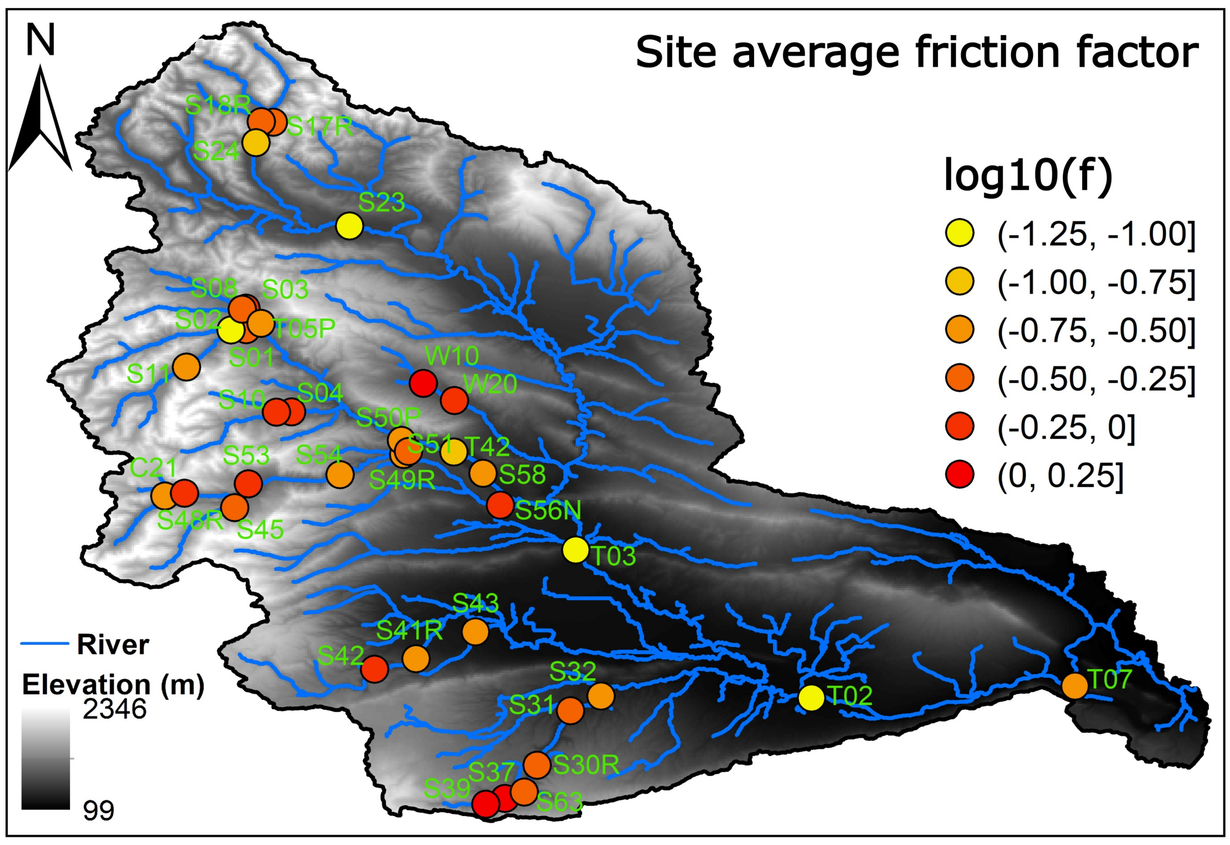
b


N
Site average manning coefficient
n (s·m$^{-1/3}$)
(0.0275, 0.0300]
(0.0300, 0.0325]
(0.0325, 0.0350]
(0.0350, 0.0375]
River
Elevation (m)
2346
99
S18R
S17R
S24
S23
S08
S03
S02
T05P
S01
S11
W10
W20
S10
S04
S50P
S51
T42
S53
S54
S49R
S58
C21
S56N
S48R
S45
T03
S43
S41R
S42
S32
S31
T02
T07
S37
S30R
S39
S63

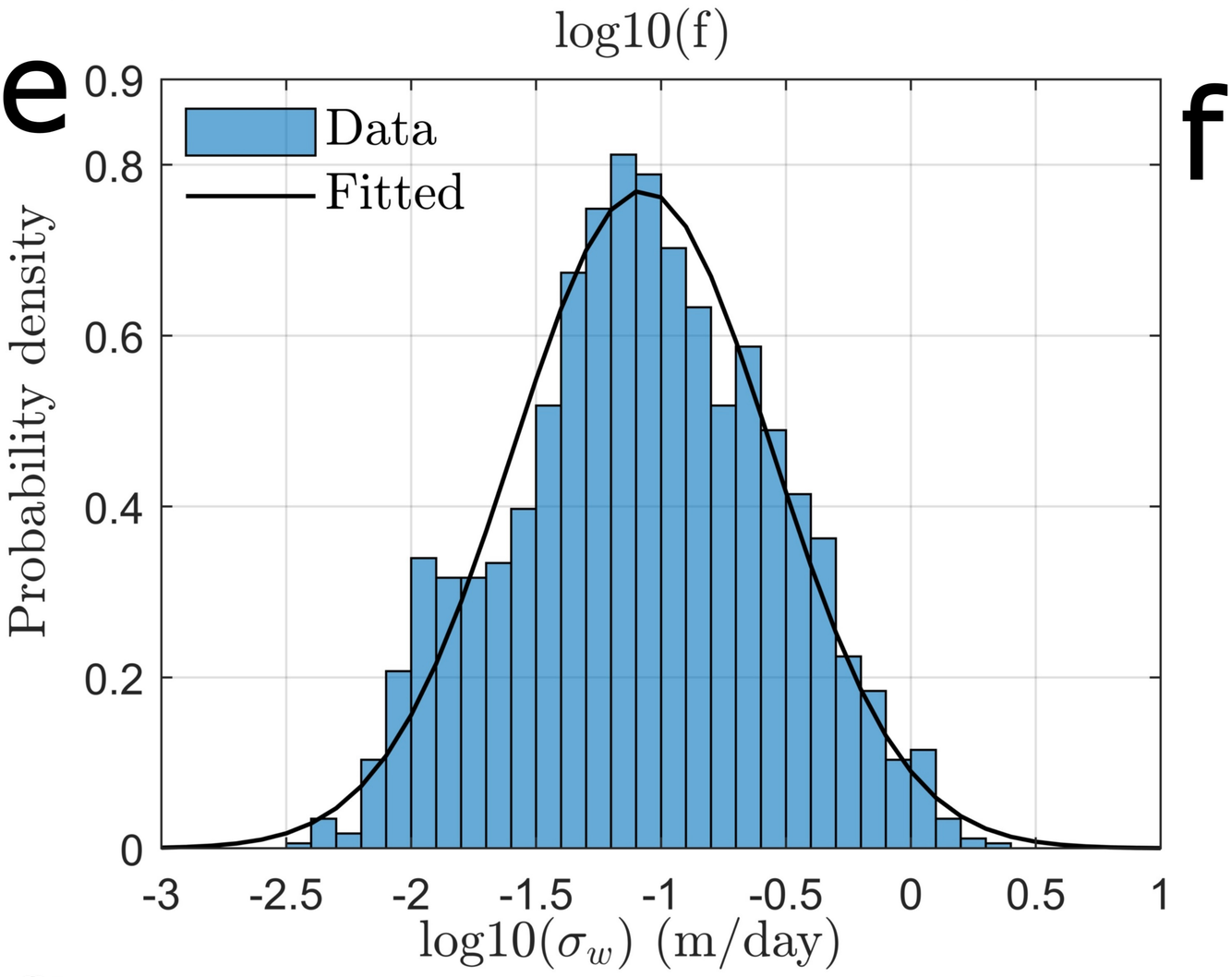
c


Probability density
0
0.2
0.4
0.6
0.8
1
1.2
1.4
-1.5
-1
-0.5
0
0.5
1

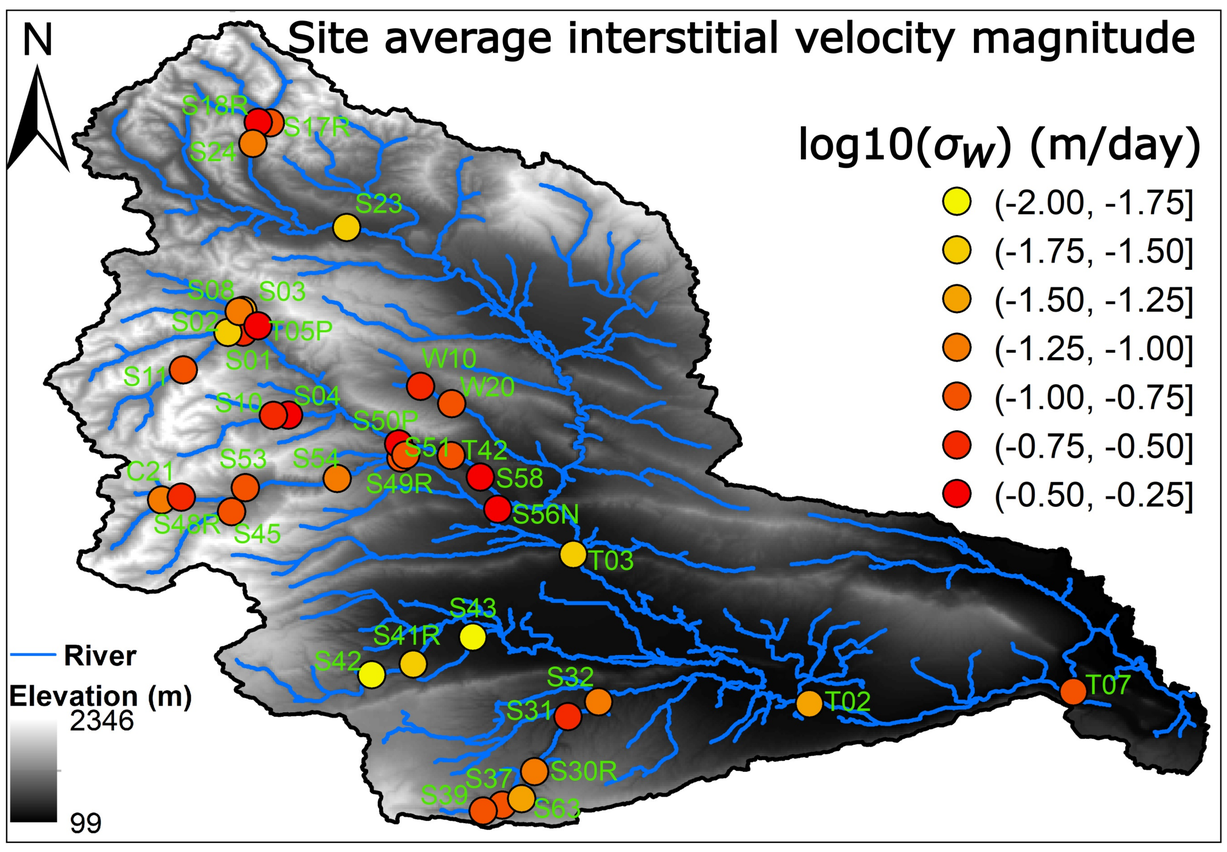
d


N
Site average friction factor
log10(f)
(-1.25, -1.00]
(-1.00, -0.75]
(-0.75, -0.50]
(-0.50, -0.25]
(-0.25, 0]
(0, 0.25]
River
Elevation (m)
2346
99
S18R
S17R
S24
S23
S08
S03
S02
T05P
S01
S11
W10
W20
S10
S04
S50P
S51
T42
S53
S54
S49R
S58
C21
S56N
S48R
S45
T03
S43
S41R
S42
S32
S31
T02
T07
S37
S30R
S39
S63

e


Data
Fitted
Probability density
log10(f)
log10($\sigma_w$) (m/day)
0
0.2
0.4
0.6
0.8
0.9
-3
-2.5
-2
-1.5
-1
-0.5
0
0.5
1

f


N
Site average interstitial velocity magnitude
log10($\sigma_W$) (m/day)
(-2.00, -1.75]
(-1.75, -1.50]
(-1.50, -1.25]
(-1.25, -1.00]
(-1.00, -0.75]
(-0.75, -0.50]
(-0.50, -0.25]
River
Elevation (m)
2346
99
S18R
S17R
S24
S23
S08
S03
S02
T05P
S01
S11
W10
W20
S10
S04
S50P
S51
T42
S53
S54
S49R
S58
C21
S56N
S48R
S45
T03
S43
S41R
S42
S32
S31
T02
T07
S37
S30R
S39
S63

g


$[NO_3^-]$ = 1e-4 mmol/L
$[NO_3^-]$ = 1e-2 mmol/L
$[NO_3^-]$ = 1 mmol/L
Probability density
log10($u_f$) (m/day)
0
0.5
1
1.5
2
2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2

h


N
Site average nitrate uptake velocity
$[NO_3^-]$ = 1e-2 mmol/L
$u_f$ (m/day)
(0, 0.5]
(0.5, 1.0]
(1.0, 1.5]
(1.5, 2.0]
(2.0, 2.5]
(2.5, 3.0]
River
Elevation (m)
2346
99
S18R
S17R
S24
S23
S08
S03
S02
T05P
S01
S11
W10
W20
S10
S04
S50P
S51
T42
S53
S54
S49R
S58
C21
S56N
S48R
S45
T03
S43
S41R
S42
S32
S31
T02
T07
S37
S30R
S39
S63

**Figure 7.**

a

Resolution (mm/pixel)

$r_{SC}$(%)

b

log2(D50) (mm)

$r_{50}$(%)

c

log2(D84) (mm)

$r_{84}$(%)

d

log2(D5) (mm)

$r_5$(%)

e

H (m)

$r_H$(%)

C21 S01 S02 S03 S04 S08 S10 S11 S17R S18R S23 S24 S30R S31 S32 S37 S39 S41R S42 S43 S45 S48R S49R S50P S51 S53 S54 S56N S58 S63 T02 T03 T05P T07 T42 W10 W20

**Figure 8.**

a

b

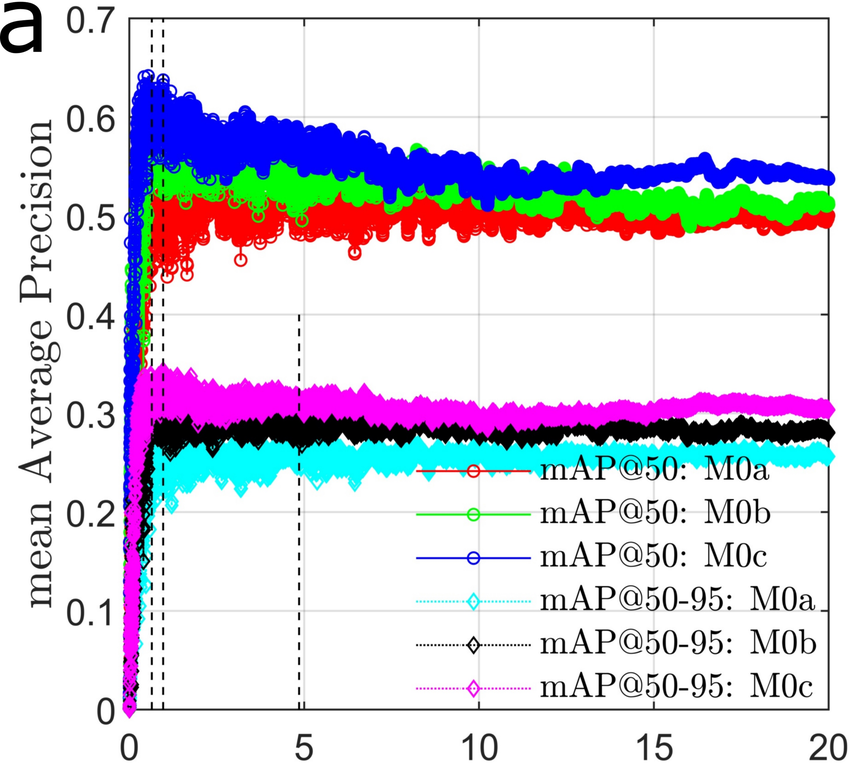
c

d

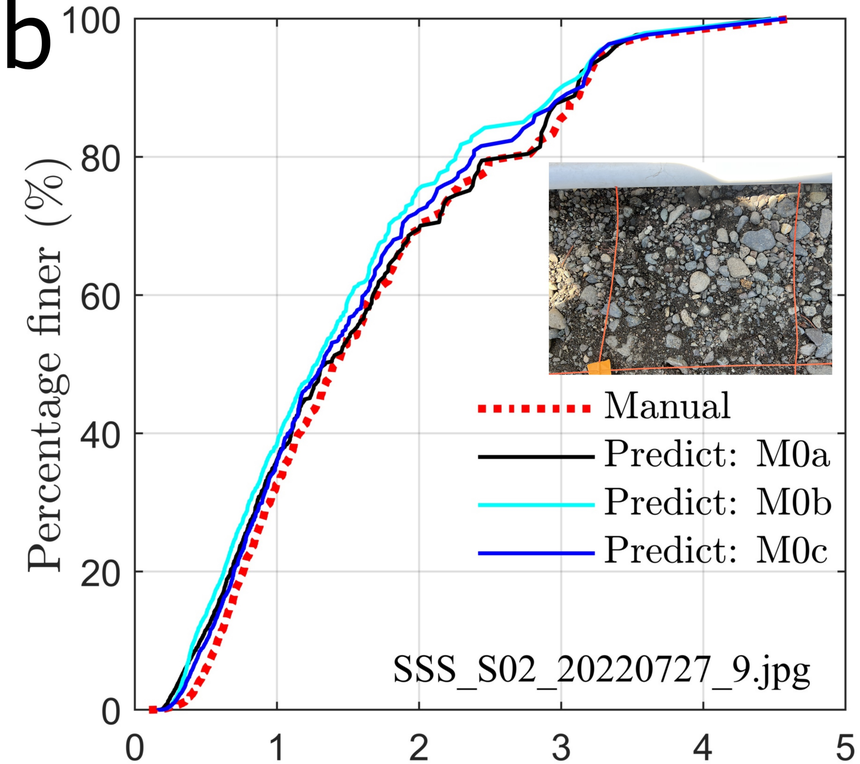
C21 S01 S02 S03 S04 S08 S10 S11 S17R S18R S23 S24 S30R S31 S32 S37 S39 S41R S42 S43 S45 S48R S49R S50P S51 S53 S54 S56N S58 S63 T02 T03 T05P T07 T42 W10 W20

**Figure 9.**

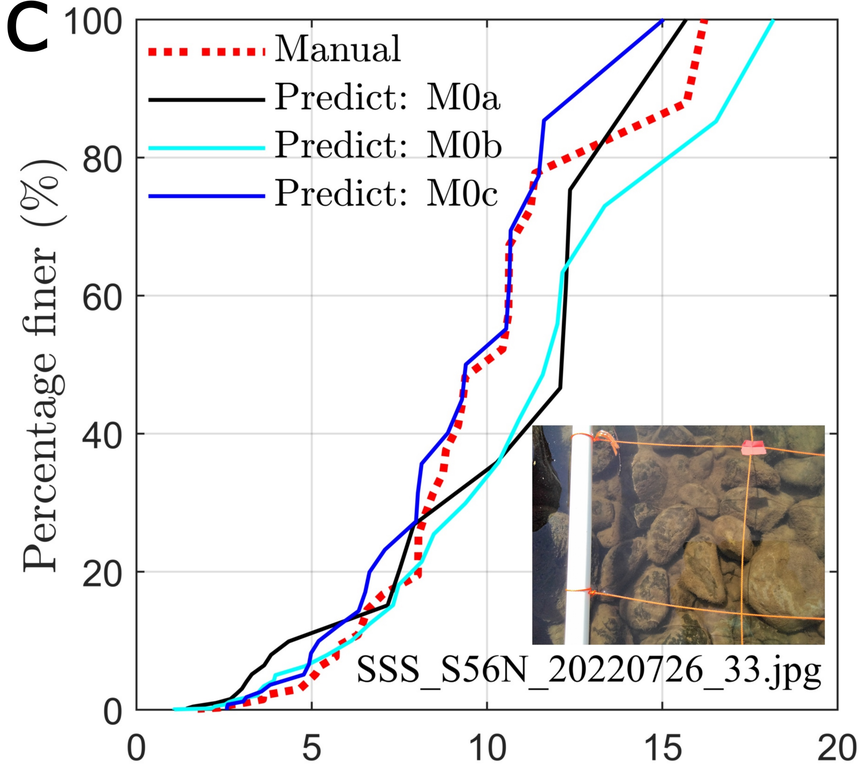
a


mean Average Precision
Training step (/1000)
mAP@50: M0a
mAP@50: M0b
mAP@50: M0c
mAP@50-95: M0a
mAP@50-95: M0b
mAP@50-95: M0c

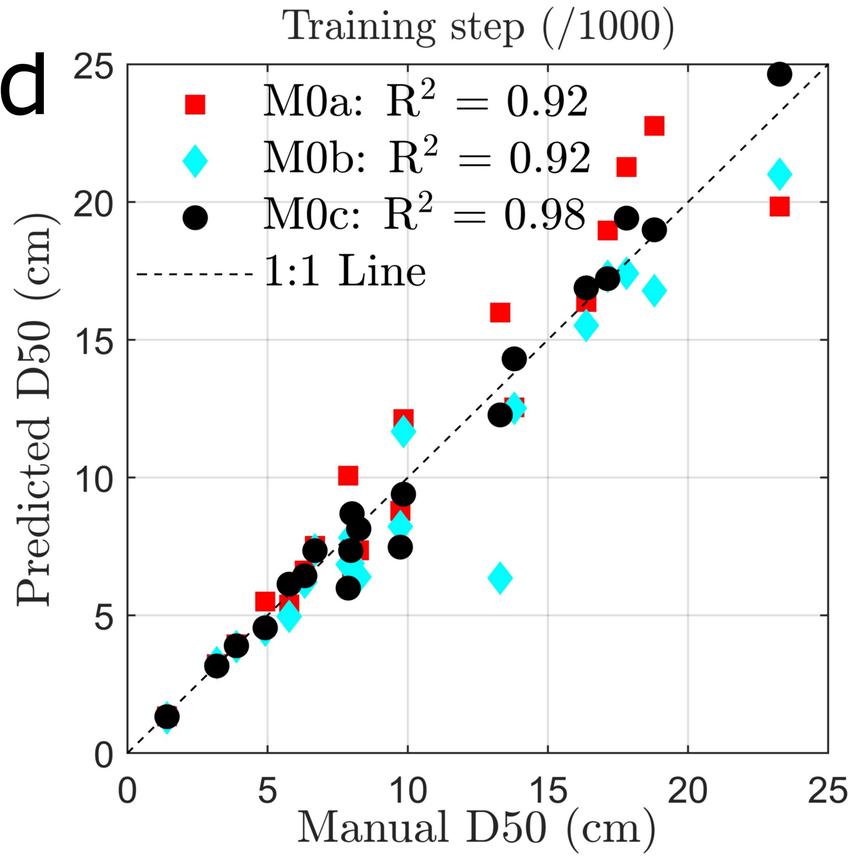
b


Percentage finer (%)
Manual
Predict: M0a
Predict: M0b
Predict: M0c
SSS_S02_20220727_9.jpg

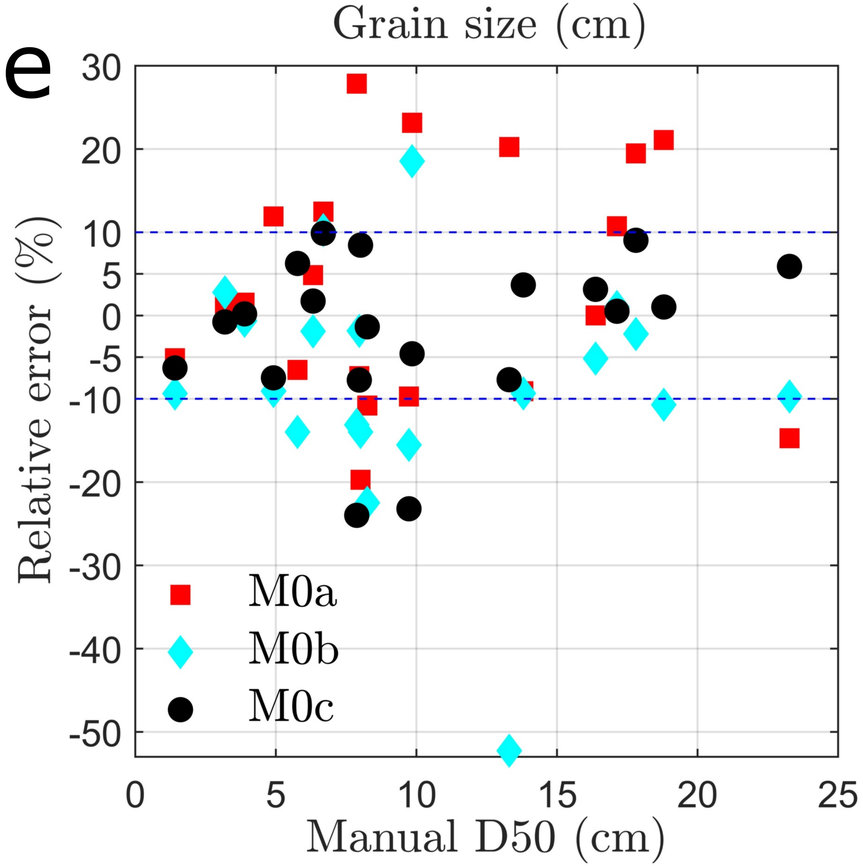
c


Percentage finer (%)
Manual
Predict: M0a
Predict: M0b
Predict: M0c
SSS_S56N_20220726_33.jpg

d


Predicted D50 (cm)
Manual D50 (cm)
M0a: R² = 0.92
M0b: R² = 0.92
M0c: R² = 0.98
1:1 Line

e


Relative error (%)
Grain size (cm)
Manual D50 (cm)
M0a
M0b
M0c

f

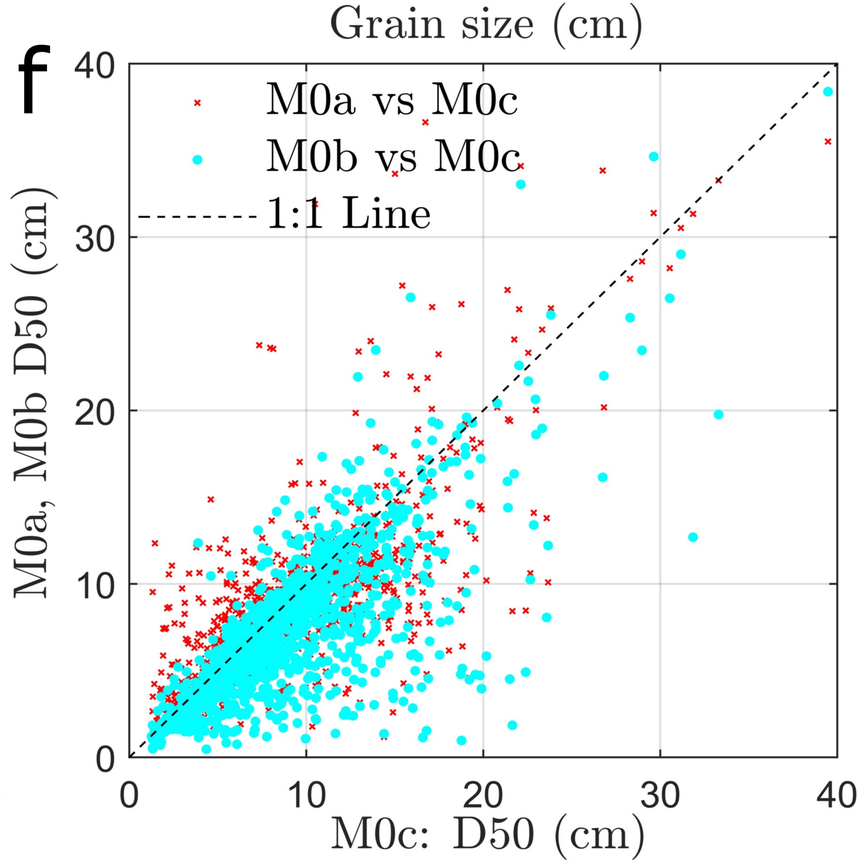

M0a, M0b D50 (cm)
Grain size (cm)
M0c: D50 (cm)
M0a vs M0c
M0b vs M0c
1:1 Line

**Figure 10.**

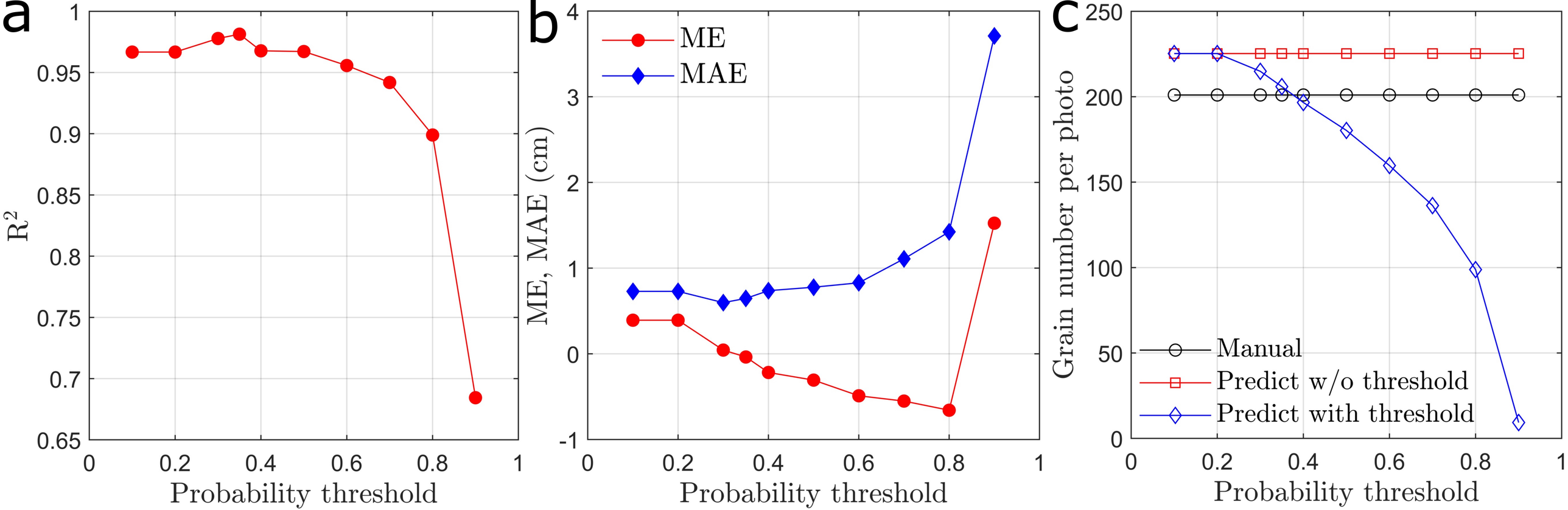

a
b
c
$R^2$
Probability threshold
ME, MAE (cm)
ME
MAE
Grain number per photo
Manual
Predict w/o threshold
Predict with threshold

**Figure 11.**

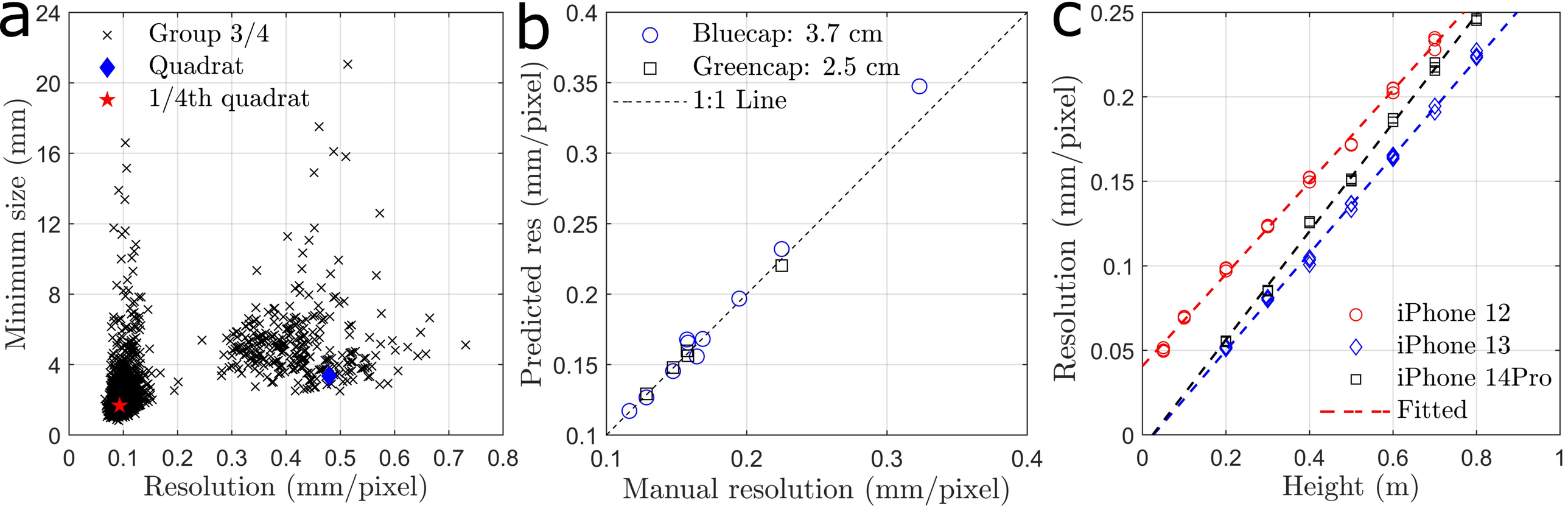

a
Group 3/4
Quadrat
1/4th quadrat
Minimum size (mm)
Resolution (mm/pixel)
0
4
8
12
16
20
24
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
b
Bluecap: 3.7 cm
Greencap: 2.5 cm
1:1 Line
Predicted res (mm/pixel)
Manual resolution (mm/pixel)
0.1
0.15
0.2
0.25
0.3
0.35
0.4
c
iPhone 12
iPhone 13
iPhone 14Pro
Fitted
Resolution (mm/pixel)
Height (m)
0
0.05
0.1
0.15
0.2
0.25
0.2
0.4
0.6
0.8
1

**Figure 12.**

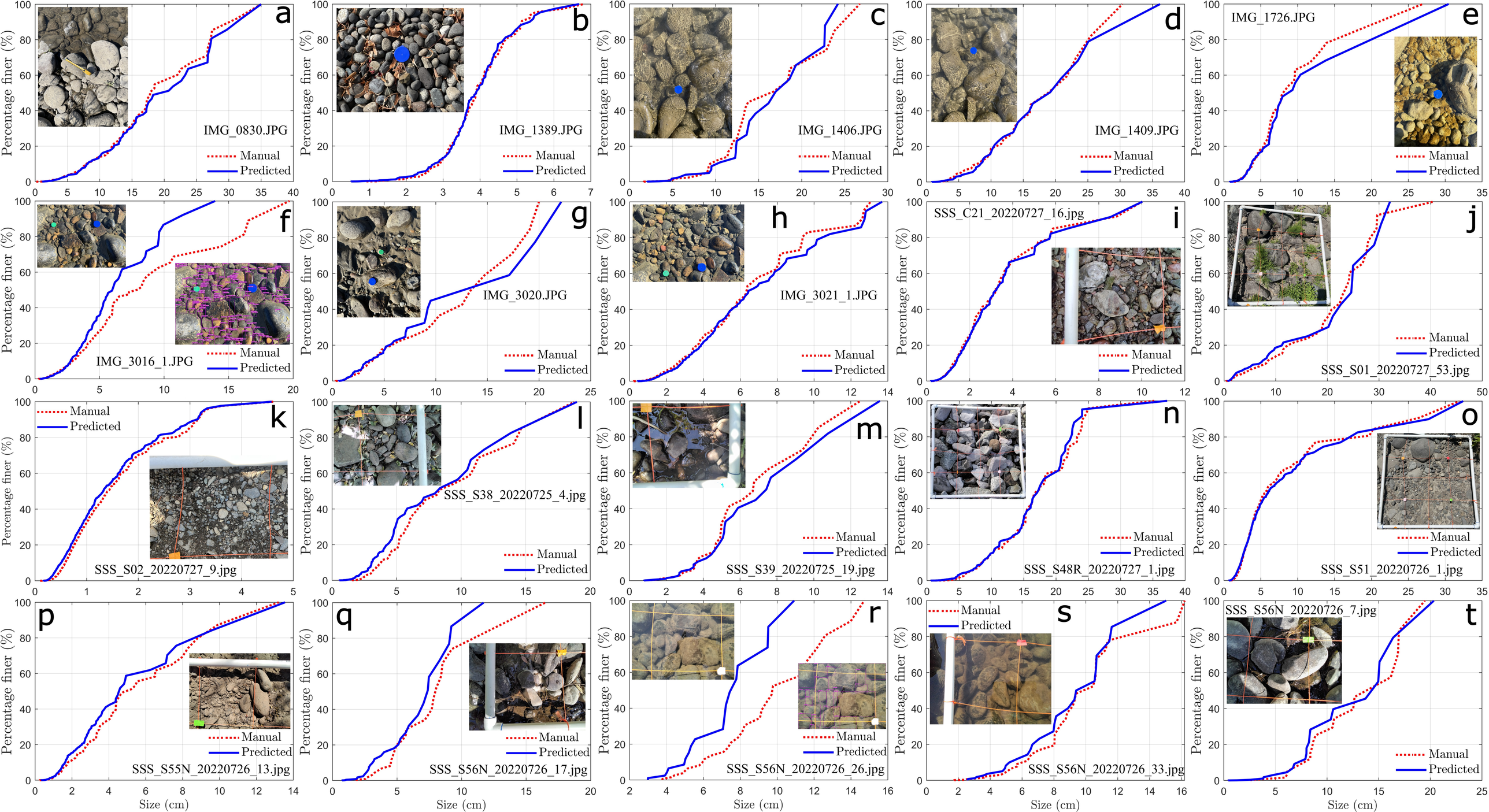
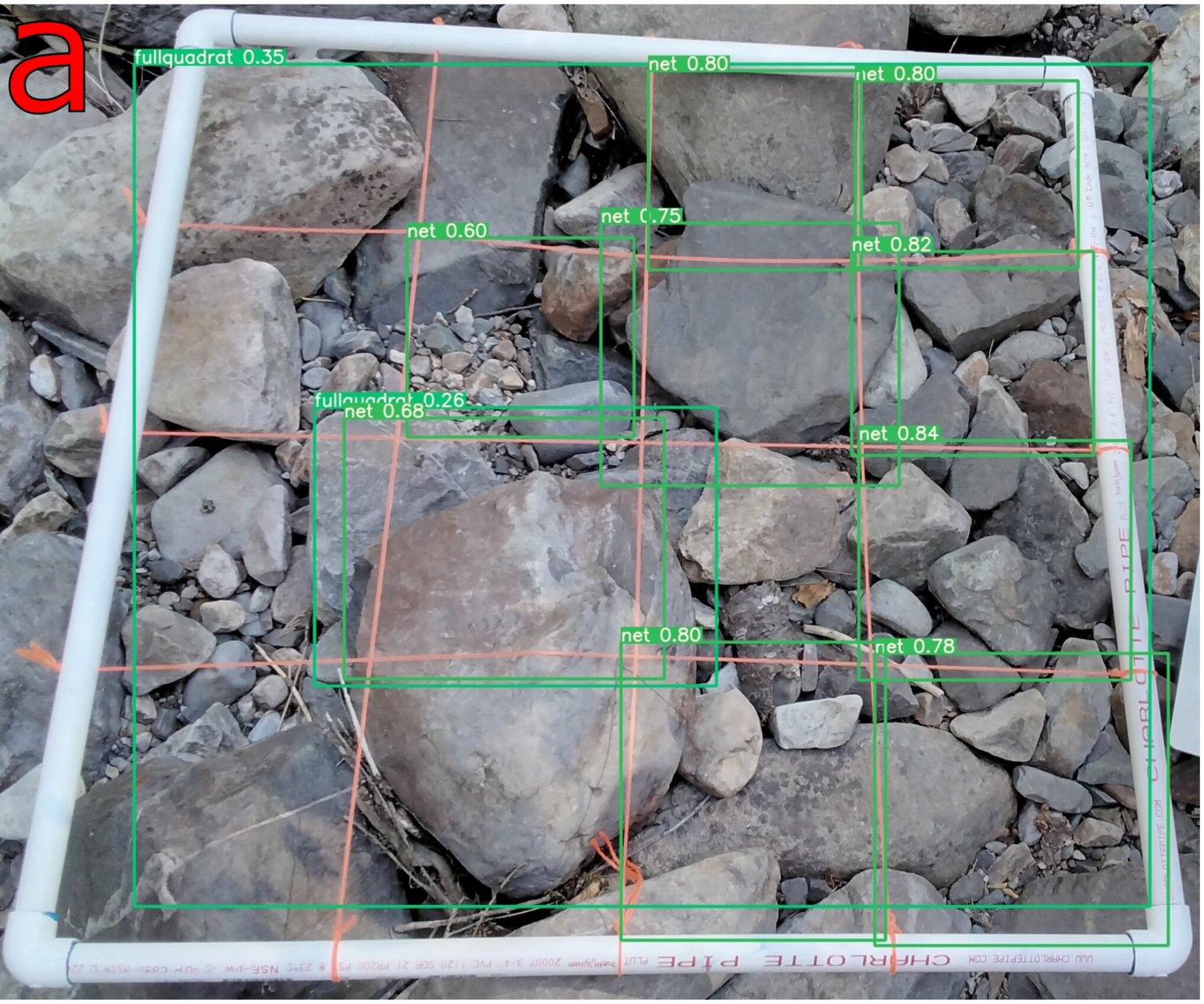
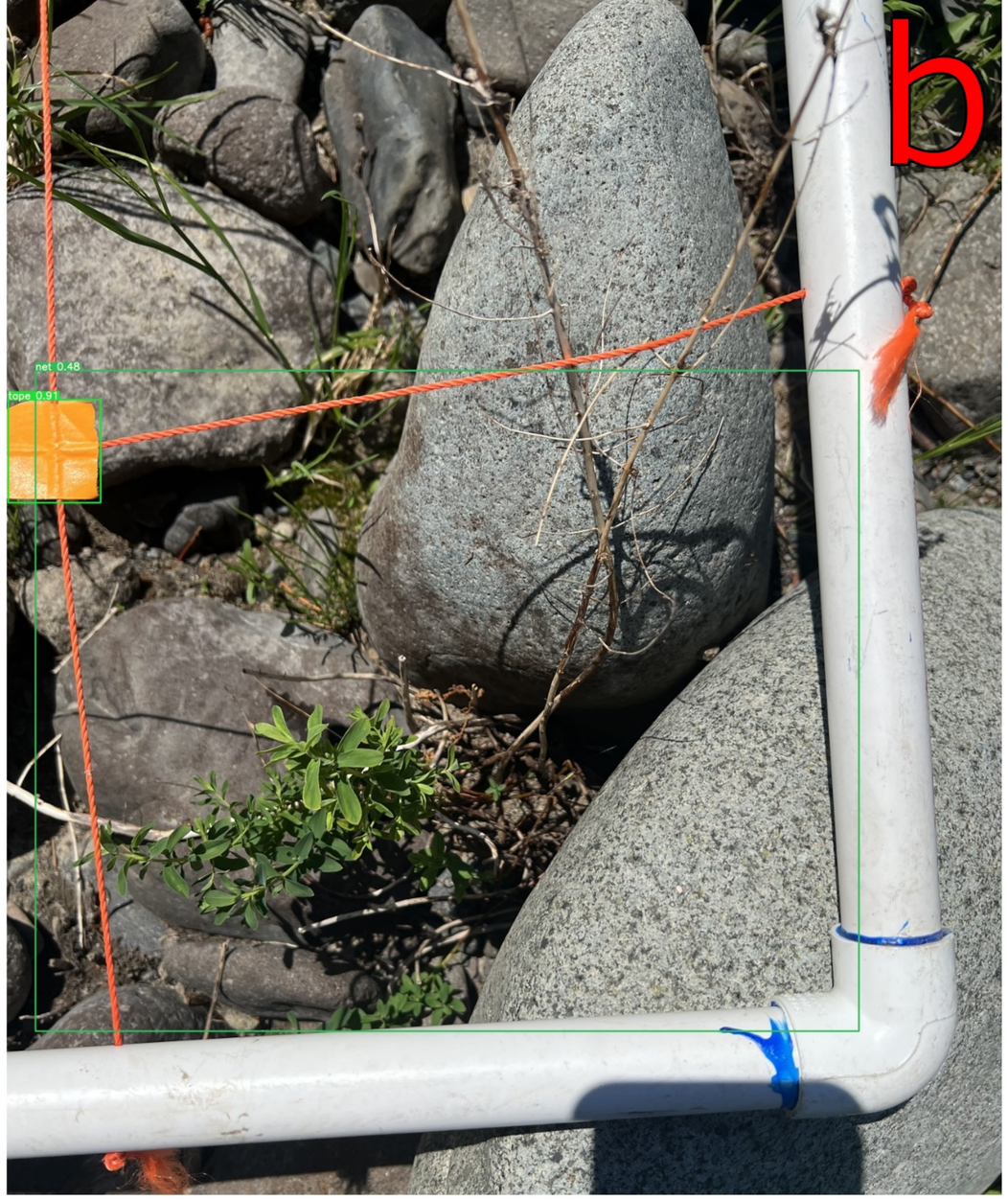

a
IMG_0830.JPG
b
IMG_1389.JPG
c
IMG_1406.JPG
d
IMG_1409.JPG
e
IMG_1726.JPG
f
IMG_3016_1.JPG
g
IMG_3020.JPG
h
IMG_3021_1.JPG
i
SSS_C21_20220727_16.jpg
j
SSS_S01_20220727_53.jpg
k
SSS_S02_20220727_9.jpg
l
SSS_S38_20220725_4.jpg
m
SSS_S39_20220725_19.jpg
n
SSS_S48R_20220727_1.jpg
o
SSS_S51_20220726_1.jpg
p
SSS_S55N_20220726_13.jpg
q
SSS_S56N_20220726_17.jpg
r
SSS_S56N_20220726_26.jpg
s
SSS_S56N_20220726_33.jpg
t
SSS_S56N_20220726_7.jpg
Manual
Predicted
Percentage finer (%)
Size (cm)

**Figure 13.**

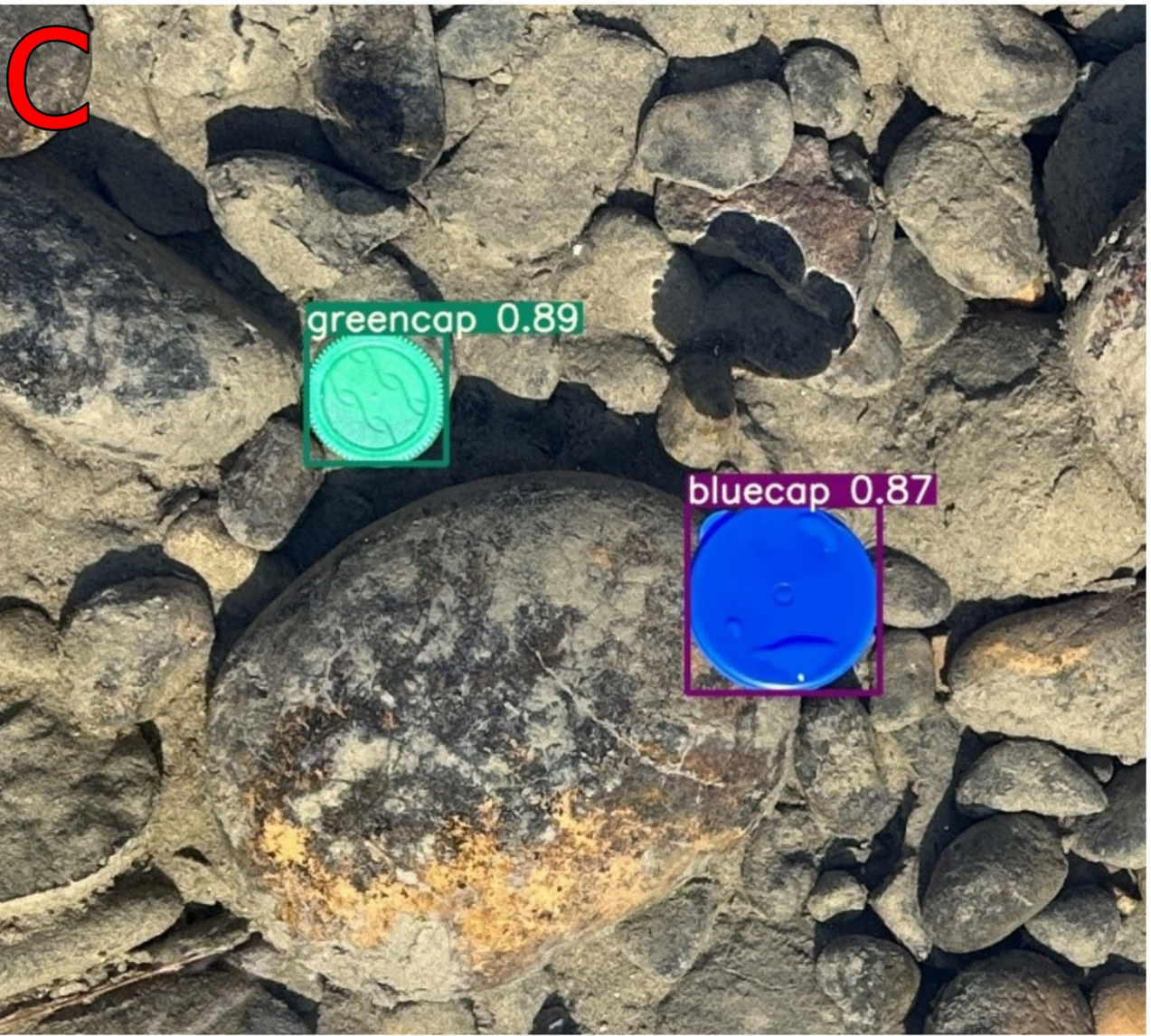
a


fullquadrat 0.35
net 0.80
net 0.80
net 0.75
net 0.60
net 0.82
fullquadrat 0.26
net 0.68
net 0.84
net 0.80
net 0.78

b


net 0.48
tape 0.91

c


greencap 0.89
bluecap 0.87

d


bluecap 0.89

# Quantifying streambed grain sizes and hydro-biogeochemistry using YOLO and photos

**Yunxiang Chen[1], Jie Bao[1], Rongyao Chen[2], Bing Li[1], Yuan Yang[3], Lupita Renteria[1], Dillman Delgado[1], Brieanne Forbes[1], Amy E. Goldman[1], Manasi Simhan[4], Morgan E. Barnes[1], Maggi Lanan[1], Sophia McKever[1], Z. Jason Hou[1], Xingyuan Chen[1], Timothy Scheibe[1], James Stegen[1,5]**

[1]Pacific Northwest National Laboratory, Richland, WA, USA
[2]School of Petroleum Engineering, Yangtze University, Wuhan, Hubei, China
[3]Center for Western Weather and Water Extremes, Scripps Institution of Oceanography, University of California San Diego, CA, USA
[4]Department of Geoscience, University of Wisconsin - Madison, Madison, WI, USA
[5]School of the Environment, Washington State University, Pullman, WA, USA

**Key Points:**

- Stream sediments bigger than 0.45 mm can be detected from smartphone photos by YOLO with a Nash–Sutcliffe efficiency of 0.98.
- Quantities, distributions, and uncertainties of streambed hydro-biogeochemistry can be determined from photos.
- We have identified sources of uncertainty in grain size measurements and proposed approaches to reduce this uncertainty.

Corresponding author: Yunxiang Chen, yunxiang.chen@pnnl.gov

## Abstract

Streambed grain sizes and hydro-biogeochemistry (HBGC) control river functions. However, measuring their quantities, distributions, and uncertainties is challenging due to the diversity and heterogeneity of natural streams. This work presents a photo-driven, artificial intelligence (AI)-enabled, and theory-based workflow for extracting the quantities, distributions, and uncertainties of streambed grain sizes and HBGC parameters from photos. Specifically, we first trained You Only Look Once (YOLO), an object detection AI, using 11,977 grain labels from 36 photos collected from 9 different stream environments. We demonstrated its accuracy with a coefficient of determination of 0.98, a Nash–Sutcliffe efficiency of 0.98, and a mean absolute relative error of 6.65% in predicting the median grain size of 20 testing photos. The AI is then used to extract the grain size distributions and determine their characteristic grain sizes, including the 5th, 50th, and 84th percentiles, for 1,999 photos taken at 66 sites. With these percentiles, the quantities, distributions, and uncertainties of HBGC parameters are further derived using existing empirical formulas and our new uncertainty equations. From the data, the median grain size and HBGC parameters, including Manning's coefficient, Darcy-Weisbach friction factor, interstitial velocity magnitude, and nitrate uptake velocity, are found to follow log-normal, normal, positively skewed, near log-normal, and negatively skewed distributions, respectively. Their most likely values are 6.63 cm, 0.0339 $s \cdot m^{-1/3}$, 0.18, 0.07 m/day, and 1.2 m/day, respectively. While their average uncertainty is 7.33%, 1.85%, 15.65%, 24.06%, and 13.88%, respectively. Major uncertainty sources in grain sizes and their subsequent impact on HBGC are further studied.


## Plain Language Summary

Streambed grain sizes control river hydro-biogeochemical function by modulating the resistance, speed of water exchange, and nutrient transport at water-sediment interface. Consequently, quantifying grain sizes and size-dependent hydro-biogeochemical parameters is critical for predicting river's function. In natural streams, measuring these sizes and parameters, however, is challenging because grain sizes vary from millimeters to a few meters, change from small creeks to big streams, and could be concealed by complex non-grain materials such as water, ice, mud, and grasses. All these factors make size measurements a time-consuming and high-uncertain task. We address these challenges by demonstrating a workflow that combines a computer vision artificial intelligence (AI),

smartphone photos, and new uncertainty quantification theories. The AI performs well across various sizes, locations, and stream environments as indicated by an accuracy metric of 0.98. We apply the AI to extract the grain sizes and their characteristic percentiles for 1,999 photos. These characteristic grain sizes are then input into existing and our new theories to derive the quantities, distributions, and uncertainties of hydro-biogeochemical parameters. The high accuracy of the AI and the success of extracting grain sizes and hydro-biogeochemical parameters demonstrate the potential to advance river science with computer vision AI and mobile devices.

## 1 Introduction

Streambed grain size is a crucial factor controlling streambed hydro-biogeochemistry (HBGC). In hydrology, hydraulics, and geomorphology, streambed flow resistance, which is parameterized by the Manning coefficient or Darcy–Weisbach friction factor, is directly linked to characteristic grain sizes such as the median, 84th, and 90th percentiles of grain size distributions (Strickler, 1923; S. Lang et al., 2004; Chaudhry, 2008; Ferguson, 2010, 2007; Rickenmann & Recking, 2011; Powell, 2014; Ferguson, 2022). In stream-groundwater interactions, the speed of water exchange through the porous sediment interface, quantified as streambed interstitial velocity, is controlled by pressure variation and subsurface permeability, both of which depend on characteristic grain sizes of streambeds (Kenney et al., 1984; Shepherd, 1989; Elliott & Brooks, 1997; Y. Chen et al., 2021). In biogeochemistry, grain sizes exert direct control over turbulent mass transfer that determines the upper limit of the total nitrate uptake velocity from streams by benthic algae, microbes, and turbulence (O'Connor & Hondzo, 2008; Mulholland et al., 2009; Grant et al., 2018). Despite the importance, measuring streambed grain sizes and size-dependent HBGC is challenging due to the multiscale and heterogeneous nature of grain size, the diversity of stream environments, and consequently the high labor costs associated with grain size quantification and HBGC estimation.

Over the past seven decades, large efforts have been made to address the aforementioned challenges. These efforts can be categorized into traditional sieve methods, grid- or area-based sediment counting or weighting methods (Wolman, 1954; Leopold, 1970; Kellerhals & Bray, 1971; Anastasi, 1984; Fehr, 1987; Fripp & Diplas, 1993), manual photo sieving method (Adams, 1979; Ibbeken & Schleyer, 1986), automated or semi-automated photo sieving methods (Butler et al., 2001; Graham et al., 2005; Detert & Weitbrecht, 2012; Purinton & Bookhagen, 2019), image texture statistics methods (Carbonneau et al., 2004; Rubin, 2004; Verdú et al., 2005; Carbonneau et al., 2005a, 2005b; Buscombe & Masselink, 2009; Buscombe et al., 2010; Buscombe & Rubin, 2012; Buscombe, 2013; Black et al., 2014), machine learning (ML) methods (Z. Chen et al., 2020; Soloy et al., 2020; N. Lang et al., 2021; Ermilov et al., 2022), point cloud methods (Vázquez-Tarrío et al., 2017; Steer et al., 2022), and ML-based in-direct grain size regression methods (Gomez-Velez et al., 2015; Ren et al., 2020; Abeshu et al., 2022). The sieve method is the oldest and most reliable approach for fine sediment characterization, however, it is not feasible for field sampling of coarse sediments due to the requirement to transport a large

number of rocks to the laboratory for drying, sieving, and weighing (Leopold, 1970). Although the grid and area based methods avoid the need to move heavy rocks, they suffer from poor reproducibility along with significant time and labor costs, due to the necessity of manually measuring and recording grain sizes in the field (Wohl et al., 1996; Bunte & Abt, 2001).

The manual photo-sieve approach was therefore developed in the late 1970s to circumvent the need for direct measurements of grains in the field, however, it remains time-consuming as it involves manual identification and digitization of grains from images (Graham et al., 2005). Consequently, automated and semi-automated techniques were developed. These approaches are based on a series of image processing algorithms such as converting colored images to grayscale, applying simple or double thresholds, edge detection, bottom-hat transformation, and finally using watershed segmentation or k-means clustering to generate individual grains (Graham et al., 2005; Detert & Weitbrecht, 2012; Purinton & Bookhagen, 2019). These methods significantly reduce the time required to generate reliable grain size distributions, but usually need considerable time to adjust key parameters used in the image processing techniques (Graham et al., 2005; Purinton & Bookhagen, 2019). Instead of directly detecting individual grains, statistical methods approximate key grain size metrics, such as the median size, by relating grain sizes to characteristic quantities of image texture derived from auto-correlation (Rubin, 2004), one-dimensional (1D) and two-dimensional (2D) semi-variance (Carbonneau et al., 2004; Verdú et al., 2005), co-occurrence matrix-derived entropy (Carbonneau et al., 2005b), spectrum decomposition (Buscombe et al., 2010), wavelets (Buscombe & Rubin, 2012; Buscombe, 2013), and their combinations (Buscombe & Masselink, 2009; Black et al., 2014). Among these methods, the spectrum decomposition and the global wavelet approaches are especially important because they provide good estimates for the median size (with root-mean-square relative errors of 9.5% to 16%) and the full grain size distribution without the need for calibration (Buscombe et al., 2010; Buscombe, 2013). Despite these successes, it is worth noting that mean sizes obtained from statistical methods are conceptually similar but different from the sizes obtained from sieve or photo-sieve approaches.

In addition to image processing and statistical methods, machine learning methods implicitly learn the relationship between input images and desired targets using data and neural networks. Examples include learning median size and grain size distribution

(N. Lang et al., 2021), individual grains (Soloy et al., 2020; Z. Chen et al., 2020), and clustered grains (Ermilov et al., 2022) using convolutional neural networks (CNNs), Mask regional CNN (R-CNN) (He et al., 2017), and atrous separable convolution (L.-C. Chen et al., 2018), respectively. The Mask R-CNN is the most similar to the traditional sieve and photo-sieve methods, however, its accuracy, which stands at approximately a 50% detection rate in predicting overlapping rocks, needs further improvement before being deployed for practical applications (Soloy et al., 2020). All of the image-based methods mentioned above use images as input, therefore, the grain sizes are three dimensional (3D)-sediment projected 2D sizes. The point-cloud based grain size characterization is more similar to actual 3D grain sizes (Steer et al., 2022), but obtaining accurate 3D point cloud poses a larger challenge than grain size quantification. There also exist ML-based in-direct methods to estimate grain sizes by learning the relationship between median grain size and large-scale geomorphological and hydrological attributes such as elevation, slope, depth, velocity, etc. (Gomez-Velez et al., 2015; Ren et al., 2020; Abeshu et al., 2022). These estimates, however, are not actual measurements and require careful validation against direct measurements before their use in large-scale models.

In summary, past efforts have tackled challenges related to accuracy, reproducibility, cost, multi scales, and heterogeneity. These methods are expected to yield satisfactory results when applied to streambeds primarily composed of granular sediments, such as sand, cobble, gravel, and boulders (Buscombe, 2013). However, they may encounter challenges in stream riparian zones where non-granular materials like grass, mud, ice, wood, and both static and flowing water overlie granular sediments. New methods that can detect sediments hidden beneath these non-granular and non-sediment objects are needed. Another aspect that is not well resolved by previous efforts is photo resolution estimation. Though photo resolution can be manually measured from reference scales, this process is usually time-consuming when dealing with a large number of images. Therefore, there is a need for fully automated photo resolution estimation method.

Our first goal is to address these needs by developing two ML models, one for grain detection and one for scale detection, using the You Only Look Once (YOLO) version 5 framework (Redmon et al., 2016) with 11,977 and 121 labels of grains and reference scales. The YOLO framework is selected because it is a general, real-time, object detection algorithm (Redmon et al., 2016) with the capability to detect hidden grains covered by non-sediment objects with much higher detection rate, compared to regional CNN

approach (He et al., 2017; Soloy et al., 2020). Our second goal is to estimate streambed hydro-biogeochemical parameters based on YOLO-derived characteristic grain sizes and empirical equations for Manning coefficient (Rickenmann & Recking, 2011), Darcy–Weisbach friction factor (Ferguson, 2007, 2022), streambed interstitial velocity magnitude (Kenney et al., 1984; Y. Chen et al., 2021), and nitrate uptake velocity (Grant et al., 2018). Our third goal is to quantify uncertainties in both characteristic grain sizes and their propagation to the estimated HBGC parameters as well as the dominant sources of uncertainties in grain sizes and HBGC.

To achieve these goals, the paper is organized as follows: Section 2 introduces the study site, photo collection and grouping, training label generation, YOLO framework setup, as well as the equations used for HBGC and uncertainty calculation; Section 3 evaluates the YOLO model accuracy and reports the distributions and uncertainties of grain sizes and HBGC parameters; a thorough discussion covering the accuracy of grain sizes and HBGC, their major sources of uncertainty, the effects of photo number and probability threshold on model accuracy, potential automated photo resolution estimation strategy, as well as the limitations and future directions, is included in Section 4; the major results and implications are summarized in Section 5.

## 2 Methods

### 2.1 Photo acquisition and grouping

We obtained 2,121 photos from 75 sites at the Yakima River Basin (YRB) and the Columbia River section near the Port of Benton (Figure 1d) during 2021 – 2023. In 2021, we collected 383 photos from 47 sites; in 2022, we obtained 1,688 photos across 41 sites; in 2023, we took 50 photos from 3 sites near the Boat Ramp (BR) of the Leslie Groves Park. 6 camera types were used, including Samsung's SM-T500 tablet and Apple's iPhone 7, 12, 13, 13 Pro Max, and 14 Pro.

From these photos, we selected 61 photos as our training (36), validation (5), and testing (20) datasets. These datasets are mutually exclusive and labeled as 0, 1, and 2, respectively, for convenience (Figure 1a). To study the effects of the number of photos on model accuracy, we further divided the 36 training photos into three mutually inclusive groups, each having 11, 21, and 36 photos, respectively. For convenience, models trained with these groups are termed as model M0a, M0b, and M0c, respectively. In addition,

we trained a fourth model for scaling, termed as model Msc, to convert pixel size to real-world size using 50 photos (23 photos are from the 2,121 photos).

The 4 trained AI models were applied to predict both individual grains and reference scales for 2,143 photos. These photos were divided into 7 groups, labeled as 3 to 9, and each had 144, 1855, 24, 20, 21, 21, and 58 photos, respectively. Their roles are described as follows: the photos in group 3 and 4 are used to predict grain sizes of photos obtained in 2021 and 2022 (Figure 1b); the 20 photos in group 6 (same photos as group 2 in Figure 1a) are used to the test the accuracy of model M0a – M0c for predicting grain sizes; the photos in groups 5 (from iPhone 12), 7 (iPhone 13), 8 (iPhone 14 Pro), and 9 (Figure 1c) are used to evaluate the sensitivity of grain sizes and scaling to camera types and height as well as the accuracy of model Msc in predicting scales, respectively. The number of photos taken at each site is visualized in Figure 1d for reference. Details of site coordinates, grain sizes, and photo number can be found from our accompanying data package (Y. Chen et al., 2023).

### 2.2 Label generation

We manually generated labels (see label definition in Section 2.3) for both the grain detection AI models (M0a - M0c) and the scale detection AI model (Msc). For the grain detection models, we manually generated 16,951 labels from 61 photos, resulting in an average of 278 labels per photo (with a minimum of 19 and a maximum of 3,315). Out of these labels, 5,272 were used for training M0a, 10,154 for M0b, and 11,977 for M0c, respectively. For the scale detection model, we generated 121 labels from 50 photos representing 10 types of scales. These photos represent diverse flow, vegetation, and geological conditions in natural streams. 9 photos for the grain detection models and 6 photos for the scale detection model are illustrated in Figure 2 to visualize the environmental conditions and manually-generated labels (green dots bounded boxes). Photos a to i represent the following 9 conditions: dry bed, dry bed with high grain size ratio, dry bed with grass, dry bed with mud, partial-dry partial-wet mud, dry bed with ice, submerged bed with static water, submerged bed with flowing water and waves, hybrid rock/water/grass bed. Photos j to o represent 10 reference scales with known sizes, including, yellow tape 1, yellow tape 2, blue cap, green cap, tape measure, yellow paper board, quadrat net, color tapes, full quadrat, and white paper board. Their sizes are 7.05 cm × 1.7 cm, 7.1 cm × 2 cm, 3.7 cm, 2.5 cm, readable from tape measure, 11 cm × 11 cm, 20 cm × 20

cm, 2.54 cm in width, 80 cm × 80 cm, and 30.48 cm × 22.86 cm, respectively. The rest 52 photos for grain detection AI models and 44 photos for scale detection AI model and their labels can be found in the accompanying data package (Y. Chen et al., 2023).

### 2.3 YOLO framework

You Only Look Once (YOLO) is an object detection AI algorithm that is widely used for computer vision tasks (Redmon et al., 2016). In this study, the fifth major updated version was used and called YOLOv5. The Python implementation of YOLOv5 algorithm was open-sourced in 2020 by Ultralytics on GitHub (Ultralytics, 2020). YOLOv5 is a state-of-the-art real-time object detection system that is faster and more accurate than its predecessors.

A brief sketch of the YOLOv5 network flowchart is shown in Figure 3, which is summarized from GitHub (Ultralytics, 2020). Generally, it is constructed by a series of convolutional layers (Conv in Figure 3) (W. Zhang et al., 1990), modified bottleneck cross stage partial network layers (C3 in Figure 3) (Wang et al., 2020), a spatial pyramid pooling-fast layer (SPPF in Figure 3) (He et al., 2014), concatenate layers (Concat in Figure 3), and up-sampling layers. The fractional numbers, such as 1/2, 1/4, 1/8 and so on, in Figure 3 represent the relative image resolutions to the input image. For the convolutional layers in Figure 3, $Ch_i$, $Ch_o$, $k$, and $s$ stand for input image's number of channels, output image's number of channels, kernel size, and stride size, respectively. For the C3 layer, it reduces the number of convolutional layers from 4 to 3 in bottleneck cross stage partial network, which is originally connected to the output of bottleneck block (Wang et al., 2020). The value $n$ in Figure 3 stands for the number of bottleneck blocks in C3 layer. The spatial pyramid pooling-fast layer is a modified spatial pyramid pooling layer specifically designed for YOLOv5 with higher computational efficiency (Ultralytics, 2020). It concatenates several MaxPool layers (PyTorch, 2022) with different sizes for resolving the difficulties of detecting objects with various sizes.

The final outputs of YOLO, also called as labels, are the centroid ($x$ and $y$ in Figure 3), width ($w$ in Figure 3), height ($h$ in Figure 3), and class ($c$ in Figure 3) of the anchor box and the probability of the detected object in each class. The centroid and sizes of the anchor box are all normalized by the dimension of the original input image. In this study, we have 10 classes for reference scales (Section 2.2) and only one class for grain.

The network input is the image and the outputs are the corresponding labels. To avoid over-fitting, 5 labeled images were used for validation. During training, the optimizer does not consider the loss between the prediction of the validation images and true labels. The loss for the validation images is only used as a training termination criterion. With the predicted width and height of individual grains, we define the diagonal length of the grain, i.e., $D_p = \sqrt{w^2 + h^2}$, as the final grain size in pixel length, which can be converted to real size ($D$) by multiplying it with the estimated photo resolution.

### 2.4 Streambed hydro-biogeochemistry estimation equations

With given water depth ($H$) and flow velocity ($U$) as well as the photo-derived characteristic grain sizes, e.g., 5th ($D_5$), 50th ($D_{50}$), and 84th ($D_{84}$) percentiles of grain size distributions, key streambed hydro-biogeochemical parameters, including Manning's coefficient ($n$), Darcy–Weisbach friction factor ($f$), shear velocity ($u_\tau$), streambed interstitial velocity magnitude ($\sigma_w$), and streambed nitrate uptake velocity ($u_f$) can be estimated by Equations 1 (Rickenmann & Recking, 2011), 2 (Ferguson, 2007, 2022), 3 (Y. Chen et al., 2021; Kenney et al., 1984), and 4 (Grant et al., 2018), respectively. The water depth is a reach average depth, which was estimated using a wading-based depth transect procedure. The details of such a procedure can be found in the field protocol described in our data package published in US DOE's Environmental System Science Data Infrastructure for a Virtual Ecosystem (ESS-DIVE) (Delgado et al., 2023). The velocity is the average velocity for August between February 1979 and December 2020, which was computed by Kaufman et al. (2023) from the National Oceanic and Atmospheric Administration's National Water Model version 2.1 (NOAA, 2023).

$$n = \frac{D_{84}^{1/6}}{20.4} \tag{1}$$

$$\sqrt{\frac{8}{f}} = \frac{U}{u_\tau} = \frac{c_1 c_2 H/D_{84}}{\sqrt{c_1^2 + c_2^2 (H/D_{84})^{5/3}}} \tag{2}$$

$$\sigma_w = c_3 \frac{g k_I}{\nu} \frac{U^2}{g D_{50}} (\frac{H}{D_{50}})^{c_4}, k_I = c_5 D_5^2 \tag{3}$$

$$u_f = k_m \phi, k_m = 0.17 u_\tau Sc^{-2/3}, Sc = \frac{\nu}{D_m}, \phi = c_6 [\mathrm{NO_3^-}]^{c_7} \tag{4}$$

The constants used in the above equations are: $c_1$ = 6.5, $c_2$ = 2.5 (Ferguson, 2022); $c_3$ = 0.88 (range 0.62 – 1.11), $c_4$ = -0.66 (Y. Chen et al., 2021); $c_5 = 1{\times}10^{-9}$ (Kenney et al., 1984); $c_6$ = 0.0032, $c_7$ = -0.49 (Grant et al., 2018); gravity acceleration $g$ = 9.81 m/s$^2$, water viscosity $\nu = 1 \times 10^{-6}$ m$^2$/s, nitrate molecular diffusion in water $D_m$ =

$1.7\times10^{-9}$ $m^2/s$ (Picioreanu et al., 1997). Non-constant parameters include subsurface intrinsic permeability $k_I$ ($m^2$), hydrogeology-biochemistry interaction efficiency $\phi$, Schmidt number $Sc$, and stream nitrate concentration $[NO_3^-]$ ($mol/m^3$, equivalent to 62 mg/L). Our field survey in 2021 shows that the nitrate concentration in YRB varies between 0.0005 and 0.1 with a mean of 0.008 $mol/m^3$ (Grieger et al., 2022). In 2022, stream nitrate concentrations are not available for all locations where depth were measured, therefore, we select three values, 0.0001, 0.01, and 1 $mol/m^3$, to represent the typical magnitudes reported at the YRB and in the literature (Mulholland et al., 2008; Grant et al., 2018; X. Zhang et al., 2021; Sadayappan et al., 2022).

### 2.5 Uncertainty quantification for grain sizes and hydro-biogeochemistry

Uncertainties occur in grain detection, scaling, and the propagation from grain sizes to hydro-biogeochemical parameter estimations. For any given photo, the real grain size $D_x$ ($x$ = 5, 50, and 84) are calculated by $D_x = D_{xp}SC$ with the $D_{xp}$ and $SC$ denoting the grain size measured by pixel number and the photo resolution measured by real size per pixel. The $D_{xp}$ is determined by YOLO and its uncertainty $r_{xp}$, quantified by the average absolute relative error of testing photos, can be directly estimated by comparing the YOLO-predicted and manually measured grain sizes. For photo-resolution uncertainty, we manually draw two straight lines for all photos following the scales showing in Figure 2 and then calculate the relative error ($r_{SC}$) between the photo resolution calculated from the two lines. With the estimation of pixel-based grain size uncertainty and scale uncertainty, the real-world grain size uncertainty and its propagation to HBGC parameters can be estimated by Equations 5 – 9 based on the law of propagation of uncertainty (Ku, 1966). The detailed mathematical derivation of these equations can be found in Appendix. The $r_H$ is the mean absolute relative difference between the measured water depth ($H$) and its time-average value over the observation period (around 1 month in August 2022). The uncertainty measurement for flow velocity ($r_U$) and stream nitrate concentration ($r_N$) are not available for the study sites. However, existing literature report that velocity measurement uncertainty by Acoustic Doppler current profilers (ADCPs) could range 1% to 25% depending on the distance away from the ADCPs (Mueller et al., 2007) and stream nitrate concentration uncertainty is 12% on average across 7 watersheds in US (Jiang et al., 2014). Therefore, we choose 10% as a rough

estimation of the typical measurement uncertainty for stream velocity and nitrate concentration in this work.

$$r_x = \sqrt{{r_{xp}}^2 + {r_{SC}}^2}, x = 5, 50, 84 \tag{5}$$

$$r_n = r_{84}/6 \tag{6}$$

$$r_f = 2\{1 - \frac{5}{6}[1 + \frac{c_1^2}{c_2^2}(\frac{H}{D_{84}})^{-5/3}]^{-1}\}\sqrt{r_H^2 + r_{84}^2} \tag{7}$$

$$r_w = \sqrt{4r_U^2 + (1 - c_4)^2 r_{50}^2 + c_4^2 r_H^2} \tag{8}$$

$$r_{uf} = \sqrt{r_U^2 + r_f^2/4 + c_7^2 r_N^2} \tag{9}$$

## 3 Results

### 3.1 YOLO performance

We evaluate the performance of YOLO through four metrics: the mean average precision (mAP) of the YOLO training, the accuracy of grain size distribution, median grain sizes, and their relative error (Figure 4). The mAP@50 and mAP@50-95 are two typical metrics used to quantify the accuracy of object detection AI algorithm. The symbol @50 means the prediction is correct if the intersection over union (IoU) larger than 50%. The IoU stands for the relative overlapping area between the predicted object bounding box and the ground truth object bounding box. Similarly, the symbol @50-95 means the prediction is correct if the IoU larger than 50% to 95% with 5% increase interval. Additional 5 photos with 954 labeled grains are used as validation data set. The accuracy of the prediction on the 5 validation photos are not seen by the optimizer, and it is only used to track the model accuracy during training and helps on determination of the best model, as shown in Figure 4(a). The weighted mAP (10% of mAP@50 and 90% of mAP@50-95) is used as final accuracy metric, and it reaches the maximum at 968 steps (Figure 4a: vertical dashed line). The corresponding mAP@50 and mAP@50-95 at this step is 0.64 and 0.34, respectively (Figure 4a: horizontal dashed lines). After 968 training steps, both mAP@50 and mAP@50-95 decrease, with no indication that the accuracy can increase within 20,000 training steps. Therefore, the trained model, which is used for all the results in the study, is the model stored at 968 training steps. For Microsoft Common Objects in Context (COCO) dataset, a commonly used benchmark dataset for object detection AI, typical values for mAP@50 and mAP@50-95 fall in the range

0.46 – 0.73 and 0.28 – 0.56, respectively (Ultralytics, 2020). In our case, the shape, sizes, color, transparency, lighting, and environmental conditions are more complex than those photos used in COCO (Figure 2), however, the model still achieves 0.64 and 0.34 values for mAP@50 and mAP@50-95 on the validation photos, respectively (Figure 4a). This means the YOLO training achieved a good performance.

To illustrate the model's capability in extracting grain size distributions (GSDs), Figure 4b shows a comparison of the area-weighted GSD between the model prediction (blue line) and manual labels (red line). The cumulative probability in calculated by $P_i = \sum A_i(D \leq D_i)/\sum A_i$ with $A_i$ and $D_i$ denoting the area and size of each grain. The minimum difference between the two lines demonstrates that the area-weighted GSD is accurately reproduced by the trained model. Similar comparisons for the remaining 19 photos used for testing are not included here for simplicity, however, can be found in Figure 12. These comparisons demonstrate that the GSDs can be well reproduced by YOLO algorithms for most (18 of 20) photos.

Based on the GSD curves, the median grain size D50, defined as the grain size corresponding to 50% finer grain sizes, can be calculated from the GSDs of the 20 testing photos. Figure 4c shows a one-to-one plot between the predicted D50 and manually estimated D50. The result shows that YOLO predicts D50 with an accuracy of 0.98, 0.98, -0.037 cm, and 0.91 cm in terms of R-squared, Nash–Sutcliffe efficiency (NSE), mean error, and root-mean-square between the prediction and manual measurements. To further examine such accuracy, Figure 4d shows the relative error between the predicted D50 and manually estimated D50. The result shows 90% (18 dots) of the data points demonstrate a relative error less than 10% and 10% (2 dots) show a relative error larger than 20%. On average, the mean absolute relative error is 6.65% for the 20 testing photos. The result also shows the relative error does not correlate with the grain size, which suggests the accuracy of YOLO is stable for both small and large grains.

### 3.2 Characteristic grain size distributions

With the confirmed high accuracy of the YOLO model, we apply the model to extract the grain size distributions (GSDs) from 1,999 photos (66 sites) in groups 3 and 4, and then calculate the characteristic grain sizes, e.g., D5, D50, and D84, from the GSDs. As valid water depth measurements are available at only 41 sites, Figure 5 shows only

the results of characteristic grain sizes from 1,745 photos obtained at the 41 sites to make a consistent evaluation for HBGC parameters in Section 3.3. In general, the three grain size distributions follow log-normal distributions (black solid lines in Figure 5a-c are fitted Gaussian distributions) with the log2-transformed mean of 4.15, 6.05, 6.75 and standard deviation of 0.86, 0.87, and 0.81 for D5, D50, and D84, respectively. This means the most likely sizes of D5, D50, and D84 are around 1.78 cm, 6.63 cm, and 10.76 cm, respectively. As D5, D50, and D84 represent different importance of grain sizes in controlling HBGC, Figure 5d further shows the relationship between D5 and D50 and that between D84 and D50. The result shows that D5 and D84 increase linearly with D50, although there are some large residuals.

### 3.3 Streambed hydro-biogeochemistry distributions

With the photo-derived characteristic grain sizes (D5, D50, and D84), measured water depth, extracted velocity, and assumed typical stream nitrate concentration (see details in Section 2.4), the HBGC parameters can be estimated using Equations 1 - 4. To mitigate the uncertainty resulting from an insufficient number of photos, we show results only from sites with more than 3 photos. Consequently, we are showing the results from 1,737 photos at 37 sites (refer to site locations in Figure 6b).

Overall, HBGC parameters demonstrate different distribution patterns compared to grain sizes. Specifically, the Manning coefficient follows a normal distribution (black line in Figure 6a) with a mean and standard deviation of 0.0339 and 0.0031 $s \cdot m^{-1/3}$, respectively. The log10-transformed friction factor, log10(f), shows a positively skewed distribution (Figure 6c) with its skewness (defined as the adjusted Fisher-Pearson skewness coefficient), mean, median, mode, and standard deviation of 0.43, -0.54, -0.58, -0.75, and 0.37, respectively. This suggests the friction factor has the most likely value of 0.18 ($=10^{-0.75}$), which falls in the range of 0.13 – 0.32 calculated from high-resolution computational fluid dynamics simulations for natural gravel bed rivers with median grain size of 6 cm (Y. Chen et al., 2019). The log10-transformed streambed interstitial velocity magnitude, log10($\sigma_w$), follows a near-Gaussian distribution (Figure 6e) with skewness, mean, median, mode, and standard deviation of -0.03, -1.07, -1.08, -1.15, and 0.52, respectively. This suggests the streambed interstitial velocity magnitude has a high likelihood at the scale of 0.07 ($=10^{-1.15}$) m/day for the study region, which is close to the value (0.11 m/day) estimated by a temperature-based data assimilation approach applied at the Hanford reach of the

Columbia River (K. Chen et al., 2023). The distribution of the nitrate uptake velocity is more complex. Firstly, the distribution is strongly affected by the concentration of stream nitrate. It may decrease 3 orders of magnitude if the nitrate concentration increases from 1e-4 mmol/L (=0.0062 mg/L) (Figure 6g blue histogram) to 1 mmol/L (=62 mg/L) (Figure 6g gold histogram). The median and mean values of stream nitrate concentration were reported at the order of 1e-2 mmol/L (=0.62 mg/L) over 72 agriculture and urban sites in US (Grant et al., 2018). The mean nitrate concentration in the YRB was also reported at a similar magnitude of 0.008 mmol/L (Grieger et al., 2022). Therefore, it is reasonable to use 0.01 mmol/L as the most likely magnitude of nitrate concentration in US. Using such a concentration, the nitrate uptake velocity varies between 0.23 and 5.6 m/day and shows a negatively skewed distribution with the skewness, mean, median, mode, and standard deviation of -0.23, 0.013, 0.036, 0.075, and 0.22, respectively (Figure 6g gold histogram). This means the nitrate uptake velocity has a high chance to be 1.2 ($=10^{0.075}$) m/day with a US median or mean nitrate conditions. This value is in the range between measured median (0.6 m/day) and mean (2.5 m/day) uptake velocity across the US (Grant et al., 2018).

The left panels of Figure 6 illustrate the overall distributions of HBGC parameters but not their spatial variations. To visualize the spatial variations, the right panels show the spatial distributions of site average HBGC parameters. The number of photos at each site can be found on Figure 1d. Figure 6b shows that the site average Manning coefficient mostly clusters at red (0.035 - 0.0375 $s{\cdot}m^{-1/3}$) and light red (0.0325 - 0.035 $s{\cdot}m^{-1/3}$), which means the site average Manning coefficient has a low spatial heterogeneity. Such a behavior can also be observed in Figure 8a where the site average value (black line) of Manning coefficient shows small variation across the sites. In contrast, the site-average friction factor exhibits greater heterogeneity, as indicated by the diverse range of colors in Figure 6d. The highest log10-transformed friction factor values (0 – 0.25) occur at site S37, S39, and W10, followed by 8 sites (W20, S04, S03, S42, S10, S53, S56N, and S48R) in the group -0.25 – 0. The lowest values (yellow dots at group -1 – -0.75) occur at S02, T02, T03, and S23, and the rest of the data points share similar colors. This behavior can also be observed in Figure 8b (see black line). Different from the friction factor, the log10-transformed interstitial velocity magnitude has maximum values at sites S04, S58, S18R, T05P, S50P, and S56N (Figures 6f dark red and 8c black line), followed by the value group -0.75 – -0.25 (red) at 5 sites (S48R, S10, S01, W10, and S31). The

lowest interstitial velocity occurs at the sites S42 and S43 with a value of around -2 (Figures 6f yellow and 8c black line). Compared to the friction factor and interstitial velocity, the uptake velocity distribution demonstrates obvious hot spot at site S04 (dark red) and cold spots (yellow) at sites T02, S41R, S42, and S43 with a value of 2.8 m/day and a range of 0.3 – 0.5 m/day, respectively. Interestingly, the cold spots are all within or downstream of the Yakama Indian Reservation region. It is also interesting to mention that the hot (S04) and cold (S42 and S43) spots in nitrate uptake velocity are also the hot and cold spots in the interstitial velocity. This suggests the hot/cold spots in denitrification are likely affected by the water exchange between stream and groundwater in the YRB. This is consistent with the work of Son et al. (2022) that shows hyporheic exchange flux is the most important factor controlling nitrate removal based on data from basin-scale numerical simulations and random forest relative importance analyses.

### 3.4 Uncertainty in characteristic grain sizes

With the uncertainty quantification equations introduced in Section 2.5, the uncertainty or variability associated with manually-measured photo resolution, YOLO-derived grain sizes, and water depth observations can be estimated for each photo. Figure 7a shows the manually-measured photo resolution (blue cross) and the relative error $r_{SC}$ (yellow line) associated with each resolution. The results shows that around 90% of the photos have a resolution of around 0.1 mm/pixel (corresponding to 1/4 of the quadrat in Figure 2n, o), and 10% of the photos have a resolution between 0.2 and 0.7 mm/pixel (corresponding to the full quadrat in Figure 2n, o). The relative error for these scales, however, are mostly in the range -10% – 10% and have an overall mean and mean absolute error of 0.13% and 2.3%, respectively. This means the photo resolution estimation has no systematic bias and the manual measurement uncertainty is low enough for further grain size quantification.

With the photo resolution uncertainty ($r_{SC}$), the uncertainty in D50, D84, and D5 can be calculated by Equation 5 with the YOLO-associated grain size uncertainty $r_{50p}$ (=6.65%), $r_{84p}$ (=10.65%), and $r_{5p}$ (=11.88%) directly estimated from the average absolute relative error of testing photos as discussed in Section 3.1. Figures 7b,c,d show the combined effects of photo resolution uncertainty and YOLO accuracy uncertainty for D50, D84, and D5, respectively. The result shows the uncertainty of D50 varies between 6.65% and 13.53% with a mean value of 7.33%. For D84 uncertainty, its minimum,

maximum, and mean are 10.65%, 15.88%, and 11.11%, respectively. For D5 uncertainty, these values are 11.88%, 16.73%, and 12.30%, respectively.

The water depth is estimated every 1 minute during July 28 and August 31 2022 (see details in data package (Delgado et al., 2023)). With these data, the depth ($H$) is calculated as the time averaged depth over the whole measurement period. The uncertainty or variability ($r_H$) of such a depth is calculated as the average absolute relative difference between the actual depth and the calculated mean depth. Figure 7e shows the variations of the mean depth and its variability at each site. The result shows the depth varies between 0.14 m and 2.11 m, with a mean of 0.45 m across all the sites. Highest depth occurs at sites T02 and T03 while depth less than 0.25 m are found at 9 sites (S63, S53, S04, S37, S39, S03, W10, W20, and S42). The depth variability varies between 0.66% and 30.2% with a mean 6.6%. High depth uncertainty is observed at sites S56N, S24, and S18R.

### 3.5 Uncertainty in hydro-biogeochemistry

With the quantification of uncertainties for grain sizes, depth, and assumed typical measurement uncertainty in velocity and nitrate concentration (see details in Section 2.5), Figure 8 shows all calculated values (blue cross dots), site-average values (black lines), and estimated uncertainty (yellow lines) for Manning's $n$, friction factor $f$, streambed interstitial velocity magnitude $\sigma_w$, and streambed nitrate uptake velocity $u_f$. It is observed that the Manning coefficient varies in a range 0.0245 – 0.0455 $\mathrm{s{\cdot}m^{-1/3}}$ with low uncertainty range of 1.78% – 2.61% (Figure 8a). The friction factor, by contrast, spans over 2 order of magnitude (0.04 – 9) and its uncertainty has minimum, maximum, and average of 3.63%, 58.36%, and 15.65%, respectively. The highest uncertainty occurs at site S56N (Figure 8b yellow line). The interstitial velocity magnitude spans even larger ranges from 0.0038 to 2.31 m/day. However, its uncertainty range is lower than the friction factor, which has minimum, maximum, and average of 22.84%, 32.11%, and 24.06%, respectively. The highest uncertainty is observed at site S56N (Figure 8c yellow line). The nitrate uptake velocity shows a lower variation range between 0.23 and 5.6 m/day. The highest uptake velocity occurs at site S04 while the lowest values occur at sites S42 and S43 (Figure 8d black line). The highest uncertainty occurs at site 56N (Figure 8d yellow line), which is similar to those observed for friction factor and interstitial velocity magnitude. Overall, the uptake velocity uncertainty is estimated as 11.28%, 31.23%,

and 13.88% in terms of the minimum, maximum, and average value. It is worth noting that the results for uptake velocity are based on US mean nitrate concentration (0.01 mmol/L). Therefore, the uptake velocity variation range will change with nitrate concentration at other sites, however, its uncertainty may be similar if the depth and grain size conditions are similar.

# 4 Discussion

## 4.1 Accuracy of grain sizes and hydro-biogeochemistry parameters

To apply the present approach to other rivers, it is important to evaluate the accuracy of the YOLO-derived grain sizes and grain size-based HBGC estimations. As percentile-based grain sizes are derived from the grain size distribution (GSD) curve, the accuracy of GSD determines the accuracy of characteristic grain sizes, e.g., D50, D84, and D5. As demonstrated in Figure 4b and Figure 12, the pre-trained YOLO can reproduce the GSDs with high accuracy for 90% (18 out of 20) of the testing photos that represent 9 different streamed conditions. Under these diverse conditions, the median grain sizes calculated from these GSDs demonstrate relative errors less than 10% (Figure 4d). These results indicate that GSDs and subsequently derived characteristic grain sizes are accurate, at least, for the majority (90%) of the photos. Even though two (10%) testing photos (Figure 12(f,r)) show larger error in GSD, the overall accuracy of all the testing photos, as indicated by an R2 value of 0.98, an NSE value of 0.98, and a mean absolute relative error of 6.65%, is still suitable for practical applications. A closer examination of the two photos (Figure 12(f,r)) with higher error shows that the error is likely caused by the unclear boundaries between the largest grains and ambient smaller sediments, due to light reflection and flocculation on wet grain surface and water surface. Future work may be needed to address these challenges to further improve grain size accuracy.

With the YOLO-derived characteristic grain sizes, using the equations introduced in Section 2.4 to estimate the streambed HBGC parameters will undoubtedly bring errors, partially from the limitation of the equations themselves, and partially from the propagation of uncertainties in input parameters. Though it is challenging to measure HBGC at all study sites, we are able to identify measured or calibrated data for HBGC from existing literature, and can evaluate the accuracy of the photo-driven, AI-enabled, and theory-based estimations for HBGC. Firstly, the well-calibrated Manning's coeffi-

cients from a two-dimensional hydraulic model for the Columbia River vary between 0.027 – 0.038 s/m$^{1/3}$ (Niehus et al., 2014), which is close to the range calculated from all photos (Figure 6a: 0.0245 – 0.0455 s/m$^{1/3}$) and site average value (Figure 6b: 0.0281 – 0.0373 s/m$^{1/3}$). Secondly, the flow resistance from 2,890 field measurements vary between 0.02 and 200 for rivers with $H/D_{84} < 200$ (Rickenmann & Recking, 2011), which covers the range derived from all photos (Figure 6c: 0.04 – 9) and site-average values (Figure 6d: 0.06 – 1.5). Meanwhile, the maximum likelihood of friction factor occurs at 0.18 (=$10^{-0.75}$) (Figure 6c), which falls in the range of 0.13 – 0.32 computed from high-resolution computational fluid dynamics simulations for natural gravel bed rivers with a median grain size of 6 cm (Y. Chen et al., 2019), a value very close to the most likely median size (6.63 cm) observed in our study area (Section 3.2). Regarding the interstitial velocity, direct field measurements are rare. However, by using a temperature-based data assimilation approach, K. Chen et al. (2023) were able to estimate the time series of vertical hydrological exchange flux at the Hanford Reach of the Columbia River. Using their data (Figure S5a in K. Chen et al. (2023)), the interstitial velocity magnitude is estimated as 0.11 m/day by calculating the ratio of the standard deviation of estimated hydrological exchange flux time series to the subsurface porosity (0.43) reported in their work. As demonstrated in Section 3.3, the most likely value of interstitial velocity is around 0.07 m/day (Figure 6e). This suggests most of the estimated interstitial velocity magnitude falls in the observation range. For the streambed nitrate uptake velocity, if the stream nitrate concentration is at the US mean or median level, i.e., 0.01 mmol/L (Grant et al., 2018), the estimated uptake velocity is most likely at the scale of 1.2 m/day, which is between the median (0.6 m/day) and mean (2.5 m/day) uptake velocity measured at 72 sites in US (Grant et al., 2018). The above comparisons, therefore, suggest that photos can be used to make reasonable estimates of HBGC parameters, using AI and empirical equations.

### 4.2 Major sources of uncertainty

Though Section 4.1 demonstrates the accuracy of estimating grain sizes and HBGC, it is still important to quantify potential uncertainties in these estimations. This is necessary to reduce measurement uncertainties in field work and evaluate their impacts on large-scale watershed models. With the use of explicit mathematical formulas, the uncertainties in grain sizes and HBGC can be mathematically accurately derived as shown

in Equations 5 - 9. From these equations, we can see that the uncertainty of YOLO model ($r_{xp}$) and photo resolution ($r_{SC}$) are propagated to the characteristic grain sizes ($r_x$). As demonstrated in Section 3.4, the overall uncertainty for YOLO model is 6.65%, 10.65%, and 11.88% in predicting D50, D84, and D5 pixel sizes, while that for photo resolution is 2.32%. Therefore, the average compounding uncertainty (based on Equation 5) in D50, D84, and D5 are 7.33%, 11.11%, and 12.30%, respectively. Such grain size uncertainties are further propagated to Manning coefficient through $r_n = r_{84}/6$, which results in low uncertainty (mean value 1.85%) in estimating Manning coefficient. The uncertainty in friction factor is more complex because it depends on not only input parameter uncertainty (depth uncertainty $r_H$ and grain size uncertainty $r_{84}$), but also the ratio of water depth to grain size. Despite such complexity, its uncertainty should vary between 1/3 to 2 times of the compounding uncertainty of water depth and D84 ($r_{H_{D84}}$) because the depth/grain size dependent term reduces to 1/3 and 2 for very deep ($H \gg D_{84}$) and shallow water ($H \ll D_{84}$). As the average uncertainty in depth and D84 are 6.6% (Section 3.4) and 11.11%, respectively, their compounding uncertainty is 12.92% ($=\sqrt{r_H^2 + r_{84}^2}$. Therefore, the overall uncertainty of friction factor should vary between 4.31% and 25.85%, which agrees with the average friction factor uncertainty of 15.65% as mentioned in Section 3.5. The uncertainty in interstitial velocity magnitude is simpler because it only depends on the uncertainties of three input parameters: velocity, grain size, and depth. In this work, as the velocity uncertainty is not available, we assume an uncertainty level of 10% based on previous work on velocity measurements with ADCPs (Mueller et al., 2007). As the overall uncertainty in grain size D50 and depth are 7.33% and 6.6%, the overall compounding uncertainty from the three input parameters is around 23.81% (computed from Equation 8) which is close to the average uncertainty (24.06%) calculated from Figure 8c (see Section 3.5).

The uncertainty in nitrate uptake velocity is much more complex because it depends on the uncertainty in velocity, nitrate, and the friction factor that further depends on the values and uncertainties in depth and grain sizes. Such complexity can be verified by Figure 8d where large changes in uptake velocity uncertainty (yellow line) are observed. As the mean uncertainty in friction factor can be estimated by $r_f^m = c_0\sqrt{r_H^2 + r_{84}^2}$ with $c_0$ in the range 1/3 – 2, the mean uncertainty in uptake velocity ($r_{uf}^m$) can be estimated by Equation 10. As the measured nitrate uptake uncertainty is not available, a 10% uncertainty is assumed based on previous work on nitrate measurement uncertainty (Jiang

et al., 2014). With the overall uncertainty for velocity (10%), depth (6.6%), pixel D84 (10.65%), photo resolution (2.32%), and nitrate (10%), the overall uptake velocity uncertainty should fall in the range of $r_{uf}^{md}$ and $r_{uf}^{ms}$ with $r_{uf}^{md}$ and $r_{uf}^{ms}$ representing the mean characteristic uncertainty in deep and shallow rivers. Here the two terms are calculated by $r_{uf}^{md} = r_{uf}^{m}(c_0 = 1/3)$ and $r_{uf}^{ms} = r_{uf}^{m}(c_0 = 2)$ and their values are 11.34% and 16.92%, respectively. As mentioned in Section 3.5, the average uncertainty in uptake velocity calculated from Figure 8d (yellow line) is 13.88%, which falls in the range of characteristic uncertainty. Therefore, the Equation 10 can be used as a fast estimate of the uncertainty in uptake velocity if the uncertainty of 5 inputs are available.

Equation 10 also suggests that the final uncertainty depends on whether the constant $c_0$ leans to the upper bound (2) or the lower bound (1/3), which is mainly determined by the ratio of water depth to grain size D84. In shallow water ($c_0 = 2$) condition, the dominant sources of uncertainties will be velocity, depth, and YOLO-accuracy for D84 because $c_0^2/4 = 1$ and $c_7^2 \approx 0.24$. In deep water ($c_0 = 1/3$), the main sources will be velocity and nitrate concentration because $c_0^2/4 \approx 0.03$. Another important aspect of such an equation is that the uncertainties in velocity, depth, and nitrate concentration represent clear physical meaning, while the uncertainties in pixel D84 and photo resolution are instead associated with AI model and photo induced uncertainties. With further improvements of AI training and photo resolution estimation, these nonphysical uncertainties can likely be reduced to a negligible level (see details in Sections 4.3 - 4.5), and Equation 10 can be reduced to Equation 11 that represents physics-driven uncertainty for uptake velocity. Furthermore, in very dynamic unsteady processes, the uncertainty terms, $r_U$, $r_H$, and $r_N$, more represent the deviation of the actual physical processes away from their time average values, therefore, the compounding uncertainty in Equation 11 can be treated as a metric to quantify the magnitude of the dynamics in nitrate uptake processes.

$$r_{uf}^{m} = \sqrt{r_U^2 + \frac{c_0^2}{4}(r_H^2 + r_{84p}^2 + r_{SC}^2) + c_7^2 r_N^2} \tag{10}$$

$$r_{uf}^{mp} = \sqrt{r_U^2 + \frac{c_0^2}{4} r_H^2 + c_7^2 r_N^2} \tag{11}$$

### 4.3 Effects of photo number

To minimize non-physical uncertainties from the AI model, one way is to increase the number of training photos and labels. Figure 9 shows the effects of photo number on AI-training convergence and accuracy in predicting grain size distribution and characteristic size such as D50. Here the M0a, M0b, and M0c represent three models trained with 11 (5,272 labels), 21 (10,154 labels), and 36 (11,977 labels) photos (see photo locations in Figure 1a and label preparation in Section 2.2). The results show that increasing the number of photos improves the accuracy of the YOLO model, with mAP@50 increasing from 0.54 to 0.64 and mAP@50-95 increasing from 0.28 to 0.34.

Though the model metrics are improved, their accuracy improvements in predicting grain size distributions and D50 depend on the complexity of the streambed. For the dry bed with large grain size ratio (Figure 9b), all three models provide accurate prediction of the GSD though the M0c model (blue line) performs better in capturing smaller grains (<50% percentage finer) and M0a model (black line) performs better in capturing larger grains (>50% percentage finer) when compared to manual measurements (dashed red line). For submerged bed with static water (Figure 9c), the M0c model outperforms M0a and M0b for most of the sizes (<80% percentage finer).

A systematic evaluation of the model accuracy is illustrated in Figure 9d-e in terms of the 1:1 plot between the model-predicted and measured D50 as well as the relative error of predicted D50. The result shows that the M0c model outperforms M0a and M0b in terms of higher R2 (0.98 vs 0.92) and closer alignment with the 1:1 line for all the points (Figure 9d). The closer alignment of model M0c can also be verified in Figure 9e where we can observe 18 points (black circle dots) in the range ± 10% for M0c while those for M0a and M0b are 13 points despite including the points outside but close to the range. The mean absolute relative error for M0a, M0b, and M0c, with values of 11.88%, 11.20%, and 6.65%, also point to the much better performance in M0c.

With available manual labels, it is straightforward to evaluate the model's accuracy. However, it is impractical to manually draw grain sizes for all 1,999 photos used in groups 3 and 4 for prediction purpose (see Section 2.1). Nevertheless, we can evaluate the differences in predicted D50 between the higher accuracy model M0c and the lower quality models as shown in Figure 9f. Statistically, the bias and root-mean-square between M0a and M0c are -0.26 and 2.85 cm; and that between M0b and M0c are -1.22

and 3.14 cm, respectively. As the most likely D50 is 6.63 cm (obtained from M0c model; Section 3.2) and 47% (821 out of 1743 points) of the grain sizes are less than such a value, the uncertainty induced by lower quality models is likely important. Therefore, it is critical to train the YOLO with sufficient data in order to avoid systematic impacts on grain size quantification and subsequent HBGC estimation. In the context of grain size prediction, the number of sufficient data may be determined by checking if the mean absolute relative error between the model prediction and testing labels becomes smaller or comparable to typical uncertainties in field observations or other manual approaches.

### 4.4 Effects of YOLO probability threshold

Another factor that affects the YOLO accuracy is the selection of the probability threshold built in YOLO. A probability threshold is required because the YOLO uses a probability, in the range $0 - 1$, to determine whether an object (grain, grass, water, etc.) in a photo is the target object (e.g., grain in this work). Under-estimation (small value) of the threshold will select too many objects that are not the target, but over-estimation (high value) will ignore objects that are desired. To identify a proper way of selecting the threshold, Figure 10 shows the variation of R2, mean error (ME), mean absolute error (MAE), and the average detected grain number per photo between the prediction (from model M0c) and manual labels, with respect to probability threshold. The best probability threshold should maximize R2, minimize ME and MAE, and identify the number of grain sizes closest to manual measurements. Following these rules, 0.35 is selected as the final probability threshold because R2 reaches maximum (Figure 10a), ME is nearest 0, MAE is at its minimum (Figure 10b), and the number of grains per photo is closest to the manually measured number (Figure 10c). Grains with a YOLO probability less than 0.35 are excluded from the grain size quantification. It is worth mentioning that selecting the probability threshold is a well constrained problem because simultaneously minimizing the ME and identifying the closest number of grains will likely lead to a unique value.

### 4.5 Estimation of photo resolution

How to properly estimate the photo resolution affects not only the accuracy of grain sizes, HBGC parameters, and their compounding uncertainties, but also the efficiency of data collection and post-processing. In general, photo resolution could be estimated

manually or automatically. The manual approach is easy for field implementation, but prone to human error and high data processing costs. In this work, we brought full quadrats and white boards with known sizes into the field, placed them on top of the grains, took photos, manually measured the pixel length of the known scales, and finally obtained the photo resolution, represented by millimeter per pixel (Figure 11a). The manual scale measurement process for 2,121 photos involves 8 person and costs around 200 hours of human labor. Large errors occur due to the unevenness of the quadrats/boards, inaccurate recording of the pixel coordinates from the computer screen, and matching the coordinates to incorrect photo names. To mitigate such errors and reduce costs, an automated scaling approach is desired. Figure 11 illustrates how an automated scaling could be implemented and whether such approaches could be comparable to the manual approach in terms of the resolution and minimum detectable sizes.

It is observed from Figure 11a that the photo resolution clusters at two ranges, i.e., 0.066 – 0.15 and 0.3 – 0.7 mm/pixel (see scale for each photo on Figure 7a and discussion in Section 3.4) and the detectable minimum grain sizes from all photos in groups 3 and 4 vary between 0.82 mm and 21 mm. The typical reference scales for the higher (red star) and lower (blue diamond) photo resolution are visualized in Appendix Figure 13(a,b), respectively. From these figures, we can see that the pixel lengths of the quadrat (white pipes) and strings (red lines) are skewed, which brings errors to resolution estimation and difficulties in manual measurements.

To expedite the photo resolution estimation, a potential way is to train a scale AI model, e.g., model Msc (see details in Sections 2.1 – 2.2), and then use it to measure the pixel sizes of the reference scales automatically. The trained Msc model can detect 10 different scales as mentioned in Section 2.2. However, the accuracy is low for all non-circular shaped reference scales because the YOLO can only use horizontally-placed rectangular boxes (see green line bounding boxes in Appendix Figure 13a,b) to capture the reference scales which could be non-horizontally placed and non-rectangular shape. Interestingly, all the scales with circular shape (e.g., green and blue caps) are accurately detected by the trained scale model at both submerged and dry conditions (Figure 13c,d). For those photos in group 9 (used for scale AI validation) with green/blue caps, we manually measured the photo resolution and then compared their values with those predicted by the scale AI model as shown in Figure 11b. The result verifies the visual observation in Figure 13c,d and provides an accuracy estimation of such an automated approach. For

the blue caps (3.7 cm diameter): the mean error (ME), mean absolute error (MAE), mean absolute relative error (MARE), minimum relative error, and maximum relative error are 0.0039 mm, 0.0065 mm, 3.2%, -5.3%, and 7.5%, respectively. For the green caps (2.5 cm diameter), their values are -0.0006 mm, 0.002 mm, 1.1%, -2.1%, and 1.4%, respectively. Note that the photo resolution uncertainty from manual estimation varies between ±10% and has an overall MARE of 2.32%. This means the cap-based automated scaling approach has a better overall accuracy and a much smaller uncertainty range than the manual approach. Meanwhile, the automated scaling can provide photo resolution of 0.12 – 0.35 mm/pixel, which is also better than the range obtained in the manual approach. Overall, the cap-based automated scaling approach is an efficient alternative to the manual approach in terms of accuracy and resolution.

Both the manual and automated approaches mentioned above are limited for locations we have site accessibility and working permits where we are able to deploy reference scales and use hand-held cameras. These limitations restrict the spatial scale we can observe. Overcoming such limitations necessitates the use of fast remote sensing techniques, such as drones, and requires an approach to reliably estimate the photo resolution captured by the drone cameras. Here we show that the photo resolution can be estimated based on camera height and camera-specific resolution-height relationships. Figure 11c shows the variation of photo resolution (from manual measurements) with respect to height for 3 smartphones, i.e., iPhone 12, 13, and 14 Pro (see photo taken locations in Section 2.1). These relationships provide an additional way to estimate photo resolution for both hand-held and unmanned devices if height information is available.

### 4.6 Limitations

Despite the promise of the proposed approach, limitations exist in photo collection, training data preparation, and HBGC empirical formulas. First of all, by using hand-held devices (e.g., smartphones, tablets, and cameras), the maximum spatial scale and the highest photo resolution are limited. In this work, the actual photo area is limited to be 2.81 $m^2$ (minimum 0.03 and mean 0.26 $m^2$; see details in data package (Y. Chen et al., 2023)). Such a limitation is mainly caused by how high a user can hold a camera. Also, the highest photo resolution is 0.05 mm/pixel and the minimum detectable grain size by YOLO is 0.45 mm. This means that sediments smaller than medium (0.25 – 0.5 mm) or coarse (0.5 – 1 mm) sands may not be reliably detected. Due to these lim-

itations, a much large number of photos are required in order to fully characterize the stream grain sizes and HBGC at watershed scales. The second limitation is the high labor costs required to prepare the training data. Due to the diversity of natural streams, a large number of labels with high quality are needed for reliable prediction of grain sizes (see effects of insufficient training data in Section 4.3). In this work, we spent around 200 hours to label around 17,000 grains to represent most of the stream conditions. Despite such effort, the trained AI still has 20% – 25% relative error for 2 photos (Figure 4d; Figure 9e; Figure 12(f,r)). More data and improved YOLO algorithms may be needed to better capture very large grains at the boundary of the photos.

Additionally, there are limitations in the empirical formulas for HBGC estimations. Due to the low uncertainty and good agreement with calibrated values (Sections 4.1 – 4.2), the Equations 1 and 6 are likely reliable for estimating Manning coefficient and its uncertainty. For friction factor, though it demonstrates large variations and uncertainty (Sections 3.3 – 3.5), the accuracy of Equation 2 has been comprehensively studied and was recognized as the second best formula for resistance estimation with depth and grain size as inputs (Powell, 2014). The Equation 3 for estimating streambed interstitial velocity magnitude is derived from 17 high-resolution CFD simulations driven by structure-from-motion reconstructed streambeds (Y. Chen et al., 2019, 2021). Though it successfully estimates the most likely magnitude of interstitial velocity (Section 4.1), further simulations or experiments with more streambed conditions may be needed to further evaluate its applicability for diverse streambed conditions, especially the relationship between subsurface permeability and the 5th percentile grain size distribution. For uptake velocity, the hydrogeology-biochemistry interaction efficiency term ($\phi$ in Equation 4) is fitted based on field measured data and thus its applicability in diverse streambed conditions also requires further evaluation.

### 4.7 Future directions

As discussed in Section 4.6, the scale and resolution are limited by hand-held approaches. A natural solution is to replace hand-held devices with drones. By using drones it is possible to increase the number of photos and videos with much higher temporal resolution (e.g., 4K and 5.4K videos) and also increase spatial scales. This is primarily due to their high speed (e.g., Skydio 2 and DJI could fly upto 15 – 27 m/s). With available high-resolution streambed data from drones and hand-held devices, an important

future direction is to directly integrate photo-derived high-resolution streambed data with pore-resolved surface-subsurface coupled models and use the simulated pressure, exchange velocity, and turbulence data to improve the empirical formulas for HBGC estimations. With both the improved formulas and high-resolution data, a further step is to integrate the photo-derived streambed grain sizes and HBGC parameters into watershed-scale models aimed at predicting hydro-biogeochemical dynamics.

## 5 Conclusions

This work presents a workflow to extract the quantities, distributions, and uncertainties of streambed grain sizes and hydro-biogeochemistry from photos using YOLO and empirical formulas. The YOLO, an object detection AI model, is firstly trained with 11,977 grain labels from 36 photos representing 9 stream environments, and demonstrates an accuracy of 0.98, 0.98, and 6.65% in terms of the coefficient of determination, the Nash–Sutcliffe efficiency, and mean absolute relative error in predicting the median grain size D50. The model is then used to predict the grain size distributions (GSDs) for 1,999 photos collected at 66 sites in the Yakima River Basin. Three characteristic grain sizes, including the 5th, 50th, and 84th percentiles of GSDs, are subsequently calculated and used to estimate key hydro-biogeochemical parameters, including Manning coefficient, Darcy-Weisbach friction factor, interstitial velocity magnitude, and nitrate uptake velocity.

From the data, the characteristic grain sizes, Manning coefficient, friction factor, interstitial velocity magnitude, and uptake velocity are found to follow log-normal, normal, positively skewed, near log-normal, and negatively skewed distributions, respectively. Their most likely values, i.e., the mode of the distributions, are 6.63 cm (for D50), 0.0339 $s{\cdot}m^{-1/3}$, 0.18, 0.07 m/day, and 1.2 m/day, respectively. And their average uncertainty or variability are reported as 7.33% (for D50), 1.85%, 15.65%, 24.06%, and 13.88%, respectively. The major sources of uncertainties in grain sizes and hydro-biogeochemical parameters are also identified. Specifically, the accuracy of YOLO is the main factor controlling grain size uncertainty. Both YOLO accuracy and stream depth control friction factor uncertainty. The interstitial velocity magnitude uncertainty is determined by both velocity uncertainty and YOLO accuracy. For the uptake velocity uncertainty, it is controlled by uncertainties in velocity, depth, and YOLO accuracy in shallow streams, while controlled by velocity and nitrate concentration uncertainties in deep rivers.

Further analyses of the effects of training data size on YOLO accuracy show that training data with an insufficient number of photos and stream environment types can cause considerable errors in extracting grain size distributions and the statistics of characteristic grain sizes. The selection of a proper class probability threshold is important for avoiding missing or incorrectly selecting individual grains as desired. The photo resolution analyses demonstrate that the integration of circular caps with an AI model can provide an automated scaling approach better than the manual approach in terms of the accuracy and resolution. We also identified the limitations in photo resolution and spatial scale using hand-held cameras, the high labor costs in training data preparation, and the necessity to further improve the empirical formulas for hydro-biogeochemistry estimations. These limitations may be addressed in future research by integrating drone-derived high-resolution streambed data with pore-scale models, and incorporating photo-derived grain sizes and hydro-biogeochemistry parameters to watershed-scale models.

**Acknowledgments**

This research was supported by the United States Department of Energy (DOE) Office of Biological and Environmental Research (BER), Environmental System Science program, through the PNNL River Corridor Science Focus Area (SFA) project (http://www.pnnl.gov/projects/river-corridor). PNNL is operated for the DOE by Battelle Memorial Institute under Contract No. DE-AC05-76RL01830.

## Data availability

All data are available at the ESS-DIVE repository with DOI: 10.15485/1999774 (Y. Chen et al., 2023).

# Appendix

## Derivations of uncertainty propagation equations

As mentioned in Section 2.5, uncertainties occur in YOLO-predicted pixel grain sizes ($D_{xp}$), photo scale measurement ($SC$), and measurements for water depth ($H$), velocity ($U$), and nitrate concentration ($[NO_3^-]$). These uncertainties can further propagate to real grain sizes ($D_x$) and HBGC parameters such as Manning coefficient ($n$), friction factor ($f$), interstitial velocity magnitude ($\sigma_w$), and nitrate uptake velocity ($u_f$). All these uncertainty can be quantified by the ratio of the absolute uncertainty of these quantities to their representative values, for example, manually measured grain sizes and scales, spatial and/or temporal average of depth and velocity, and direct measurement of nitrate concentrations. If denoting the input parameters and subsequently derived grain sizes/HBGC parameters as $x_i$ (i = 1,2,...) and $y_j$ (j = 1,2,...), then the absolute uncertainty can be quantified by $\delta x_i$ and $\delta y_j$ and the relative uncertainty can be calculated as $r_{x_i} = |\delta x_i|/x_i$ and $r_{y_j} = |\delta y_j|/y_j$, respectively. Statistically, such relative uncertainty can be mean absolute relative error (MARE), root-mean-square of the relative error (RMSRE), and the standard deviation of the relative error (STDRE). Here we choose MARE as the reporting metric, however, it can be easily replaced by RMSRE and STDRE. In general, the target $y_j$ is a function of the input parameters $x_i$, which has the form of $y_j = F_j(x_1, ..., x_i, ..., x_n)$. Based on the multi-variable chain rule and the error propagation law (Ku, 1966), the uncertainty of $y_j$ can be computed through Equation 12.

$$r_{y_j}^2 = \frac{(\delta y_j)^2}{y_j^2} = y_j^{-2}\Big[\sum_{i=1}^{n}(\frac{\partial F_j}{\partial x_i})^2(\delta x_i)^2 + \sum_{i=1}^{n}\sum_{k=1,k\neq i}^{n}\frac{\partial F_j}{\partial x_i}\frac{\partial F_j}{\partial x_k}\delta x_i \delta x_k\Big] \tag{12}$$

The last term in Equation 12 represents the correlation among input variable uncertainty and could be assumed as 0 if the uncertainty of input variables are independent to each other. With such an assumption, Equation 12 can be rewritten as Equation 13.

$$r_{y_j} = \sqrt{y_j^{-2}\sum_{i=1}^{n}(\frac{\partial F_j}{\partial x_i})^2(\delta x_i)^2} = \sqrt{\sum_{i=1}^{n}(\frac{\partial F_j}{\partial x_i})^2\frac{(\delta x_i)^2}{x_i^2}\frac{x_i^2}{y_j^2}} = \sqrt{\sum_{i=1}^{n}\Big(\frac{\partial F_j}{\partial x_i}\frac{x_i}{y_j}\Big)^2 r_{x_i}^2} = \sqrt{\sum_{i=1}^{n} s_{x_i}^2 r_{x_i}^2} \tag{13}$$

where $\frac{\partial F_j}{\partial x_i}\frac{x_i}{y_j}$ is the uncertainty propagation scale of $y_j$ to input $x_i$, and is denoted by $s_{x_i}$ for convenience. With such a general form of uncertainty propagation equation, we apply it for real grain size $D_x$ and the four HBGC parameters in Equations 1 – 4.

For $D_x$, it depends on two independent variables $D_{xp}$ and $SC$. Its uncertainty propagation scales are both 1 for $D_{xp}$ and $SC$, which results in Equation 5. For Manning co-

efficient, it depends on only one variable and its propagation scale is 1/6. For friction factor, if denoting $H/D_{84}$ by $H_{D84}$, then Equation 2 becomes a single variable function of $H_{D84}$. Its uncertainty can be calculated by $r_f = |s_{H_{D84}}| r_{H_{D84}}$ with $|s_{H_{D84}}|$ represented by Equation 14.

$$|s_{H_{D84}}| = |\frac{\partial f}{\partial H_{D84}} \frac{H_{D84}}{f}| = \frac{6c_1^2 + c_2^2 H_{D84}^{5/3}}{3c_1^2 + 3c_2^2 H_{D84}^{5/3}} = 2 - \frac{5}{3} \frac{c_2^2 H_{D84}^{5/3}}{c_1^2 + c_2^2 H_{D84}^{5/3}} = 2 - \frac{5}{3} \frac{1}{c_1^2/c_2^2 H_{D84}^{-5/3} + 1}$$
$$= 2\Big\{1 - \frac{5}{6}\big[1 + \frac{c_1^2}{c_2^2} H_{D84}^{-5/3}\big]^{-1}\Big\} = 2\Big\{1 - \frac{5}{6}\big[1 + \frac{c_1^2}{c_2^2}(\frac{H}{D_{84}})^{-5/3}\big]^{-1}\Big\} \tag{14}$$

For the uncertainty term $r_{H_{D84}}$, because $H_{D84} = H/D_{84}$, its uncertainty propagation scales for $H$ and $D_{84}$ are both 1, therefore, $r_{H_{D84}} = \sqrt{r_H^2 + r_{84}^2}$. Such an equation together with Equation 14 leads to Equation 7.

For interstitial velocity magnitude (Equation 3), both $D_5$ and $D_{50}$ are used as inputs. However, these two variables are not independent. To avoid using both sizes as inputs, we use a simplified $D_5$ relationship, $D_5 = 0.23 D_{50}$ (fitted from data; see Section 3.2 and Figure 5d), to replace the YOLO-derived $D_5$ for uncertainty quantification purpose. With such an simplification, Equation 3 is converted to Equation 15.

$$\sigma_w = \frac{0.23^2 c_3 c_5}{2\nu} U^2 D_{50}^{1-c_4} H^{c_4} \tag{15}$$

The uncertainty propagation scales of Equation 15 with respect to inputs $U$, $D_{50}$, and $H$ were computed as 2, (1-$c_4$), and $c_4$, respectively. Combining these scales and the uncertainty of input parameters will lead to Equation 8.

For nitrate uptake velocity, we rewrite Equation 4 in the form of Equation 16 to utilizing the uncertainty equation for friction factor. If we assume no correlation among the three inputs, then the uncertainty propagation scales of $u_f$ with respect to $U$, $f$, and $[NO_3^-]$ are 1, 1/2, and $c_7$, respectively. Combining these scales and the uncertainty of input parameters leads to Equation 9.

$$u_f = \frac{0.17 Sc^{-2/3} c_6}{\sqrt{8}} U f^{1/2} [NO_3^-]^{c_7} \tag{16}$$

901 **Grain size distribution of 20 test photos**

## References


Abeshu, G. W., Li, H. Y., Zhu, Z., Tan, Z., Leung, L. R., & Li, H. Y. (2022). Median bed-material sediment particle size across rivers in the contiguous US. *Earth System Science Data*, *14*(2), 929–942. doi: 10.5194/essd-14-929-2022

Adams, J. (1979). Gravel Size Analysis from Photographs. *Journal of the Hydraulics Division*, *105*(10), 1247–1255. doi: 10.1061/JYCEAJ.0005283

Anastasi, G. (1984). *Geschiebeanalysen im felde unter beruecksichtigung Von grobkomponenten [Field grain-size analyses with special attention to coarse material]* (Tech. Rep. No. 70).

Black, M., Carbonneau, P., Church, M., & Warburton, J. (2014). Mapping sub-pixel fluvial grain sizes with hyperspatial imagery. *Sedimentology*, *61*, 691–711. doi: 10.1111/sed.12072

Bunte, K., & Abt, S. R. (2001). *Sampling surface and subsurface particle-size distributions in wadable gravel- and cobble-bed streams for analyses in sediment transport, hydraulics, and streambed monitoring* (Tech. Rep.). Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station.

Buscombe, D. (2013). Transferable wavelet method for grain-size distribution from images of sediment surfaces and thin sections, and other natural granular patterns. *Sedimentology*, *60*, 1709–1732. doi: 10.1111/sed.12049

Buscombe, D., & Masselink, G. (2009). Grain-size information from the statistical properties of digital images of sediment. *Sedimentology*, *56*, 421–438. doi: 10.1111/j.1365-3091.2008.00977.x

Buscombe, D., & Rubin, D. M. (2012). Advances in the simulation and automated measurement of well-sorted granular material: 2. Direct measures of particle properties. *Journal of Geophysical Research: Earth Surface*, *117*, F02002. doi: 10.1029/2011JF001975

Buscombe, D., Rubin, D. M., & Warrick, J. A. (2010). A universal approximation of grain size from images of noncohesive sediment. *Journal of Geophysical Research: Earth Surface*, *115*, F02015. doi: 10.1029/2009jf001477

Butler, J. B., Lane, S. N., & Chandler, J. H. (2001). Automated extraction of grain-size data from gravel surfaces using digital image processing. *Journal of Hydraulic Research*, *39*(5), 519–529. doi: 10.1080/00221686.2001.9628276

Carbonneau, P. E., Bergeron, N., & Lane, S. N. (2005a). Automated grain size measurements from airborne remote sensing for long profile measurements of fluvial grain sizes. *Water Resources Research*, *41*, W11426. doi: 10.1029/2005WR003994

Carbonneau, P. E., Bergeron, N. E., & Lane, S. N. (2005b). Texture-based image segmentation applied to the quantification of superficial sand in salmonid river gravels. *Earth Surface Processes and Landforms*, *30*, 121–127. doi: 10.1002/esp.1140

Carbonneau, P. E., Lane, S. N., & Bergeron, N. E. (2004). Catchment-scale mapping of surface grain size in gravel bed rivers using airborne digital imagery. *Water Resources Research*, *40*, W07202. doi: 10.1029/2003WR002759

Chaudhry, M. H. H. (2008). *Open-channel flow* (2nd ed.). New York, NY, USA: Springer US. doi: 10.1007/978-0-387-68648-6

Chen, K., Chen, X., Stegen, J. C., Villa, J. A., Bohrer, G., Song, X., ... Zheng, C. (2023). Vertical hydrologic exchange flows control methane emissions from riverbed sediments. *Environmental Science and Technology*, *57*(9), 4014–4026. doi: 10.1021/acs.est.2c07676

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. *Lecture Notes in Computer Science*, *11211*, 833–851. doi: 10.1007/978-3-030-01234-2_49

Chen, Y., Bao, J., Chen, Y., Yang, Y., Stegen, J., Renteria, L., ... Scheibe, T. (2023). *Artificial intelligence models, photos, and data associated with the manuscript "Quantifying streambed grain sizes and hydro-biogeochemistry using YOLO and photos".* Richland, WA, USA: River Corridor and Watershed Biogeochemistry SFA, ESS-DIVE repository. doi: 10.15485/1999774

Chen, Y., Bao, J., Li, B., Liu, X., DiBiase, R., & Scheibe, T. (2021). Estimating hydrological exchange flows from streambed images. In *Agu fall meeting* (pp. B25D–1480). New Orleans, LA.

Chen, Y., DiBiase, R. A. R. A., McCarroll, N., & Liu, X. (2019). Quantifying flow resistance in mountain streams using computational fluid dynamics modeling over structure-from-motion photogrammetry-derived microtopography. *Earth Surface Processes and Landforms*, *44*(10), 1973–1987. doi: 10.1002/esp.4624

Chen, Z., Scott, T. R., Bearman, S., Anand, H., Keating, D., Scott, C., ... Das, J. (2020). Geomorphological analysis using unpiloted aircraft systems, structure from motion, and deep learning. In *2020 ieee/rsj international conference on intelligent robots and systems (iros)* (pp. 1276–1283). IEEE. doi: 10.1109/IROS45743.2020.9341354

Delgado, D., Barnes, M., Boehnke, B., Chen, X., Cornwell, K., Forbes, B., ... Stegen, J. (2023). *Spatial study 2022: surface water samples, cotton strip degradation, and hydrologic sensor data across the Yakima River Basin, Washington, USA (v2)* (Tech. Rep.). Richland, WA, USA: Pacific Northwest National Laboratory. doi: 10.15485/1969566

Detert, M., & Weitbrecht, V. (2012). Automatic object detection to analyze the geometry of gravel grains - a free stand-alone tool. In *River flow 2012 - proceedings of the international conference on fluvial hydraulics* (pp. 595–600). San Jose, Costa Rica: Taylor and Francis.

Elliott, A. H., & Brooks, N. H. (1997). Transfer of nonsorbing solutes to a streambed with bed forms: Theory. *Water Resources Research*, *33*(1), 123–136. doi: 10.1029/96WR02783

Ermilov, A. A., Benkő, G., & Baranya, S. (2022). Automated riverbed material analysis using Deep Learning on underwater images. *Earth Surf. Dynam. Discuss*(November), 1–34. doi: 10.5194/esurf-2022-56

Fehr, R. (1987). Geschiebeanalysen in Gebirgsflüssen [Bed material analyses in mountain streams. Transformation and comparison of various analyses procedures]. *Mitteilungen der Versuchsanstalt für Wasserbau, Hydrologie und Glaziologie*(92), 139.

Ferguson, R. (2007). Flow resistance equations for gravel- and boulder-bed streams. *Water Resources Research*, *43*, W05427. doi: 10.1029/2006WR005422

Ferguson, R. (2010). Time to abandon the Manning equation? *Earth Surface Processes and Landforms*, *35*(15), 1873–1876. doi: 10.1002/esp.2091

Ferguson, R. (2022). Reach-scale flow resistance. *Treatise on Geomorphology*, *6.1*, 110–132. doi: 10.1016/B978-0-12-409548-9.09386-6

Fripp, J. B., & Diplas, P. (1993, apr). Surface sampling in gravel streams. *Journal of Hydraulic Engineering*, *119*, 473–490. doi: 10.1061/(ASCE)0733-9429(1993)119:4(473)

Gomez-Velez, J. D., Harvey, J. W., Cardenas, M. B., & Kiel, B. (2015). Denitrification in the Mississippi River network controlled by flow through river bedforms. *Nature Geoscience*, *8*(12), 941–945. doi: 10.1038/ngeo2567

Graham, D. J., Reid, I., & Rice, S. P. (2005, jan). Automated sizing of coarse-grained sediments: image-processing procedures. *Mathematical Geology*, *37*(1), 1–28. doi: 10.1007/s11004-005-8745-x

Grant, S. B., Azizian, M., Cook, P., Boano, F., & Rippy, M. A. (2018, mar). Factoring stream turbulence into global assessments of nitrogen pollution. *Science*, *359*(6381), 1266–1269. doi: 10.1126/science.aap8074

Grieger, S., Barnes, M., Borton, M., Chen, X., Chu, R., Farris, Y., ... Stegen, J. (2022). *Spatial study 2021: sample-based surface water chemistry and organic matter characterization across watersheds in the Yakima River Basin, Washington, USA (v2)* (Tech. Rep.). Richland, WA, USA: Pacific Northwest National Laboratory. doi: 10.15485/1898914

He, K., Gkioxari, G., Dollar, P., & Girshick, R. (2017, oct). Mask R-CNN. In *2017 ieee international conference on computer vision (iccv)* (pp. 2980–2988). IEEE. doi: 10.1109/ICCV.2017.322

He, K., Zhang, X., Ren, S., & Sun, J. (2014). Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. In *European conference on computer vision* (pp. 346–361). Retrieved from `http://link.springer.com/10.1007/978-3-319-10578-9_23` doi: 10.1007/978-3-319-10578-9_23

Ibbeken, H., & Schleyer, R. (1986). Photo-sieving: A method for grain-size analysis of coarse-grained, unconsolidated bedding surfaces. *Earth Surface Processes and Landforms*, *11*(1), 59–77. doi: 10.1002/esp.3290110108

Jiang, Y., Frankenberger, J. R., Bowling, L. C., & Sun, Z. (2014). Quantification of uncertainty in estimated nitrate-N loads in agricultural watersheds. *Journal of Hydrology*, *519*(PA), 106–116. doi: 10.1016/j.jhydrol.2014.06.027

Kaufman, M., Delgado, D., Barnes, M., Boehnke, B., Chen, X., Cornwell, K., ... Stegen, J. (2023). *Spatial study 2022: water column, sediment, and total ecosystem respiration rates across the Yakima River Basin, Washington, USA* (Tech. Rep.). Richland, WA, USA: Pacific Northwest National Laboratory. doi: 10.15485/1987520

Kellerhals, R., & Bray, D. I. (1971, aug). Sampling procedures for coarse fluvial sed-

iments. *Journal of the Hydraulics Division*, *97*(8), 1165–1180. doi: 10.1061/JYCEAJ.0003044

Kenney, T. C., Lau, D., & Ofoegbu, G. I. (1984, nov). Permeability of compacted granular materials. *Canadian Geotechnical Journal*, *21*, 726–729. doi: 10.1139/t84-080

Ku, H. H. (1966). Notes on the use of propagation of error formulas. *Journal of Research of the National Bureau of Standards, Section C: Engineering and Instrumentation*, *70C*(4), 263–273. doi: 10.6028/jres.070C.025

Lang, N., Irniger, A., Rozniak, A., Hunziker, R., Wegner, J. D., & Schindler, K. (2021, may). GRAINet: mapping grain size distributions in river beds from UAV images with convolutional neural networks. *Hydrology and Earth System Sciences*, *25*, 2567–2597. doi: 10.5194/hess-25-2567-2021

Lang, S., Ladson, T., & Anderson, B. (2004). A review of empirical equations for estimating stream roughness and their application to four streams in Victoria. *Australasian Journal of Water Resources*, *8*(1), 69–82. doi: 10.1080/13241583.2004.11465245

Leopold, L. B. (1970). An improved method for size distribution of stream bed gravel. *Water Resources Research*, *6*(5), 1357–1366. doi: 10.1029/WR006i005p01357

Mueller, D. S., Abad, J. D., García, C. M., Gartner, J. W., García, M. H., & Oberg, K. A. (2007). Errors in acoustic doppler profiler velocity measurements caused by flow disturbance. *Journal of Hydraulic Engineering*, *133*(12), 1411–1420. doi: 10.1061/(asce)0733-9429(2007)133:12(1411)

Mulholland, P. J., Hall, R. O., Sobota, D. J., Dodds, W. K., Findlay, S. E. G., Grimm, N. B., ... Thomasn, S. M. (2009, may). Nitrate removal in stream ecosystems measured by 15N addition experiments: Denitrification. *Limnology and Oceanography*, *54*(3), 666–680. doi: 10.4319/lo.2009.54.3.0666

Mulholland, P. J., Helton, A. M., Poole, G. C., Hall, R. O., Hamilton, S. K., Peterson, B. J., ... Thomas, S. M. (2008). Stream denitrification across biomes and its response to anthropogenic nitrate loading. *Nature*, *452*(7184), 202–205. doi: 10.1038/nature06686

Niehus, S., Perkins, W., & Richmond, M. (2014). *Simulation of Columbia River Hydrodynamics and Water Temperature from 1917 through 2011 in the Han-*

*ford Reach* (Tech. Rep. No. PNWD-3278). Richland, WA: Battelle-Pacific Northwest Division. doi: 10.13140/RG.2.1.5146.8409

NOAA. (2023). *The National Water Model.* Retrieved from `https://water.noaa.gov/about/nwm`

O'Connor, B. L., & Hondzo, M. (2008). Dissolved oxygen transfer to sediments by sweep and eject motions in aquatic environments. *Limnology and Oceanography*, *53*(2), 566–578. doi: 10.4319/lo.2008.53.2.0566

Picioreanu, C., van Loosdrecht, M. C. M., & Heijnen, J. J. (1997, jul). Modelling the effect of oxygen concentration on nitrite accumulation in a biofilm airlift suspension reactor. *Water Science and Technology*, *36*(1), 147–156. doi: 10.2166/wst.1997.0034

Powell, D. M. (2014). Flow resistance in gravel-bed rivers: progress in research. *Earth-Science Reviews*, *136*, 301–338. doi: 10.1016/j.earscirev.2014.06.001

Purinton, B., & Bookhagen, B. (2019). Introducing PebbleCounts: a grain-sizing tool for photo surveys of dynamic gravel-bed rivers. *Earth Surface Dynamics*, *7*, 859–877. doi: 10.5194/esurf-7-859-2019

PyTorch. (2022). *PyTorch.* Retrieved from `https://pytorch.org/`

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: unified, real-time object detection. In *2016 ieee conference on computer vision and pattern recognition (cvpr)* (pp. 779–788). IEEE. doi: 10.1109/CVPR.2016.91

Ren, H., Hou, Z., Duan, Z., Song, X., Perkins, W. A., Richmond, M. C., ... Scheibe, T. D. (2020). Spatial mapping of riverbed grain-size distribution using machine learning. *Frontiers in Water*, *2*, 551627. doi: 10.3389/frwa.2020.551627

Rickenmann, D., & Recking, A. (2011). Evaluation of flow resistance in gravel-bed rivers through a large field data set. *Water Resources Research*, *47*, W07538. doi: 10.1029/2010WR009793

Rubin, D. M. (2004). A Simple Autocorrelation Algorithm for Determining Grain Size from Digital Images of Sediment. *Journal of Sedimentary Research*, *74*(1), 160–165. doi: 10.1306/052203740160

Sadayappan, K., Kerins, D., Shen, C., & Li, L. (2022, nov). Nitrate concentrations predominantly driven by human, climate, and soil properties in US rivers. *Water Research*, *226*, 119295. doi: 10.1016/j.watres.2022.119295

Shepherd, R. G. (1989). Correlations of permeability and grain Size. *Ground Water*, *27*(5), 633–638. doi: 10.1111/j.1745-6584.1989.tb00476.x

Soloy, A., Turki, I., Fournier, M., Costa, S., Peuziat, B., & Lecoq, N. (2020). A Deep Learning-Based Method for Quantifying and Mapping the Grain Size on Pebble Beaches. *Remote Sensing*, *12*, 3659. doi: 10.3390/rs12213659

Son, K., Fang, Y., Gomez-Velez, J. D., Byun, K., & Chen, X. (2022, dec). Combined effects of stream hydrology and land use on basin-scale hyporheic zone denitrification in the Columbia River Basin. *Water Resources Research*, *58*, e2021WR031131. doi: 10.1029/2021WR031131

Steer, P., Guerit, L., Lague, D., Crave, A., & Gourdon, A. (2022, dec). Size, shape and orientation matter: fast and semi-automatic measurement of grain geometries from 3D point clouds. *Earth Surface Dynamics*, *10*(6), 1211–1232. doi: 10.5194/esurf-10-1211-2022

Strickler, A. (1923). *Contributions to the question of a velocity formula and roughness data for streams, channels and closed pipelines* (Tech. Rep.). Bern, Switzerland: Bureau of Water Affairs, Swiss Department of the Interior.

Ultralytics. (2020). *YOLO V5.* Retrieved from `https://github.com/ultralytics/yolov5`

Vázquez-Tarrío, D., Borgniet, L., Liébault, F., & Recking, A. (2017). Using UAS optical imagery and SfM photogrammetry to characterize the surface grain size of gravel bars in a braided river (Vénéon River, French Alps). *Geomorphology*, *285*, 94–105. doi: 10.1016/j.geomorph.2017.01.039

Verdú, J. M., Batalla, R. J., & Martínez-Casasnovas, J. A. (2005, dec). High-resolution grain-size characterisation of gravel bars using imagery analysis and geo-statistics. *Geomorphology*, *72*, 73–93. doi: 10.1016/j.geomorph.2005.04.015

Wang, C.-Y., Mark Liao, H.-Y., Wu, Y.-H., Chen, P.-Y., Hsieh, J.-W., & Yeh, I.-H. (2020). CSPNet: a new backbone that can enhance learning capability of CNN. In *2020 ieee/cvf conference on computer vision and pattern recognition workshops (cvprw)* (pp. 1571–1580). IEEE. doi: 10.1109/CVPRW50498.2020.00203

Wohl, E. E., Anthony, D. J., Madsen, S. W., & Thompson, D. M. (1996). A comparison of surface sampling methods for coarse fluvial sediments. *Water Resources*

*Research*, *32*(10), 3219–3226. doi: 10.1029/96WR01527

Wolman, M. G. (1954). A method of sampling coarse river-bed material. *Transactions, American Geophysical Union*, *35*(6), 951. doi: 10.1029/TR035i006p00951

Zhang, W., Itoh, K., Tanida, J., & Ichioka, Y. (1990). Parallel distributed processing model with local space-invariant interconnections and its optical architecture. *Applied Optics*, *29*(32), 4790. doi: 10.1364/AO.29.004790

Zhang, X., Zhang, Y., Shi, P., Bi, Z., Shan, Z., & Ren, L. (2021). The deep challenge of nitrate pollution in river water of China. *Science of the Total Environment*, *770*, 144674. doi: 10.1016/j.scitotenv.2020.144674

1143 ## Figure captions

**Figure 1.** The locations, site-average median grain sizes, and labels of photos used for AI training/validation/testing (a), prediction (b), scaling sensitivity and accuracy purposes (c), as well as the number of photos at each site (d). The site locations of group 1 (green circles) are invisible due to too close to group 0 and 2. Their locations are described with a character "V" following the site names in (a).

**Figure 2.** The labels of individual grains (a – i) and scales (j – o) in representative river corridor environments.

**Figure 3.** The sketch of the YOLO version 5 network. Modified from Ultralytics (2020).

**Figure 4.** The convergence history of YOLO training (a) and the accuracy of YOLO predicted grain size distribution (b), median grain size D50 (c) as well as the relative error of D50 prediction (d). NSE in (a) is Nash–Sutcliffe efficiency.

**Figure 5.** The probability density distributions of D50 (a), D5 (b), D84 (c), and the relationship between D5/D84 and D50 (d).

**Figure 6.** The probability density distribution of Manning coefficient (a), Darcy-weisbach friction factor (b), fluctuation magnitude of vertical exchange flux (c), and total nitrate uptake velocity attributed to microbes and turbulence mass transfer (d).

**Figure 7.** The in-site variations (blue cross dots), site average values (horizontal black lines), and estimated relative variations (yellow dot lines) of photo resolution (a), log2-transformed D50 (b), log2-transformed D84 (c), log2-transformed D5 (d), and water depth (e) for 32 sites. The site name is reordered in an alphabetical order for convenience. The nearest region to the right of site name represents the data within the site. The site-average value in (b), (c), and (d) are first averaged over the actual data and then log2-transformed.

**Figure 8.** The in-site variations (blue cross dots), site average values (horizontal black lines), and estimated relative variations (red dot lines) for Manning's coefficient (a), log10-transformed friction factor (b), log10-transformed streambed interstitial velocity magnitude (c), and streambed nitrate uptake velocity (d). The site-average value in (b) and (c) are first averaged over the actual data and then log10-transformed.

**Figure 9.** The effects on training photo number on YOLO precision (a), individual grain size distributions (b,c), median grain size (d) and relative error (e) of testing photos, as well as the prediction of median grain size of prediction photos (f). M0a, M0b, and M0c represent models trained with 11, 21, and 36 photos.

**Figure 10.** The effects of probability threshold on model performance metrics R2 (a), mean and mean absolute error (b), and the average number of grains detected by the model (c.)

**Figure 11.** The values of photo resolution and associated detected minimum grain sizes using square quadrats and manual measurements of resolution (a), the comparison of automatically predicted photo resolution to the manually measured values using circular caps (b), and the relationship between photo resolution and camera height (c).

**Figure 12.** The comparison of grain size distribution between YOLO (M0c) prediction and manual measurements for 20 testing photos.

**Figure 13.** The typical scales and YOLO (Msc) predicted scales for the full quadrat (a), 1/4 of the quadart (b), green and blue caps in flowing water (c), and blue cap in dry bed (d).