Title
Kayak drifter surface velocity observation for 2D hydraulic model validation

Journal
River Research and Applications, 34(2)

ISSN
1535-1459

Authors
Barker, JR
Pasternack, GB
Bratovich, PM
et al.

Publication Date
2018-02-01

DOI
10.1002/rra.3238

Peer reviewed

Kayak drifter surface velocity observation for 2D hydraulic model validation

Running head: Kayak drifter surface velocity for 2D model validation

J. Rusty Barker[a], Gregory B. Pasternack[b]*, Paul M. Bratovich[c], Duane A. Massa[d], Joshua R. Wyrick[e], Thomas R. Johnson[f]

[a]Cbec eco engineering, 2544 Industrial Blvd, West Sacramento, CA 95691; email: jrbarker11@gmail.com

[b]Department of Land, Air, and Water Resources, University of California, One Shields Avenue, Davis, CA 95616-8626, USA; voice: (530) 302-5658; email: gpast@ucdavis.edu; ORCID 0000-0002-1977-4175

[c]HDR, Inc., 2365 Iron Point Road, Suite 300, Folsom, CA 95630; voice: (916) 817-4899; email: Paul.Bratovich@hdrinc.com

[d]Pacific States Marine Fisheries Commission, 2711 Highway 20, Marysville, CA 95901; voice: (530) 570-3474; email: DMassa@psmfc.org

[e]Department of Engineering and Computer Science, York College of Pennsylvania, 441 Country Club Road, York, PA, 17403; email: jwyrick@ycp.edu

[f]Thomas R. Johnson, LLC; 7090 Wells Avenue, Loomis, CA 95650; voice: (916) 764-2268; email: trjllc@zetabroadband.com

**Corresponding Author:*

ABSTRACT

Advances in remote sensing, informatics, software and microprocessors enable meter-resolution two-dimensional (2D) hydrodynamic models that produce nearly a census of ecohydraulic conditions over long river segments with 10^5 to 10^8 computational elements. It is difficult to test statistical and spatial model performance at such scope using fixed-point velocity measurements, because field methods are so expensive, laborious, slow, and restricted by safety factors. This study evaluated low-cost water surface particle tracking by kayak with real time kinematic GPS for 2D model validation using 7.2 km of the lower Yuba River in California. Observed flows were between 15 to 140 m^3/s, which were in-channel up to and including bankfull conditions. The coefficients of determination between 5780 observations and 2D model predictions were 0.79 and 0.80 for velocity magnitude and direction, respectively. When surface speed was downscaled and compared to modeled depth-averaged velocity, median unsigned difference was 15.5%. Standard hydrological model performance metrics affirmed satisfactory validation. Surface tracking provided the novel benefit of enabling validation of velocity direction, and that testing found satisfactory performance using all metrics. Having 10 to 1000 times more data enables robust statistical testing and spatial analysis of both speed and direction, which outweighs the loss of depth-averaged data. Both fixed-point and kayak particle tracking methods are useful tools to help evaluate 2D model performance.

Keywords: 2D hydraulic modeling; hydraulic validation; river velocity; river drifters; particle tracking

1. INTRODUCTION

For over 3,000 years people have recorded water velocity according to its literal definition by tracking the displacement of surface particles along a path through time. According to Lumpkin and Pazos (2007), surface particle tracking of ocean currents has been a common practice for centuries, but real-time radio and satellite tracking modernized the technique to produce continuous velocity vector data (Swallow, 1955; Davis, 1991). Today there are ~2,000 ocean drifters perpetually. Surface drifter data is taken as observational truth to calibrate other ocean observation systems (Ducet et al., 2000) and validate ocean circulation models (Lumpkin & Garzoli, 2005; Blockley et al, 2012). Overall, surface drifters are a legitimate, accurate, and widely accepted technology for surface water velocity observation (Fratantoni, 2001). Two journal articles (Stockdale et al., 2007; Han et al., 2016) and some conference proceedings tested drifters in nontidal rivers. Given the utility of drifters in validating ocean circulation models, this study evaluates that for two-dimensional (depth-averaged) hydrodynamic (2D) models of rivers.

1.1. Need For More 2D Model Validation

Today, 2D modeling is essential for river research, management, and rehabilitation, because it addresses spatial patterns of geomorphic processes and ecological conditions. This ascent synergizes with rapid progress in meter-resolution remote sensing (Mandlburger et al., 2015), faster microprocessors, large informatics systems, and computational parallelization (Huxley & Syme, 2016). Studies aiding 2D model development that address boundary conditions (Casas et al., 2010), wetting/drying schemes (Tchamen & Kahawita, 1998), and mesh discretization (Horritt et al., 2006) guide our understanding of model capabilities.

One aspect of 2D modeling is being left behind – validation of velocity vectors. Airborne, boat, and ground technologies provide an abundance of depth data, while velocity validation remains reliant upon fixed-point, scaler speed measurement. Few studies report vector validation, as it requires precise orientation and a multidimensional sensor. Discussion of velocity methods and validation, including advantages and limitations, are in the supplementary file.

1.2. Large-Scale Particle Image Velocimetry

Large-scale particle image velocimetry (LSPIV) can map surface velocity vectors in a fixed, small domain by querying image frames through time (Fujita et al., 1998). Commonly, river LSPIV involves mounting cameras onto telescoping poles, helicopters, or electric drones, depending on cost, stability, flight time, and hazard considerations. Le Coz et al. (2010) demonstrated both the potential and numerous challenges with this technology (see supplementary file for further details).

1.3. Kayak Surface Particle Tracking

A handful of studies have piloted surface drifters with differential GPS in low-wind rivers to obtain velocity along floating paths, but for meter-scale 2D models, real-time kinematic (RTK) GPS is required (see supplementary section 1.3 for details). Drifters have fewer weather, flow, and accessibility limitations than LSPIV and data processing is simpler. They can travel long distances, operate for many hours, and sample a greater diversity of hydraulic conditions. For example, in one six-hour float of 14.3 km, two people recorded 10,250 observations from 0–4.2 m/s, including in gyres and rapids.

When a river has gradually varying changes to velocity, then depth and the thickness of the surface layer moving at the same velocity are key differences between oceans and rivers that change the technological specifications for a drifter. Ocean drifters use drogues that are submerged, tethered weights (or parachutes) to resist wind and move with the thick surface layer. However, rivers are usually too shallow for drogues. Even without drogues, passive drifters get stuck on the bed (Stockdale et al., 2007) or in vegetation.

Often 2D models simulate complex hydraulics with abrupt current accelerations that drifters do not adjust to for some distance. This occurs at riffle/rapid entrances and exits, dissipating jets, in shear zones between fast and slow currents, and in flow recirculations (i.e. eddies or gyres) (Figure 1). The heavier a drifter is, the more it deviates when currents change abruptly. Consequently, drifters in these settings cannot be passive. They need help to move with accelerations. Today, the best solution is manual course correction.

This study investigated such challenging hydraulics. For a 100-m wide river, a manned kayak is a suitable drifter, because it is < 1% of channel width and as wide as a computational element. Hull displacement depends on boat specifications and operator weight, but can be ~10 cm. Operator weight aids drifter submergence deeper into the surface layer.

We have performed seven years of testing kayak drifter observation with people of different paddling skills. Most of the time no manual correction is needed. People can be quickly trained to make infrequent minor adjustments in complex hydraulics. Even a small vector deviation is readily evident and countered almost instantly. Abundant bubbles and particles move with the current for a kayaker to monitor current pattern. More may be added as needed.

Having a mindful operator enables beneficial operations that improve data collection, processing, and analysis. An operator can start/stop data collection in response to GPS precisions or an unsatisfactory kayak vector. An operator has control over the conditions being sampled to sample diverse conditions in complex hydraulics at question in 2D model performance (Brown & Pasternack, 2012).

1.4. Study Objectives

Given the goal of enhancing 2D model validation, study objectives involved (i) evaluating the capabilities of kayak surface velocity vector measurement for 2D model validation; (ii) comparing model-prediction errors of this technique with those of traditional fixed-point validation; (iii) testing for discharge dependence of steady-state model performance; and (iv) using drifter data to better evaluate spatial uncertainties in 2D models. Even though data from drifters and fixed-point observations are different and the metrics to decide what constitutes better performance vary between scientists, putting both types of data through thorough validation procedures provides insights about their advantages and disadvantages. Secondarily, the supplementary file presents a model validation sensitivity analysis of the parameter used to convert surface velocity to depth-averaged velocity. Regardless of surface particle tracking method, this study develops and tests how such velocity data is used to evaluate 2D models. Given strict journal page limits, full details and extra information for all sections of this study are provided in a supplementary file, including figures referred to as “Figure S#” below.

2. STUDY AREA

The regulated 37.5 km lower Yuba River (LYR) begins downstream of Englebright Dam, fed by a 3,490 km^2catchment (Figure 2). It is a wandering cobble/gravel-bed river (Figure 1) with sparse woody vegetation, little entrenchment, an average channel slope of 0.16%, and an average bankfull width of 97 m. Daguerre Point Dam (DPD) is a low-height, historic, partial sediment barrier present in the LYR that influences the longitudinal profile of the river. Water diversions at DPD remove ~ 1–10 m^3/s of the flow. River and diversion flows are quantified by gaging stations along the river, as explained in the supplementary materials file. Due to the storage of historical hydraulic mining sediments in the river valley and a relatively unconstrained winter flood regime, the LYR is unique in the region for having rejuvenation of diverse geomorphic units and streambed substrate (Wyrick & Pasternack, 2014; Pasternack & Wyrick, 2016).

3. METHODS

3.1. 2D Hydrodynamic Modeling

Topography and bathymetry were mapped to produce a 1-m resolution DEM using ground-based, boat-based, and airborne methods (Carley et al., 2012). To exemplify data resolution, within and beyond the wetted area at 24.92 m^3/s, overall point density downstream of the Highway 20 Bridge was 59 and 554 points per 100 m^2, respectively. As previous published in ten journal articles (e.g. Wyrick & Pasternack, 2014; Gonzalez & Pasternack, 2015, Gibson & Pasternack, 2016), SRH-2D (v. 2.1) was used to simulate meter-scale, non-vegetated, in-channel river hydraulics in the LYR from 8.50 to 3,126 m^3/s (0.06 to 22 times bankfull discharge).

3.2. Fixed-Point Hydraulic Data

Fixed-point hydraulic data were obtained using standard methods detailed in the supplementary materials. Locations were within close radio range (~ 0-3 km) of an RTK GPS base station to keep time latency low and base station corrections more accurate. A total of 199 fixed-point observations of depth and velocity were made at ~ 2 m intervals along 17 cross sections (Figure S1), while a different set of 199 points captured water surface elevations. Data were collected on different dates at steady, regulated discharges of 15.26, 22.51, and 22.57 m^3/s.

3.3. Kayak Velocity Measurement

Kayak drifter velocity vector data (5780 observations) were collected at six discharges (17.61–141.9 m^3/s) between December 15, 2009 and July 1, 2010 along 7.2 km, but this whole length was not traversed in one day. One path was done each day either upstream or downstream of DPD. There was no attempt to replicate or differentiate paths for different days or flows. The 7.2-km length passed through 10 of the cross-sections. Others were located further upstream but still in the same river conditions.

Kayakers simply got on a path and stayed on it. If the current took the boat closer to or father away from a bank, then so be it- that was left to chance. Today, we pause data collection frequently to move to different velocity streamlines. This maximizes the range of velocities observed, improves equal-effort sampling across the velocity range, and allows mapping of recirculations.

To evaluate the efficacy of the drifting strategy to sample diverse hydrogeomorphic conditions, kayak velocity point locations were analyzed for proximity to the riverbank toe (point distance tool in ArcGIS® 10.x) and presence within each of the in-channel morphological units (spatial join tool in ArcGIS® 10.x) mapped by Wyrick and Pasternack (2014). For the entire LYR, the average width between bank toes is 59.4 m, so the average distance from bank toe to centreline is 29.7 m. Relative to those metrics, the range of distances of kayak velocity observations away from the bank toe was 0.03 – 37.6 m (median of 13.7 m, standard deviation of 6.9 m). When "close to bank" is defined as within 1/5 of the average half width of the river (< 5.94 m), then 13% of kayak velocity data met this criterion. Turning to morphological representation, fast glide (1711 points) and run (1134) were sampled most. Chute, pool, slow glide, riffle, and riffle transition each had 393-802 observations. The least sampled unit was slackwater, which still had 97 observations. These data substantiate that repeatedly drifting along a single streamline for several kilometers per day for a few days will capture hydraulic diversity.

Velocity vector equations are in the supplementary materials file. Velocity magnitude was calculated as the horizontal displacement from one GPS position to the next, divided by the change in time between position measurements. This value was assigned to the midpoint coordinates. Velocity direction was assigned using the differences in northing and easting of consecutive position measurements.

Kayak position was measured on a fixed time interval using a Trimble® R7 RTK GPS rover linked by radio to a local R7 base station GPS within 3 km of the rover, yielding centimetre-to-decimetre scale horizontal accuracy. GPS satellite clocks are specified to record time to within 40 ns of Coordinated Universal Time (Allan et al., 1997), but GPS data collectors only store values to the nearest 1 s. Initially, measurements were recorded at 3-s intervals, but the time step was adjusted to 5 s in later runs (17.6 m^3/s, 23.1 m^3/s, and 141.9 m^3/s) to reduce the relative effect of GPS time-recording precision limitations on velocity calculations. For a fixed sampling interval, it does not matter if the GPS clock is exactly on the integer second for each sample, only that the sampling keeps a constant interval. Because the method of determining time does not change over time, the sampling interval should remain fixed. In this case, slower velocities are more prone to error than fast velocities, because they involve shorter distances, so positional error is a larger fraction of distance. Velocity precision was calculated as the sum of the GPS horizontal precisions for the two positions used to compute a velocity divided by the time to traverse between the two positions. Any velocity with absolute precision worse than 0.03 m/s was discarded, so a strict quality criterion was applied.

3.4. Data Analysis

3.4.1. *Validation Data Preparation*

Validation data consisted of three types: fixed-point observations, unadjusted surface velocity vectors, and adjusted (depth-averaged) velocity magnitude. For each location of an observation, GPS coordinates were imported to ArcGIS®. Then model values there were interpolated from triangulated irregular networks built using the irregular, meter-scale point clouds of model outputs.

3.4.2. *Model Performance Indicators*

There are no standards for 2D model validation. Common metrics are reviewed by Pasternack (2011), while others exist for hydrological validation (Moriasi et al., 2007). Concepts and algorithms for validating 2D model velocity direction in river eddies are non-existent whereas atmospheric and ocean studies often report visual comparisons. Metrics and their threshold values are detailed in the supplementary file.

Metrics used herein include (i) basic statistical measures for signed and unsigned deviations and percent differences; (ii) regression analysis metrics of slope, y-intercept, r^2, p-value, regression slope standard error, and regression intercept standard error; and (iii) the hydrological metrics Nash-Sutcliffe efficiency (NSE), percent bias (PBIAS), and the root mean square error-observations standard deviation ratio (RSR). Regression slope standard error is the average distance that observed values fall from the regression line and thus indicates how wrong the regression model is on average using the units of the response variable.

3.4.3. *Direct 2D Model Validation*

Validation metrics were computed using fixed-point depth and velocity data collected by wading. They were also computed using observed kayak surface velocity direction versus modeled velocity direction. For raw kayak surface velocity magnitude, the only viable test involved regression analysis as surface and averaged velocities are related and different only by a scaling coefficient (Hulsing et al., 1966; Rantz, 1982). That does not affect the regression's coefficient of determination or p-value.

3.4.4. Adjusting Kayak Surface Velocity To Mean Velocity

A depth-averaged velocity constant (DAVC) of 0.71 was taken from unpublished profiles collected at one site on the LYR by the authors, which also matched vertical velocity profile data collected in a similar setting (Pasternack et al., 2006). Similar values of 0.72–0.79 were reported by Dramais et al. (2011) for similar conditions. A DAVC sensitivity analysis is in supplementary section 3.4.4.

3.4.5. Testing Discharge Dependence of Model Error

The number of flows to be observed at steady state and over what range for 2D model validation is an important open question. Flood hazard models are used to protect society from extreme events never observed, yet parameters are necessarily calibrated and validated during safe, steady in-channel flows. If any aspect of a 2D model is discharge-dependent, it is the bed-roughness parameter.

The deeper water is relative to bed roughness, the lower Manning's n might be. Alternately, the increase in very rough, shallow areas along channel flanks that comes with increasing discharge can offset the increased bed submergence in the thalweg to keep the overall cross-sectional roughness the same. Increased roughness along the flanks affects thalweg hydraulics (Abu-Aly et al., 2013).

There are no 2D model validation datasets as of yet involving meter-scale resolution topography in which Manning's n was held constant, and then velocity and depth errors were inversely related to each other and both a function of relative submergence. Thus, fixed-point and kayak validations were checked for any trend in velocity error as a function of depth. There are also no specific tests of discharge-dependence of model performance, either. The kayak method enabled such a test infeasible by wading. Validation was performed using only the subset of velocities recorded during the lowest and highest observed flows (17.6 m^3/s and 141.9 m^3/s) to determine whether validation results would differ from those using data from all flows and see if the data plotted along the same trend line.

3.4.6. Comparing Validation Outcomes

Practitioners want to know what value they get for deploying a traditional fixed-point strategy or this new kayak drifter strategy at the holistic level. This does not require head-to-head sampling of the same points by the two methods, as it is a comparison of two strategies. One benefit in attempting comparison is not to say that one method is necessary "better" than another, but to convey what kinds of outcomes are likely in each case and why. Another benefit is to reveal how one's sense of whether a model is validated or not might depend on the type and abundance of data collected, which is currently not understood. Finally, each approach can do unique things, so what is gained by trying direction validation and more spatial analysis of speed error? Readers can judge for themselves what technical metrics and qualitative comparisons to value.

In our own assessment of fixed-point and kayak surface data collection methods for our future validation needs, we found value in four comparisons. The ability of the kayak method to yield abundant, diverse direction data is uncontested by typical fixed-point sampling, so that is not comparable. First, the range of flows and depths where velocities were measured, as well as the range of velocities, are meaningful comparative indicators of what each strategy can accomplish. Model validation data should span actual conditions. Second, head-to-head scatter plot validation tests need a roughly uniform distribution of velocity values as well as enough points in each velocity bin to represent the distribution of prediction deviations about the mean value in each bin. This was assessed using the velocity statistical data for each method. Third, all quantitative validation metrics were compared between methods. Finally, locations with > 50% model error were compared for the two techniques with a histogram analysis. Considered together, these provided a broad-based evaluation of fixed-point versus kayak 2D model validation strategies without necessarily aiming to say that one is better than another.

4. RESULTS

4.1. 2D Model Validation Using Fixed-Point Velocities

Fixed-point velocity magnitude data showed a typical 2D model validation outcome, with the points generally along a 1:1 line for observed versus modelled and the best-fit line showing a bias (Figure 3a). Model validation metrics using fixed-point data exceeded all performance standards (Table 1). Thus, the 2D model is validated.

4.2. 2D Model Validation Using Kayak Observations

4.2.1. Statistical Velocity Validation

The results of DAVC sensitivity analysis are in the supplementary file. Using the DAVC value of 0.71, performance indicators surpassed validation thresholds farther than they did using fixed-point data (Figure 3b). There is so much data in the plot that it is not possible to see the outcome that the majority of the residuals are very close to the 1:1 line; thus, it is necessary to focus on the quantitative performance metrics (Table 1). Once again, the 2D model is validated by every single metric comparing against kayak observations. Considering the standard error of the regression slope and intercept values, both are low and they help to appreciate that a regression plot with thousands of points often hides the true precision of the results, which in this case is quite high. In fact, the standard error of the slope for the kayak data was 8.4 times lower than that for the fixed-point data, while that of the intercept was 5.9 times lower. Similar to fixed-point data, kayak data showed a bias in which low values were over-predicted and high values under-predicted, indicating the typical 2D model problem of insufficient lateral velocity gradients.

Descriptive statistics revealed a mean difference of -0.095 m/s, standard deviation (SD) of 0.228 m/s, mean velocity error of 4.4%, and error SD of 29.0%. Similar to the fixed-point data and other studies, the kayak data found the largest errors at the lowest observed velocities (Figure 4). The unsigned (aka absolute value) median error was 15.5%. Only seven observations had absolute errors > 160% and they were all for very low velocities. When velocity was above 0.5 m/s, errors were almost all < 100%. Considering the cumulative distribution function of velocity error for all the kayak data, 12.7, 58.7, and 90% of the modeled values were within 0.03, 0.15 m/s and 0.34 m/s, respectively.

4.2.2. No Discharge Dependence of Model Error Found

When steady state model performance was judged using only velocities observed at the highest and lowest flows (n=1,336; DAVC=0.71), the r^2 was 0.80 and the slope was 0.90 (Figure 5). This represents a 1.4% difference in the r^2 and a 9.4% difference in slope with the validation performed for the entire dataset. The mean difference in velocities and SD was -0.072 m/s and 0.219 m/s, respectively, while the mean error was 6.1% (Table 2a). These represent 24, 4, and

39% changes, respectively, from the values obtained from the entire dataset. Considering the cumulative distribution function of velocity error using only velocities observed at the highest and lowest flows revealed that 12.5, 61.9, and 90.0% of modeled values were within 0.03, 0.15, and 0.38 m/s, respectively.

When velocity deviations between the model and fixed-point data were regressed against observed depth, there was no significant trend. The same outcome was obtained when deviations between model and kayak velocities were regressed against modeled depths. These results further affirm that there is no discharge dependence on model performance and there is no control of relative submergence on velocity error when using a constant Manning's n in a 2D model of a gravel/cobble river over the full range of in-channel flows. These findings are important, as they go against expectation. Most likely, relative submergence is too small of an effect for in-channel flows. If it matters anywhere, it is in the shallowest edge zone and in vegetated terrain, where Manning's n increases to values of 0.1-0.3.

4.2.3. Spatial Velocity Validation

Spatial analysis of velocity deviations enabled by kayak data revealed consistent locations of poor 2D model performance, either associated with localized DEM deficiencies or steep riverbed slopes. The discussion of the former is not novel and is relegated to the supplementary file. The study found consistently large differences at the entrances and exits of riffles, which is where abrupt slope changes occur (Figure 6). Because the kayak was manned and kept moving at the speed of the water in these locations, this error is attributable to the 2D model. Entrances and exits of riffles are often steeply sloped, causing a violation of the 2D model's horizontal flow assumption in exactly this fashion. The average difference for the three areas shown in Figure 6 was -0.18 m/s, with an absolute difference of 0.36 m/s, both of which are about double the values for the entire dataset. We have observed this effect before in our other unpublished validation datasets with fixed-point data, so we are confident to attribute this to the 2D model, but it is rare to have cross-sections in these locations that are challenging to holding position in (whether by wading or by motorboat), so there has been insufficient data to see this longitudinal effect clearly until now.

4.2.4. Directional Validation

Velocity direction validation plots and metrics substantially exceeded common threshold values for model performance (Table 1; Figure 7; Figure S8), though there were subpar locations. Standard error of the regression slope and intercept were extremely low. The 2D model generally performed within the 9° criterion (Table 2b). The mean raw direction difference was -0.11° with a SD of 9.7°. Histogram analysis of differences revealed that 15% of model predicted values were within 1° of observed values, 62% were within 5°, and 86% were within 10°. The mean absolute error was 2.7% with a SD of 3.8%. A similar analysis of the directional error showed that 88%, 98%, and 99.8% of predicted values had errors less than 5%, 10%, and 25%, respectively.

Although almost all differences were < 45°, there were 20 outliers. Of these, half occurred in model-predicted eddies along a complex bank that the kayak did not experience at those locations (Figures S5 and S8). This problem is attributable to mesh structure and resolution compared to submeter effects of topographic roughness along the water's edge. Modelers do not carefully design meter-scale mesh structure along banks for long distances, and bank conditions changes with discharge anyway.

Another effect observed in eddy validation involved the occurrence of a model eddy slightly shifted or scaled differently than the real eddy. This small spatial difference can cause a large apparent error in magnitude when compared at a fixed point, even if the results are correct for the same relative position in an eddy. Thus, head-to-head velocity direction comparisons are problematic in the vicinity of eddies.

4.3. Comparing Validation Outcomes

The maximum estimated depth-averaged kayak velocity was 2.44 m/s (double that measured while wading). The kayak data r^2 was 0.22 higher than that for fixed-point data, backed by an order of magnitude more points to assess statistical assumptions and confidence. The smallest kayak regression slope was 0.9 (DAVC=0.64) and even that was larger than the highest benchmark established by Lane et al. (1999). Comparison of locations with error > 50% model error is in the supplementary file.

Histograms of observed velocities between the two methods show that kayak measurements had a better distribution (Figure 8). Nearly 30% of fixed-point measurements were 0.25–0.5 m/s, while kayak-observed velocities had a peak with only 25% between 0.5–0.75 m/s. Approximately 60% of velocities were < 1 and 0.75 m/s for fixed-point and kayak data, respectively. The upper tail is noticeably larger for kayak data. In the future, if mindful effort was made to direct the kayak through diverse and unwadable conditions, then the kayak dataset would far exceed the range and value of a fixed-point dataset.

It was demonstrated in Section 4.2.3 that there were discrepancies between modeled and observed kayak velocities into and out of riffles. Velocities were measured on one of these riffles (Figure 6b) during the cross-sectional fixed-point survey (Cross-sections 3 and 4). Side-by-side visual inspection of both sets of data (Figure 9) at cross-section 4 revealed the trend that measurements made on the inside one-half of the channel (river-right) had larger differences (average absolute difference = 0.42 m/s) than the measurements made on the outside one-half of the channel (river-left) (average absolute difference = 0.20 m/s). This observation was consistent for both velocity measurement methods suggesting a 2D model problem, not a kayak method problem.

5. DISCUSSION

Using common fixed-point validation methods, our 2D model performed on par with scientific literature. The same occurred using kayak data. Most people prefer to get an outcome with less effort and lower cost. Having established satisfactory performance from both strategies, discussion focuses on the additional benefits enabled by kayak velocity reconnaissance. The supplementary file includes two additional subsections that detail limitations and future improvements.

5.1. Improved Sampling of Deep And Fast Areas

Compared to the < 1.3 m wadable depths for fixed-point measurement in the ambient currents, the kayak had no depth or velocity constraint in the testbed river. This also enabled a wide range of discharges to be sampled, including a flow above the representative bankfull discharge that covered point and medial bars. This capability is especially important for validating 2D models in floods and steeper mountain rivers (Pasternack & Senter, 2011). Kayaks use no fuel and require less maintenance and set-up time than a motorboat. Some comparison to acoustic Doppler current profiling is discussed in the supplementary materials file as is the discharge-dependence analysis.

5.2. Improved Sampling Of Statistical Structure

Statistical analysis of model performance using kayak data differed from that using fixed-point data, because the former involved an order of magnitude more data spanning a wider range of flows, depths, and velocities. Collection of kayak data is less time consuming, so the same effort by both methods yields a substantial difference in test data. These advantages provide a beneficial capability for characterizing the statistical structures of deviation residuals. Specific numerical outcomes on statistical performance are discussed in the supplementary materials file.

This study also showed that the kayak method more evenly represented velocities across a wider range (Figure 8). A 2D model can over-predict low velocities and under-predict high velocities, so a higher range, more evenly distributed dataset aids testing for this. May et al. (2009) found a significant local bias of this nature, but that study was impacted by the occurrence of most observations in the 1.5–3.0 m/s range.

Kayak velocity measurements were made along longitudinal transects in this study, but paths were not always centered on the thalweg. As a result, data spanned morphological unit types with hundreds of points each and sampled all positions across the channel. This more appropriately represented the complete range of varying hydraulics typical of low-flow velocity fields in shallow gravel-bed rivers compared to the fixed-point approach that miss important transitions and peak velocities.

5.3. Improved Spatial Testing

Another novel outcome of this study was that modelers can now map and compare observations for long river segments using kayak drifters. When mapped, areas with higher uncertainty that would benefit from multiple sampling runs can be easily identified. In particular, for this study there were consistently large differences between model and observed results at the entrances and exits of riffles using both observational methods (Figure 6). With bathymetric LiDAR and multibeam echosounding, meter to sub-meter resolution 2D models are reaching the natural limitation in accuracy imposed by their inherent structural assumptions regarding 2D flows and sub-grid scale turbulence.

5.4. Putting The 2D In 2D Model Validation

The ability of surface particle tracking to map velocity direction adds a sorely needed dimension to 2D model validation. Test results showed a strong correlation between observed and predicted flow direction. The averaged and measured directions matched up well for the ambient flow, and where vectors converged or diverged. The more flow obstructions a river has, the more eddies it will have and thus the greater need for more comprehensive spatial pattern analysis of velocity vectors.

Using a fixed-point method with a multi-axis sensor adds a significantly cumbersome requirement of precisely characterizing sensor direction in the same coordinates as used in a 2D model. The kayak, by contrast, was carried along by the water, which eliminated the uncertainty about direction of flow. GPS data collection with the kayak is natively set to collect data in the same coordinate system as the 2D model. Although some error is inevitable due to kayak momentum and GPS time-recording precision, operator course corrections was anecdotally highly beneficial. Direction validation is now practical and cost effective.

6. CONCLUSIONS

No hydrodynamic model is free of uncertainty, so modelers are responsible for evaluating and conveying limitations to stakeholders. As a community, 2D hydraulic modelers should tackle spatial patterns of velocity deviation and offer diverse statistical validation metrics. Standardized methods using drifter data can provide those advancements. With regard to the four primary objectives, this study found that:

- Kayak surface velocity tracking is well suited for 2D model validation (section 4.2).
- On a comparative basis, kayak data outperformed fixed-point data for all but one metric (sections 4.3 and 5). There is no metric that fixed-point methods can obtain that kayak methods cannot also obtain to test a 2D model.
- There is no discharge dependence of 2D model performance in most cases, because there is nothing fundamentally different about open-channel hydraulics over a wide range of flows in many settings (section 4.2.2).
- Kayak surface velocity tracking provides significantly more data for the same cost compared to fixed-point observation, it enables evaluation of velocity spatial patterns, and it addresses velocity magnitude direction (sections 4.2.3 and 4.2.4).

Each data collection approach has assumptions, uncertainties, and limitations, though people are well acclimatized to those for fixed-point methods and so tend to downplay them. Although there is uncertainty when choosing a DAVC, the numerous benefits of kayak mapping discovered in this study suggest that practitioners should add this capability to more completely validate model predictions. Nevertheless, there are times when fixed-point methods are more useful than kayak methods, too.

7. GEOLOCATION

39°13′13″ N, 121°20′7″ W

8. ACKNOWLEDGMENTS

Financial and in-kind labor support for this work was provided by Pacific Gas & Electric Company, the U.S. Fish and Wildlife Service Anadromous Fish Restoration Program (Agreement #113323J011), Yuba County Water Agency, and the Yuba Accord River Management Team. This project was also supported by the USDA National Institute of Food and Agriculture, Hatch project number #CA-D-LAW-7034-H. Non-financial acknowledgements are in the supplementary file.

9. REFERENCES

Abu-Aly, T. R., Pasternack, G. B., Wyrick, J. R., Barker, R., Massa, D. A., & Johnson, T. R. (2013). Effects of LiDAR–derived, spatially distributed vegetation roughness on two–dimensional hydraulics in a gravel–cobble river at flows of 0.2 to 20 times bankfull. *Geomorphology*, 206, 468–482. DOI: 10.1016/j.geomorph.2013.10.017.

Allan, D. W., Ashby, N., & Hodge, C. C. (1997). The science of timekeeping. Palo Alto, CA: Hewlett Packard, Application Note 1289.

Blockley, E. W., Martin, M. J., & Hyder, P. (2012). Validation of FOAM near-surface ocean current forecasts using Lagrangian drifting buoys. *Ocean Science*, 8, 551-565. DOI: 10.5194/os-8-551-2012.

Brown, R. A., & Pasternack, G. B. (2012). Monitoring and assessment of the 2010-2011 gravel/cobble augmentation in the Englebright Dam Reach of the lower Yuba River, CA. Davis, CA: University of California at Davis, Prepared for the U. S. Army Corps of Engineers, Sacramento District.

Carley, J. K., Pasternack, G. B., Wyrick, J. R., Barker, J. R., Bratovich, P. M., Massa, D. A., Reedy, G D, Johnson T R. (2012). Significant decadal channel change 58–67 years post–dam accounting for uncertainty in topographic change detection between contour maps and point cloud models. *Geomorphology*, 179, 71–88.

Casas, M. A., Lane, S. N., & Benito, G. (2010a). A method for parameterising roughness and topographic sub–grid scale effects in hydraulic modelling from LiDAR data. *Hydrology and Earth System Sciences Discussions*, 7, 2261–2299.

Davis, R. E. (1991). Lagrangian ocean studies. *Annual Review of Fluid Mechanics*, 23, 43–64.

Dramais, G., Le Coz, J., Camenen, B., & Hauet, A. (2011). Advantages of a mobile LSPIV method for measuring flood discharges and improving stage–discharge curves. *Journal of*

Hydro–environment Research, 5, 301–312.

Ducet, N., Le Traon, P. Y., & Reverdin, G. (2000). Global high-resolution mapping of ocean circulation from TOPEX/Poseidon and ERS-1 and -2. *Journal of Geophysical Research*, 105(C8), 19477–19498. DOI: 10.1029/2000JC900063.

Fratantoni, D. M. (2001). North Atlantic surface circulation during the 1990's observed with satellite-tracked drifters. *Journal of Geophysical Research*, 106(C10), 22067–22093. DOI: 10.1029/2000JC000730.

Fujita, I., Muste, M., & Kruger, A. (1998). Large–scale particle image velocimetry for flow analysis in hydraulic engineering applications. *Journal of Hydraulic Research*, 36, 397–414.

Gibson, S. A., & Pasternack, G. B. (2016). Selecting between one–dimensional and two–dimensional hydrodynamic models for ecohydraulic analysis. *River Research and Applications*, 32, 1365–1381.

Gonzalez, R. L., & Pasternack, G. B. (2015). Reenvisioning cross-sectional at–a–station hydraulic geometry as spatially explicit hydraulic topography. *Geomorphology*, 246, 394–406.

Han, E. J., Park, I., Do Kim, Y., & Seo, I. W. 2016. A study of surface mixing in a meandering channel using GPS floaters. *Environmental Earth Sciences*, 75, 901.

Horritt, M. S., Bates, P. D., & Mattinson, M. J. (2006). Effects of mesh resolution and topographic representation in 2–D finite volume models of shallow water fluvial flow. *Journal of Hydrology*, 329, 306–314.

Hulsing, H., Smith, W., & Cobb, E. D. (1966). Velocity–head coefficients in open channels (Water Supply Paper 1869–C). Washington, DC: United States Geological Survey.

Huxley, C., & Syme, W. (2016, Nov 28-Dec2). TUFLOW GPU – best practice advice for hydrologic and hydraulic model simulations. Paper presented at the Hydrology and Water Resources Symposium, Queenstown, New Zealand.

Lane, S. N., Bradbrook K. F., Richards K. S., Biron P. A., & Roy, A. G. (1999). The application of computational fluid dynamics to natural river channels: three–dimensional versus two–dimensional approaches. *Geomorphology*, 29, 1–20.

Le Coz, J., Hauet, A., Pierrefeu, G., Dramais, G., & Camenen, B. (2010). Performance of image-based velocimetry (LSPIV) applied to flash-flood discharge measurements in

Mediterranean rivers. *Journal of Hydrology*, 394, 42-52. DOI: 10.1016/j.jhydrol.2010.05.049

Lumpkin, R., & Garzoli, S. L. (2005). Near-surface circulation in the Tropical Atlantic Ocean. *Deep Sea Research Part I*, 52, 495-518. DOI: 10.1016/j.dsr.2004.09.001.

Lumpkin, R., & Pazos, M. (2007). Measuring surface currents with Surface Velocity Program drifters: The instrument, its data, and some recent results. In A. Griffa, A. Kirwan Jr., A. Mariano, T. Özgökmen, & H. Rossby, H, (Eds.), *Lagrangian Analysis And Prediction Of Coastal And Ocean Dynamics* (pp. 39-67). Cambridge: Cambridge University Press. DOI:10.1017/CBO9780511535901.003.

Mandlburger, G., Hauer, C., Wieser, M., & Pfeifer, N. (2015). Topo–Bathymetric LiDAR for Monitoring River Morphodynamics and Instream Habitats– A Case Study at the Pielach River. *Remote Sensing*, 7, 6160.

May, C. L., Pryor, B. S., Lisle, T., & Lang, M. (2009). Coupling hydrodynamic modeling and empirical measures of bed mobility to predict the risk of scour and fill of salmon redds in a large regulated river. *Water Resources Research*, 45, W05402. DOI: 10.1029/2007WR006498.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the American Society of Agricultural and Biological Engineers*, 50, 885–900

Pasternack, G. B. (2011). *2D Modeling and Ecohydraulic Analysis*. Seattle, WA: Createspace.

Pasternack, G. B., & Senter, A. E. (2011). 21st Century instream flow assessment framework for mountain streams (CEC–500–2013–059). Sacramento, CA: California Energy Commission, PIER.

Pasternack, G. B., & Wyrick, J. R. (2016). Flood–driven topographic changes in a gravel–cobble river over segment, reach, and unit scales. *Earth Surface Processes and Landforms*, 42(3), 487-502. DOI: 10.1002/esp.4064.

Pasternack, G. B., Gilbert, A. T., Wheaton, J. M., & Buckland, E. M. (2006). Error propagation for velocity and shear stress prediction using 2D models for environmental management. *Journal of Hydrology*, 328, 227–241.

Rantz, S. E. (1982). Measurement and computation of streamflow, Volume 1, Measurement of stage and discharge (Water Supply Paper 2175). Washington, DC: U.S. Geological Survey.

Stockdale, R. J., McLelland, S. J., Middleton, R., & Coulthard, T. J. (2007). Measuring river velocities using GPS River Flow Tracers (GRiFTers). *Earth Surface Processes and Landforms*, 33, 1315-1322, DOI: 10.1002/esp. 1614.

Swallow, J. C. (1955). A neutral–buoyancy float for measuring deep currents. *Deep Sea Research*, 3, 74–81.Tchamen, G. W., & Kahawita, R. A. (1998). Modeling wetting and drying effects over complex topography. *Hydrological Processes*, 12, 1151–1182.

Wyrick, J. R., & Pasternack, G. B. (2014). Geospatial organization of fluvial landforms in a gravel–cobble river: beyond the riffle– pool couplet. *Geomorphology*, 213, 48–65. DOI: 10.1016/j.geomorph.2013.12.040.

Table 1. Velocity validation metrics comparing fixed-point and kayak observations.

Validation metrics.	Velocity magnitude Fixed-point	Velocity magnitude Kayak	Velocity direction
Trendline slope	0.78	1.00	0.90
Trendline r^2	0.57	0.79	0.80
Regression slope standard error	0.0455	0.0054	0.0059
Regression intercept standard error	0.035	0.006	1.24
NSE	0.52	0.75	0.79
PBIAS	-4.40%	9.87%	0.05%
RSR	0.69	0.50	0.46

Note. NSE = Nash–Sutcliffe efficiency; PBIAS = percent bias; RSR = root mean square error−observations standard deviation ratio.

Table 2. Descriptive statistics for modeled versus kayak-observed highest and lowest velocity magnitude and direction. ABS is the absolute value.

Statistic	(A) Velocity magnitude (m/s)			
Statistic	V_{Diff}	ABS(V_{Diff})	V_{Error}	ABS(V_{Error})
Mean	-0.072	0.163	6.1%	18.2%
Standard Error	0.006	0.004	0.6%	0.4%
Median	-0.066	0.109	8.3%	14.1%
Standard Deviation	0.219	0.162	22.8%	15.0%
Range	2.011	1.238	167.0%	94.5%
Minimum	-1.238	0.000	-72.5%	0.0%
Maximum	0.772	1.238	94.5%	94.5%
	(B) Velocity direction (deg)			
Statistic	Dir_{Diff}	ABS(Dir_{Diff})	Dir_{Error}	ABS(Dir_{Error})
Mean	-0.11	5.49	-0.06%	2.66%
Standard Error	0.13	0.10	0.06%	0.05%
Median	-0.01	3.75	0.00%	1.81%
Standard Deviation	9.65	7.94	4.65%	3.82%
Range	406.06	228.48	196.68%	102.04%
Minimum	-228.48	0.00	-102.04%	0.00%
Maximum	177.58	228.48	94.64%	102.04%

Figure Captions

Figure 1. Photos illustrating the diversity of hydraulic conditions on the low Yuba River.

Figure 2. Lower Yuba River corridor below the Highway 20 Bridge, its local and regional context, 2D model reach domains, and geomorphic reaches.

Figure 3. Model predicted depth-averaged velocities versus observations from (a) fixed-point measurement and (b) kayak measured velocities using DAVC=0.71. It is not possible to convey the small residuals of the majority of kayak points, so refer to the standard error of the slope metric in the text.

Figure 4. Absolute velocity error versus observed kayak velocities (DAVC = 0.71).

Figure 5. Predicted model velocities versus kayak-based velocity results (DAVC = 0.71) for the highest and lowest flows in the study (141.9 and 17.6 m^3/s, respectively).

Figure 6. Differences between modeled and observed kayak velocities near riffles.

Figure 7. Observed versus model-predicted direction of flow. The largest errors are in model-predicted eddies that were either not present in reality or were not located where the kayak drifted. Even a small difference in model eddy size and shape can cause large directional deviations.

Figure 8. Percent distribution of 2D model velocity errors binned by observed velocity, comparing results between fixed-point and kayak datasets.

Figure 9. Differences of kayak and cross-section fixed-point velocities with modeled velocities showing consistency in trends across the channel.

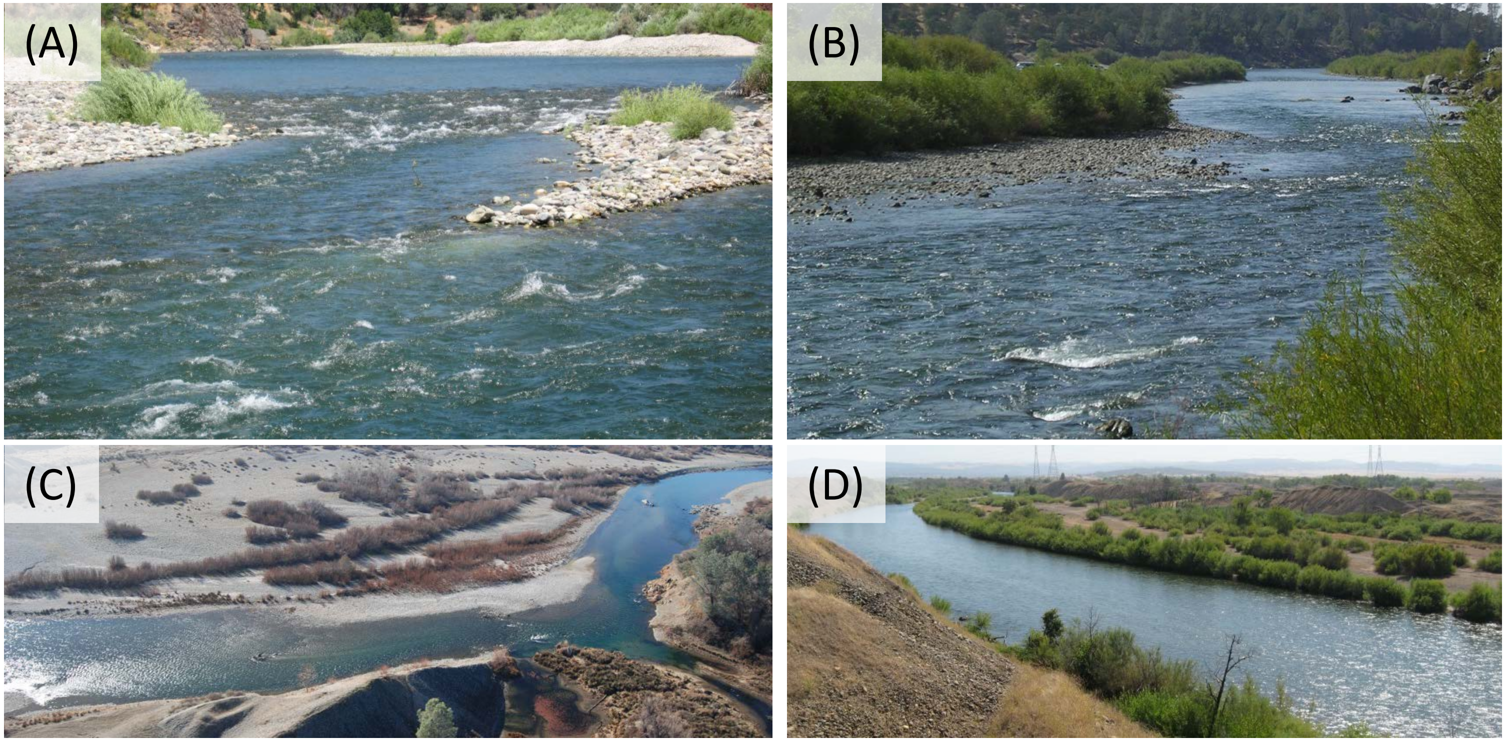


Figure 1

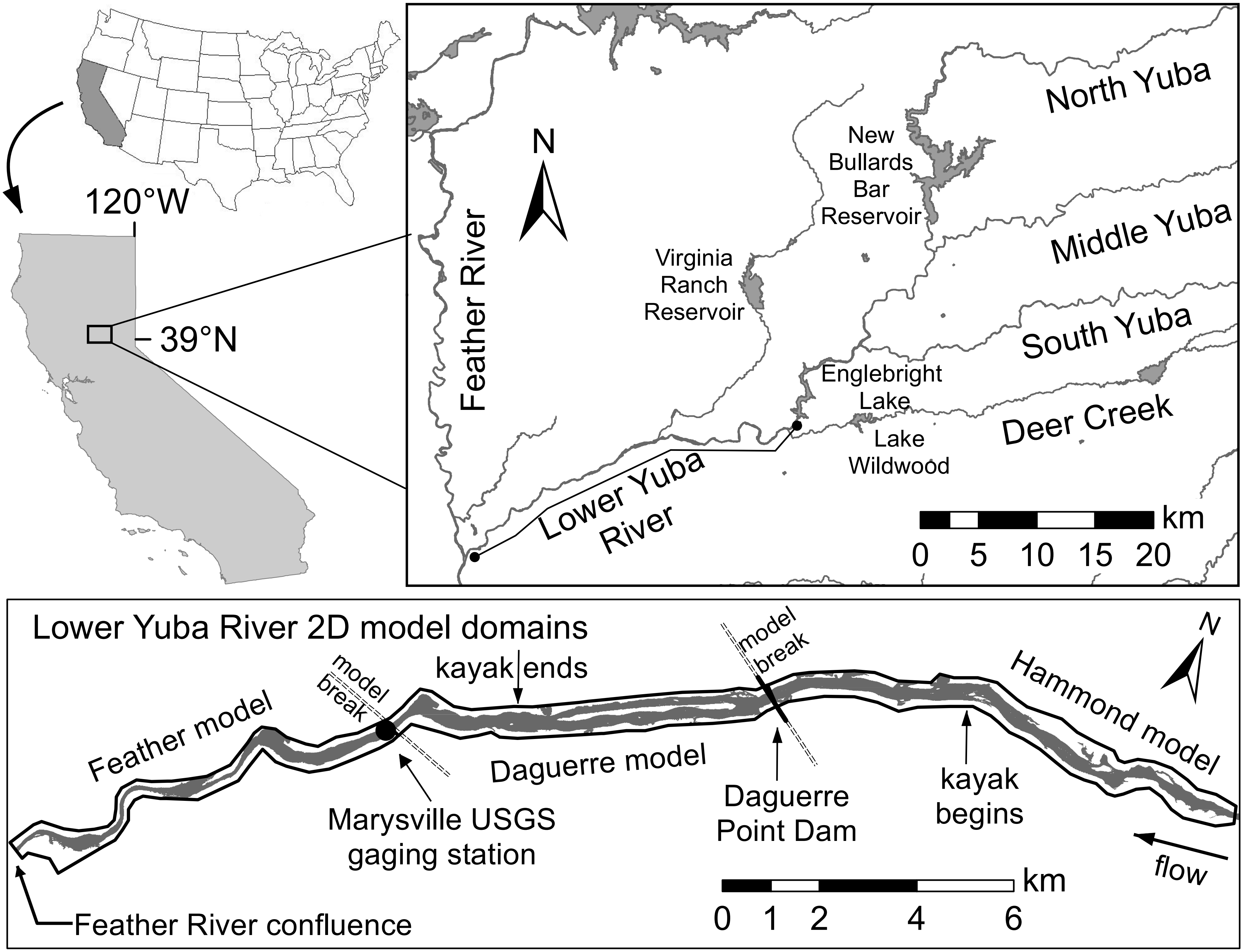

120°W
39°N
North Yuba
New
Bullards
Bar
Reservoir
N
Feather River
Virginia
Ranch
Reservoir
Middle Yuba
South Yuba
Englebright
Lake
Deer Creek
Lake
Wildwood
Lower Yuba
River
km
0 5 10 15 20
Lower Yuba River 2D model domains
model
break
kayak ends
model
break
N
Hammond model
Feather model
Daguerre model
Marysville USGS
gaging station
Daguerre
Point Dam
kayak
begins
flow
km
0 1 2 4 6
Feather River confluence

Figure 2

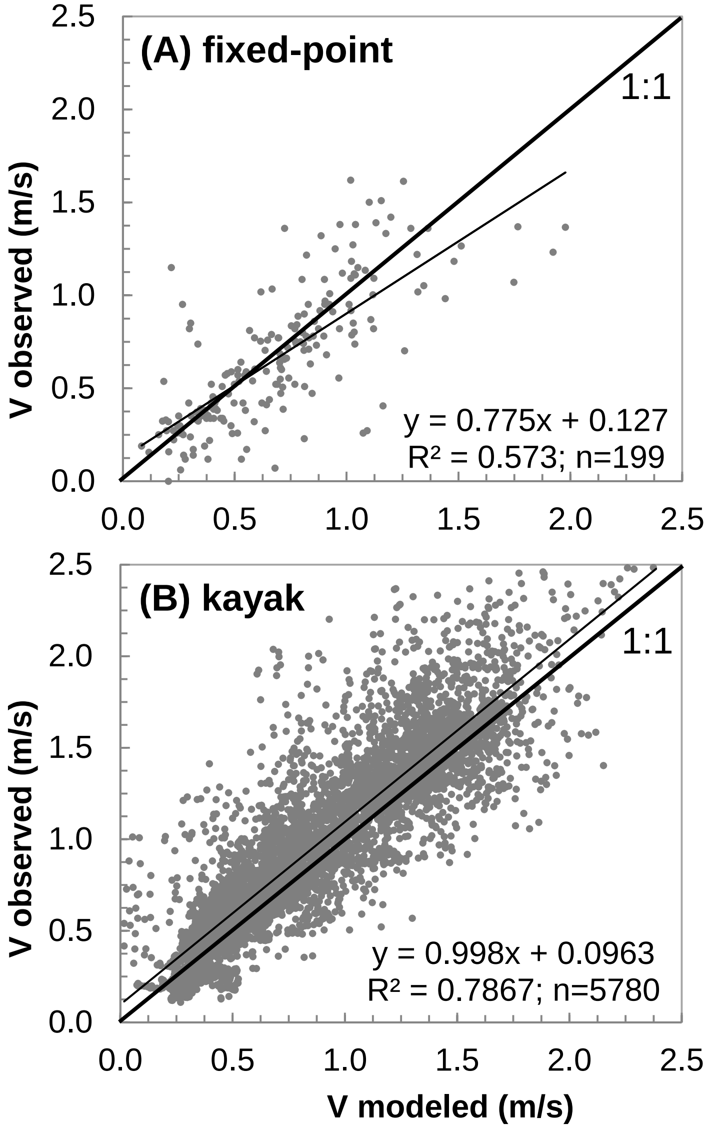
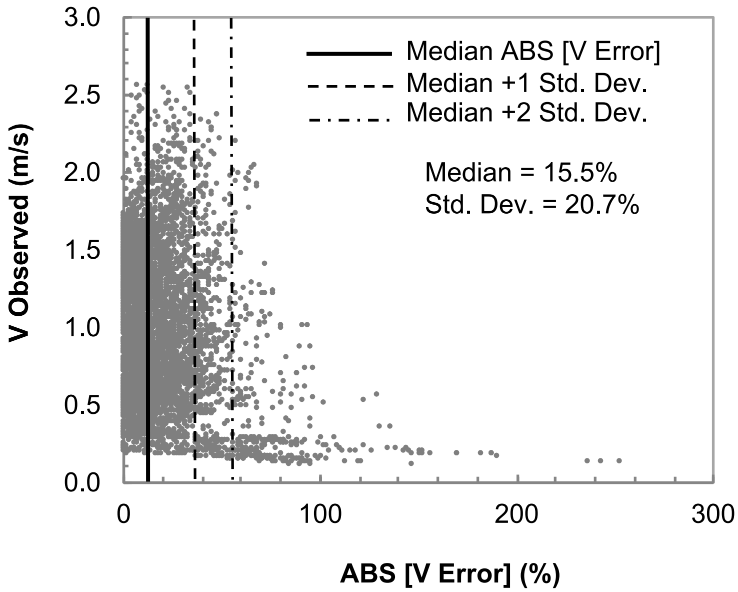
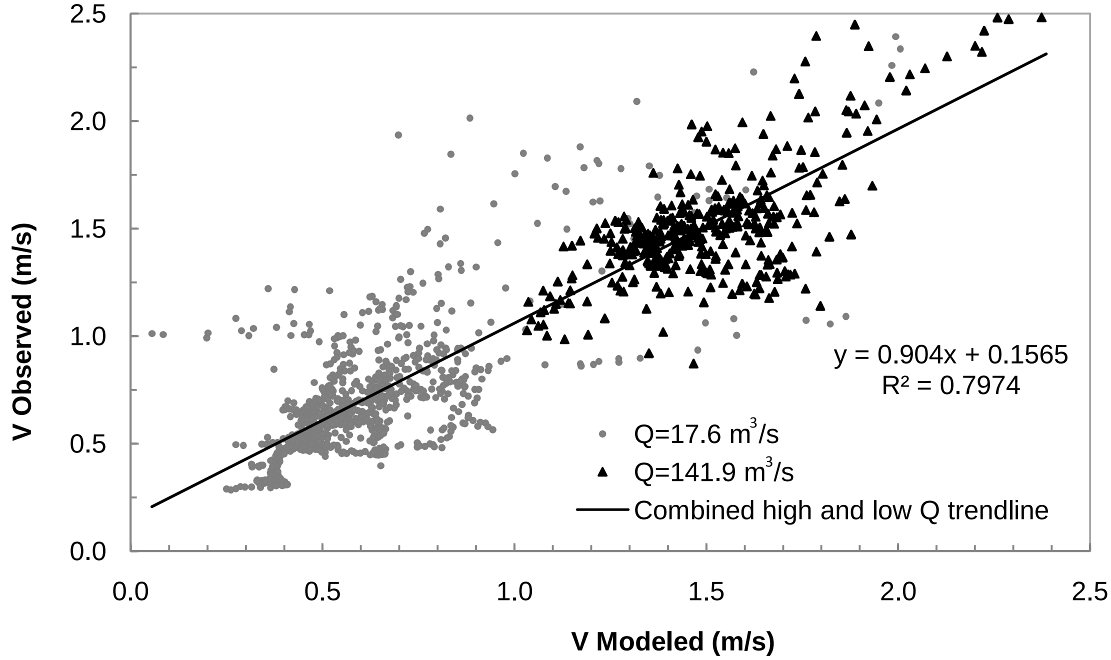

(A) fixed-point
1:1
y = 0.775x + 0.127
R² = 0.573; n=199
V observed (m/s)
0.0
0.5
1.0
1.5
2.0
2.5
(B) kayak
1:1
y = 0.998x + 0.0963
R² = 0.7867; n=5780
V observed (m/s)
V modeled (m/s)

Figure 3

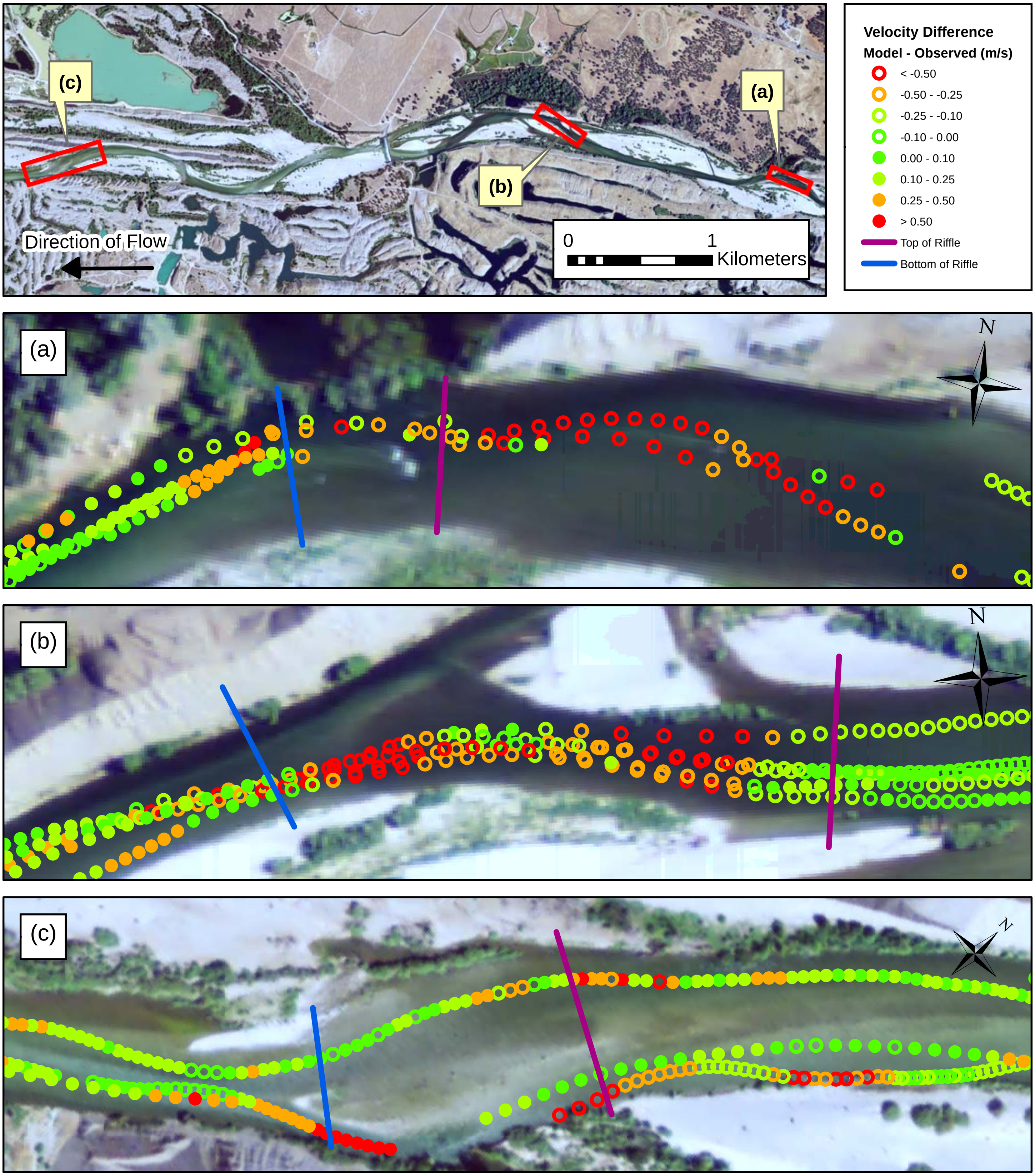


V Observed (m/s)
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
100
200
300
ABS [V Error] (%)
Median ABS [V Error]
Median +1 Std. Dev.
Median +2 Std. Dev.
Median = 15.5%
Std. Dev. = 20.7%



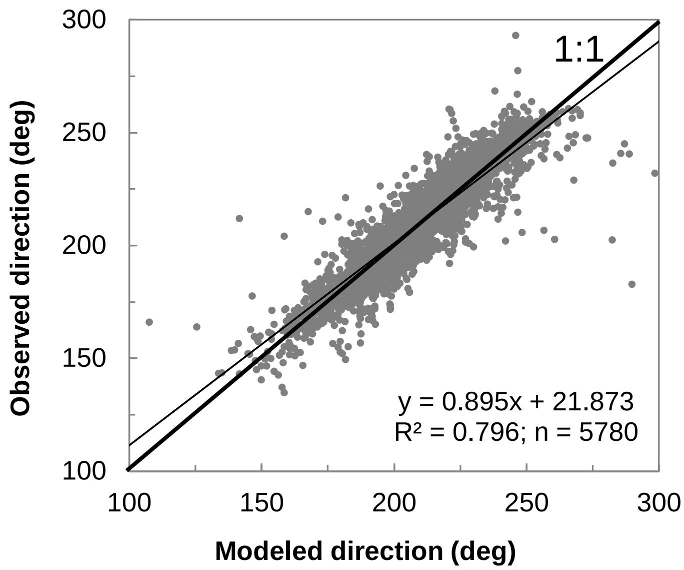
2.5
2.0
1.5
1.0
0.5
0.0
V Observed (m/s)
0.0
0.5
1.0
1.5
2.0
2.5
V Modeled (m/s)
y = 0.904x + 0.1565
R² = 0.7974
Q=17.6 m3/s
Q=141.9 m3/s
Combined high and low Q trendline

Figure 5



(c)
(a)
(b)
Direction of Flow
0
1
Kilometers
Velocity Difference
Model - Observed (m/s)
< -0.50
-0.50 - -0.25
-0.25 - -0.10
-0.10 - 0.00
0.00 - 0.10
0.10 - 0.25
0.25 - 0.50
> 0.50
Top of Riffle
Bottom of Riffle
(a)
N
(b)
N
(c)
N

Figure 6



1:1
y = 0.895x + 21.873
R² = 0.796; n = 5780
Observed direction (deg)
Modeled direction (deg)
300
250
200
150
100
100
150
200
250
300

Figure 7

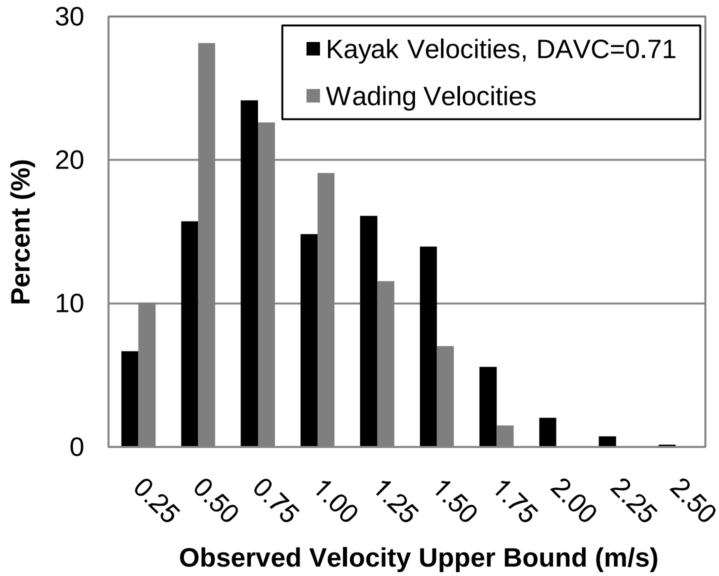

Kayak Velocities, DAVC=0.71
Wading Velocities
Percent (%)
30
20
10
0
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
2.25
2.50
Observed Velocity Upper Bound (m/s)

Figure 8

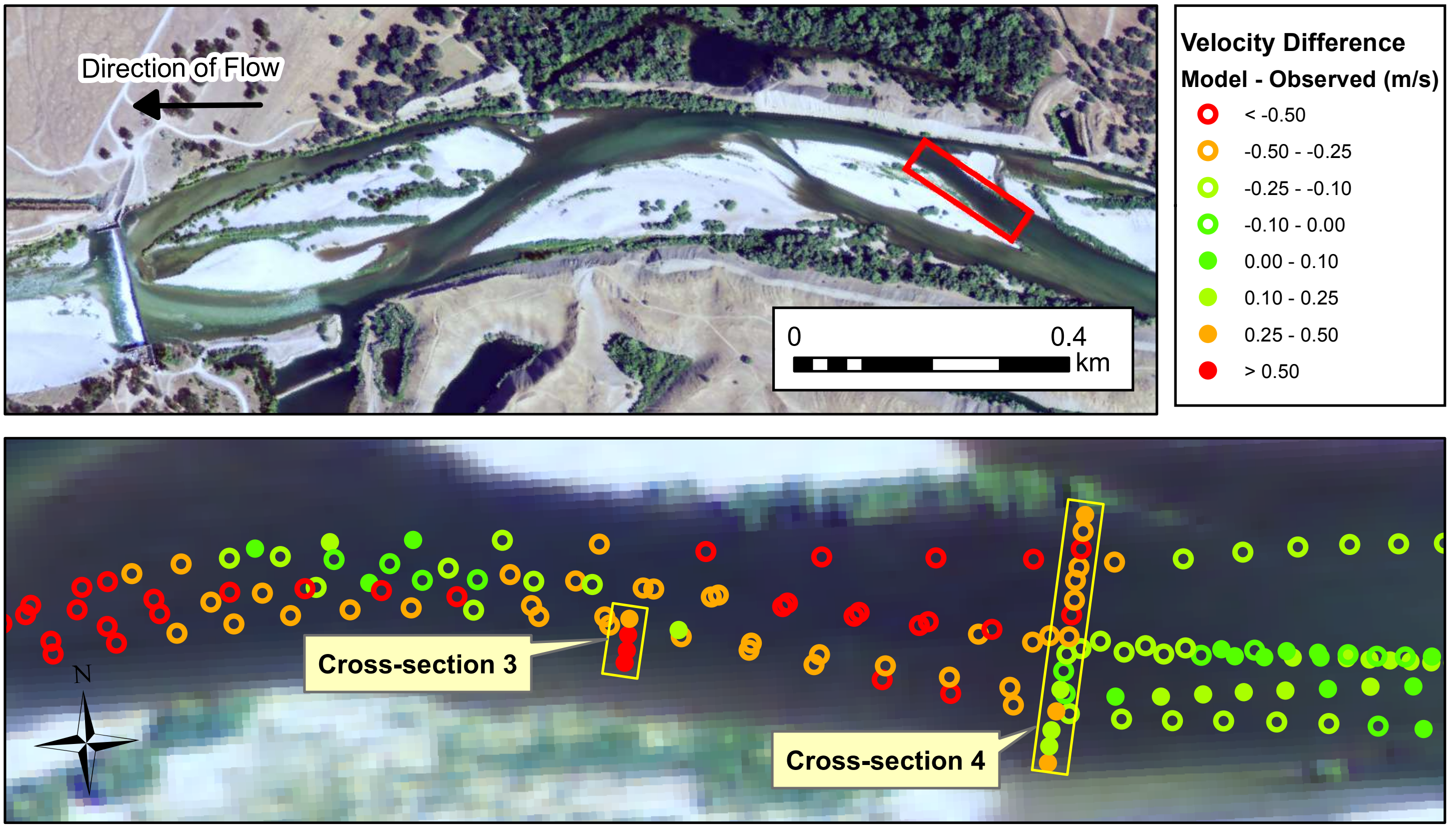

Direction of Flow
0
0.4
km
Velocity Difference
Model - Observed (m/s)
< -0.50
-0.50 - -0.25
-0.25 - -0.10
-0.10 - 0.00
0.00 - 0.10
0.10 - 0.25
0.25 - 0.50
> 0.50
Cross-section 3
Cross-section 4
N

Figure 9

Kayak drifter surface velocity observation for 2D hydraulic model validation

SUPPLEMENTARY MATERIALS

This file is organized with sections that parallel those in the article for convenience in locating text to see more details on any topic of interest.

1 Introduction Supplements

1.1 Need For More 2D Model Validation

Scientific progress benefiting from 2D modeling includes studies of river hydraulics (Waddle, 2010; Strom et al., 2016), environmental flows (Reinfelds et al., 2010), hydrogeomorphic processes (Pasternack et al., 2008; Sawyer et al., 2010), floodplain inundation (Tayefi et al., 2007), urban flooding (Yu & Lane, 2006; Mason et al., 2007), aquatic meso- and micro-habitat quality (Clark et al., 2008; Kammel et al., 2016), and river design assessment (Elkins et al., 2007; Lee et al., 2010; Brown & Pasternack, 2016). In general, 2D modeling studies are using very little data for validating velocity performance, and this means that models just aim for the simplest performance criteria with little analysis of where and when such models work well or not. Studies that exemplify what is commonly done with 2D model velocity validation include Ghanem et al., 1996; Lane et al., 1999; Stewart, 2000; Ballard & Gard, 2003; Wheaton et al., 2004; Pasternack et al., 2006; Elkins et al., 2007; Brown & Pasternack, 2008. Velocity is commonly observed at just 2-10 cross-sections, even when validating 1-10 km long models (see citations in previous paragraph for examples of this). Data collection usually involves standing in the flow with a 1D current meter or lowering a current meter on a cable from a boat to obtain time-averaged velocity measurements over 30-60 s at each of 1-3 locations in a vertical profile. This limits the locations and flows at which data are collected. Acoustic Doppler current profilers (ADCPs) are infrequently used for velocity measurements by moving across a measured section and creating a series of 5-s averaged vertical velocity profiles (e.g. Tiffan et al., 2002; May et al., 2009). From the literature, there may be 10-30 observations per cross-section, yielding ~ 50-300 total observation points to test a model, and that still leaves velocity direction untested.

Although 2D model validation using fixed-point velocity is well documented, there are important operational limitations. First, how many observations and in what pattern are needed to produce a satisfactory accounting of model performance? Many studies use fewer than 10 cross-sections and choose transects with relatively simple velocity vector patterns, as these are often locations required for discharge measurements. For example, Tiffan et al. (2002) used just two ADCP cross-sections (one “simple” and one “more complex”) at two discharges to validate a 33-km long 2D model. A few hundred points sounds like a lot, but given the breadth of uncertainty of real velocity conditions, not only are individual measurements error-prone, but the statistical sampling regime is inadequate for models spanning tens to hundreds of kilometres.

Large-scale 2D flow patterns are important to geomorphic processes and ecological conditions, and there is a pressing need to have the data to ensure that 2D models represent those correctly. The velocity vector field is the hallmark of 2D modeling, yet hardly anyone is validating such vectors in real-world applications. Velocity patterns can be significantly varied in a model by selection of a turbulence closure scheme and the associated parameter values. Overall, improvement of 2D model velocity validation is a key need to better characterize uncertainty in 2D models and improve the perception of 2D models by the public, which is often sceptical of numerical models (Ludwig, 2001; Van Asselt and Rotmans, 2002).

Second, collection of point-velocity measurements and use of an ADCP are too time-consuming and still have significant errors and uncertainties. Using a fixed-point velocity sensor, depth-averaged velocity data requires measurements made during 0.5–3 min per vertical, plus the time to set up and relocate. This creates an infeasible timeframe to collect a dense mesh of velocity points and span a large area at a reasonable cost. For ADCP, an individual vertical velocity profile only takes a minimum of 5 s (though > 30 s should be used to be consistent with standard fixed-point protocols), but transects should be made at a slower speed than the ambient velocity (20% of it), and transects need to be repeated 4–6 times to get a reasonable time-averaged estimate of the flow field. Proper recording of the orientation of the 3D vector in the flow field is challenging. Operation of boats for ADCP aimed at velocity mapping is more difficult than for discharge gaging, because it involves navigating through diverse conditions, including locations where ADCP performs poorly, such as along the bank, in shallows, and where there are air bubbles. The diversity of hydraulics also generates more challenges for and uncertainties in ADCP velocity accuracy. ADCP has much promise, if it is further developed for velocity mapping instead of discharge gaging, including thorough testing for this purpose. Mueller et al. (2007) present important errors with ADCP velocity measurements, while Lee et al. (2011) make a compelling case for the need of a low-cost velocity vector mapping solution instead of ADCP. Ultimately, ADCP remains expensive, complex, and inaccessible to many potential 2D model users, especially biologists who are increasingly using 2D models but don't want to become professional hydrographers as well.

Third, wading and motorboat approaches to data collection are limited by flow conditions precluding 2D model validation in the most important places to understand geomorphic processes. Observations by wading requires a low combination of depth and velocity as well as a lack of hazards around the wader (Milanesi et al., 2015). Motorboats have a much wider range of capabilities, but they have trouble in shallows, near flow obstacles, over rapids, and in some flood conditions. These challenges inhibit equal-effort sampling among diverse ecohydraulic conditions, which ought to be required for 2D model validation.

Finally, the scale of velocity observation and prediction are often incommensurate. Velocity sensors sample 0.5–100 mm, while 2D model grid cells are 500–10,000 mm, creating a mismatch for comparison, especially in light of turbulence, whose intensity also varies spatially and over these scales. ADCP units have a trade-off with range versus resolution (e.g. Sontek/YSI Inc, 2007), making them ineffective in rivers whose spatially variable depths and velocities transcend the selected range and resolution parameters. Historically the problem was particularly bad for shallow rivers, given that an ADCP cannot measure velocity close to the water's surface due to its blanking distance. However, range and resolution options are improving. They also require precise orientation in the 2D model's coordinate system, which is problematic with bottom tracking and differential GPS. Moreover, using a 3D velocity profiler to validate a 2D model can introduce errors due to differing averaging methods of varying spatial scales (Pasternack et al., 2006).

1.2 Large-Scale Particle Image Velocimetry

Some additional LSPIV studies include Creutin et al. (2003), Hauet et al. (2009), Fujita & Kunita (2011), Lewis & Rhodes (2015), Detert & Weitbrecht (2015), and Creëlle et al. (2016). For brevity, these citations had to be cut out of the article. More articles are likely available, especially including conference proceedings, as this a rapidly emerging area of inquiry. LSPIV has significant challenges for use in widespread velocity vector validation of 2D models. Platforms are commonly operated in clear-sky weather over smaller channels accessible to camera mounting systems, whereas 2D model validation is needed for a wide range of river sizes with limited accessibility during a wide range of flows and precipitation intensity. Fixed-view domains at the scale of <100 m long by 70 m wide used in LSPIV have limited applicability for 2D models that span 10-100 km long by 10-500 m wide channels, which is a primary reason why oceanographers use drifters. The legal and regulatory context for aerial platforms is presently uncertain. Regardless of acquisition platform, LSPIV methods require substantial data processing with many additive potential sources of uncertainty at this time. Data acquisition and processing to obtain surface velocity vectors over tens to hundreds of kilometres is not presently available.

1.3 Kayak Surface Particle Tracking

There are only two journal articles (Stockdale et al., 2007; Han et al., 2016) and a few conference proceedings (Swick and MacMahon, 2009; Emory et al., 2010; Lee et al., 2011) testing drifters in nontidal rivers, and they all used differential GPS. None of these compared drifter results against fixed-point observations, though some report past studies of error estimates. The purpose of this supplemental section is to provide background information on those past studies, help readers understand the accuracy of kayak surface particle tracking method (using RTK GPS) in more detail, and explain why there is value in selecting RTK GPS over regular or differential GPS, when possible.

Several of the past efforts have primarily served to report technological developments and secondarily to share descriptive data from test deployments. Stockdale et al. (2010) developed flat, packaged river surface drifters with on-board differential GPS. They produced a velocity vector map of a 400-m long test site. Swick and MacMahan (2009) developed cylindrical differential GPS drifters with ballast to keep them upright. They deployed 6-16 drifters in 3 settings to produce both Lagrangian and Eulerian (averaged) statistical analyses. Emery et al. (2010) developed commercial spherical river drifters. They deployed five of them at one test site and one at another test site. Lee et al. (2011) developed another differential GPS drifter technology similar to that of Stockdale et al. (2010), but then in their research they added a thorough data-processing framework. They used drifters to collect 45 trajectories with ~ 7000 observations to map velocity vectors in a 100 m long by 10-15 m wide channel. Finally, Han et al. (2016) developed a spherical drifter similar to that of Emery et al. (2010). They deployed 50 drifters to analyze riverine surface mixing characteristics.

In general, with a good satellite configuration, GPS has ~ 3-m horizontal accuracy, while differential GPS has ~ 1-m horizontal accuracy. For an ambient current of 1 m/s moving in a straight line, two positions that are off by 1 or 3 m extra away from each other (i.e., the worst-case scenario) yield a velocity error of 300% or 700%, respectively. In contrast, an RTK GPS with 0.02-m or 0.05-m accuracy has a worse-case velocity error of just 4% or 10%, respectively. That is an enormous difference, especially for 2D model validation. Our viewpoint is that most stakeholders and independent reviewers expect observational data to be of a high-quality to serve for model validation purposes. Thus, the extra cost of RTK GPS is necessary. Other potential uses of river drifters may not require this. However, the cost of RTK GPS has dropped so much over the last decade that it is getting difficult to justify avoiding the expenditure. In regions where RTK GPS is still excessively expensive or outright unavailable, more work would be required to ascertain if DGPS is satisfactory for 2D model validation purposes.

Stockdale et al. (2008) used a Royaltek™ BlueGPS (RBT3000) that samples position at 1 Hz, with a reacquisition time of 0.1 s for each reading on average. This unit a built-in differential capability using the United States' Federal Aviation Administration Wide Area Augmentation System (WAAS) as well as Europe's European Geostationary Navigation Overlay Service (EGNOS). These systems have similar specifications with their corrections. Differential corrections are provided every ~ 0.2 Hz. WAAS provides a conservative position accuracy of 3 m or better (for both lateral and vertical measurements) at least 95% of the time (Zhang et al., 2014). It is reported with different actual performances in horizontal accuracy, typically in the 1-3 m range (e.g., Ariens Specialty Brands LLC, 2014).

Lee et al. (2011) used a MediaTek GPS module FPGMMOPA1. The technical specifications for this model number are no longer available on the MediaTek website, but the authors report a positional accuracy of 3 m, which is consistent with native GPS performance absent any differential correction. Lee et al. (2011) state a measured average velocity "error" of 0.175 m/s comparing drifters versus fixed-point observations.

Both Stockdale et al. (2008) and Lee et al. (2011) both applied additional statistical methods to help diminish the effects of individual point velocity estimation errors on bulk results for larger subsets of data. Stockdale et al. (2008) simply divided the river into cells and then computed the average velocity in each section. Lee et al. (2011) also divided the river into cells, but they computed a diversity of statistics to characterize motion of the bulk flow field. Such methods allow for more relaxation of point-scale expectations as long as the bulk field is acceptable to within specified tolerances for the derived statistics. It remains to be seen what performance metrics will be acceptable to the scientific and practitioner 2D modeling community for such approaches, but this is a good start. Overall, Stockdale et al. (2008) and Lee et al. (2011) provide enough discussion of velocity accuracy to insure that the method is worthwhile to develop further, and there is now ample justification for new studies to carefully test drifters more comprehensively.

1.4 Study Objectives

None.

2 Study Site Supplements

None.

3 Methods Supplements

Field data collection efforts were explicitly intended to characterize geomorphic, hydrologic and hydraulic attributes of the LYR at roughly meter-scale resolution in support of a near-census approach to river assessment, including 2D hydrodynamic modeling. The types of data collected included topography and bathymetry (Pasternack, 2009; White et al., 2010; Carley et al., 2012) as well as hydraulic data: water surface elevation, depth, velocity magnitude and velocity direction (Barker, 2011; Pasternack et al., 2014). Details about spatial coverage, resolution and accuracy for the digital elevation model (DEM) and 2D hydrodynamic modeling used in this study are provided below.

3.1 2D hydrodynamic modeling details

This study only used the portion of the LYR topographic map from the highway 20 Bridge down to the confluence with the Feather River. On September 21, 2008 Aero-Metric, Inc. (Seattle, WA) acquired Light Detection and Ranging (LiDAR) data of the river corridor during a constant low flow typical of the period when hydro facility maintenance takes place (24.437 m^3/s (863 cfs) between Englebright Dam and Daguerre Point Dam; 17.613 m^3/s (622 cfs) below Daguerre Point Dam where irrigation diversions occur). The target point spacing was 0.74 m. Compared against 8769 ground-based RTK GPS observations of elevation along flat surfaces, 84.7% of LiDAR points were within 0.06 m, another 14.0% were within 0.12 m, and almost all of the remained of the data were within 0.18 m. All in-water LIDAR points were removed from the dataset using a shoreline boundary polygon.

Boat-based bathymetric surveys were performed during low flows in August and September of 2008 as well as during higher flows in March and May 2009 by Environmental Data Solutions. An array of four echosounders were stationed across the bow with ~1.8 m spacing and used to obtain longitudinal swaths of bathymetric points. It was cost-prohibitive to wade all wetted areas inaccessible to the boat, but a set of key “data gap” locations was identified based on iterative map production checks and filled in by ground-based RTK GPS surveying in November 2009.

A comprehensive set of uncertainty analyses were performed to ensure that the datasets were accurate and intercomparable (Barker, 2010). Ground points on the uneven natural surface were compared between ground-based and boat-based surveys, ground-based and LiDAR surveys, and boat-based and LiDAR surveys. Surveys were also inter-compared at carefully surveyed water surface elevation shots along the water’s edge, where there was less surface variability. Vertical datums were checked between survey methods to ensure compatibility.

After the final gap-fill survey, all the points were brought into ArcGIS® 9.3.1 software, visualized as a map, and edited to remove any obvious errors. In narrow backwater channels and other gap-fill areas that contained interpolation errors, breaklines were created to better represent landform features. Additionally, some large areas that contained very few points were augmented so that channel characteristics were maintained. Finally, using the spatial join function in ArcGIS, consistency between data sets was assessed to ensure there were not any major discrepancies between different mapping methods.

Basic information describing topographic and bathymetric field data in the Yuba River in the areas investigated in this study are reported in the box below.

Attribute	Description
Aerial extent	From Highway 20 Bridge to confluence with Feather River
Years of data collection	Most bathymetry was mapped in late August to early September 2008, with some high-flow data collection in March and May 2009 as well as small additional near-bank and near-DPD gaps mapped in November 2009. Ground-based topographic surveys were done in November 2008 and November 2009. LiDAR of the terrestrial river corridor was flown on September 21, 2008.
Bathymetric Resolution	Within the 24.92 m^3/s inundation area, points were collected along longitudinal lines, some cross-sections and some localized grids. The average grid point spacing is one point every 1.3 m. (59.8 pts/100 m^2).
Topographic Resolution	Outside the 24.92 m^3/s inundation area, points were mostly collected with lidar, yielding an average grid point spacing of one point every 0.43 m. (554 pts/100 m^2).
Bathymetric Accuracy	Comparison of overlapping echosounder and total station survey points at one site yielded observed differences of 50% within 0.15 m, 75% within 0.18 m and 94% within 0.3 m. Comparison of boat-based water edge shots versus RTK GPS surveyed water's edge shots yielded observed differences of 75% within 0.03 m, 91% within 0.061 m and

	99% within 0.15 m.
Topographic Accuracy	Compared against 8,769 ground-based RTK GPS observations of elevation along flat surfaces, 54% of LIDAR points were within 0.03 m , 86% were within 0.061 m and virtually all of the data were within 0.15 m. Regular total station control point checks yielded accuracies of 0.0091-0.018 m. RTK GPS observations had vertical precisions of 0.018 m. Comparison of lidar water edge points versus the same for RTK GPS yielded observed differences of 30% within 0.03 m, 57% within 0.061 m and 92% within 0.15 m.

In this study, the Sedimentation and River Hydraulics – Two-Dimensional Version 2.1 (SRH-2D v2.1) model was used to simulate river hydraulics as well as predict flow velocities and directions (Lai, 2008). SRH-2D v2.1 focuses specifically on 2D modeling of river systems by using the depth averaged St. Venant equations:

$$\frac{\partial h}{\partial t} + \frac{\partial hU}{\partial x} + \frac{\partial hV}{\partial y} = e \quad (1)$$

$$\frac{\partial hU}{\partial t} + \frac{\partial hUU}{\partial x} + \frac{\partial hVU}{\partial y} = \frac{\partial hT_{xx}}{\partial x} + \frac{\partial hT_{xy}}{\partial y} - gh\frac{\partial z}{\partial x} - \frac{\tau_{bx}}{\rho} + D_{xx} + D_{xy} \quad (2)$$

$$\frac{\partial hV}{\partial t} + \frac{\partial hUV}{\partial x} + \frac{\partial hVV}{\partial y} = \frac{\partial hT_{xy}}{\partial x} + \frac{\partial hT_{yy}}{\partial y} - gh\frac{\partial z}{\partial y} - \frac{\tau_{by}}{\rho} + D_{yx} + D_{yy} \quad (3)$$

where t is time, x and y are horizontal Cartesian coordinates, h is water depth, U and V are depth-averaged velocity components in x and y directions, respectively, e is excess rainfall rate, g is gravitational acceleration, T_{xx}, T_{xy}, and T_{yy} are depth-averaged turbulent stresses, D_{xx}, D_{xy}, D_{yx}, and D_{yy} are dispersion terms due to depth averaging, $z = z_b + h$ is water surface elevation, z_b is bed elevation, ρ is water density, and τ_{bx}, τ_{by} are the bed shear stresses (friction). Dispersion terms represent the conversion of kinetic energy into internal energy by viscous shear stress since small scale turbulence is not fully represented in this model. Bed friction is calculated using the Manning's roughness equation as follows:

$$\begin{pmatrix}\tau_{bx}\\ \tau_{by}\end{pmatrix} = \rho C_f \begin{pmatrix}U\\ V\end{pmatrix}\sqrt{U^2+V^2}; \quad C_f = \frac{gn^2}{h^{\frac{1}{3}}} \qquad (4)$$

where *n* is the Manning's roughness coefficient.

In order to solve these St. Venant equations a finite-volume numerical method was used which requires the use of a computational mesh. The Surface-water Modeling System® (SMS) version 10.1 graphical user interface (Aquaveo, LLC, Provo, UT) was used to produce the meshes. A hybrid structured-unstructured, arbitrarily-shaped mesh was produced in SMS using both quadrilateral and triangular elements. Typical nodal spacing for the mesh was ~1.5 m. Although this mesh size is smaller than the resolution of bathymetric data collected by the boat, several studies have shown that reasonable results can still be predicted in areas with sparse topographic coverage (Anderson and Bates, 1994; French and Clifford, 2000; Marks and Bates, 2000). Based on past experience with mesh-resolution testing to evaluate issues such as numerical diffusion and numerical stability, this mesh resolution is more than high enough to avoid those problems for the finite-volume method. Topographic (x,y,z) points and breaklines were then imported from ArcGIS® software into SMS software where they were then used to interpolate elevations to the mesh.

Because of the large size of the study area, the desire for computational efficiency, meter-scale model resolution commensurate with the best of the available topographic data, and the order-of-magnitude range of flows being assessed, the river between the Highway 20 Bridge and the Feather River confluence was split into three model reaches: (1) Hammon Reach [HR – HWY 20 Bridge to DPD], (2) Daguerre Reach [DGR – DPD to Marysville Gauging Station], and (3) Feather Reach [FR – Marysville Gauging Station to Feather River Confluence]. These do not correspond with the geomorphic reach delineation for the LYR, but are purely for computational balance and efficiency. Higher and lower flow meshes (corresponding with flow ranges of ~42-141 m^3/s and <42 m^3/s, respectively) were then created for each model reach to limit the amount of unnecessary dry areas being modeled in the lower-flow simulations. Models were run for 28 steady flows ranging from 8.50 to 3,126 m^3/s (0.06 to 22 times bankfull discharge).

Channel roughness was primarily addressed during the mapping effort by creating a highly detailed DEM, but unresolved roughness issues were addressed by using a global Manning's *n* coefficient. The use of a constant roughness-coefficient value for unvegetated substrate was justified on the LYR, because the New Years 2006 flood with a peak flow of ~ 3087 m^3/s (~$19.5 \cdot Q_{bankfull}$) was observed to erase much of the spatial variability in bed surface grain size distribution. This occurs because the underlying valley fill is composed of relatively homogeneous hydraulic mining sediment deposited prior to the construction of Englebright Dam (Pasternack, 2008). Substrate observation after the even larger 1997 flood found that it took 5-7 years for an armor layer to re-establish itself in riffles and chutes. Flows < 141 m^3/s are mostly (but not always) constrained inside the bankfull channel dimensions devoid of vegetation, rendering it unnecessary to account for vegetation in the roughness coefficient in this study. It is possible that adding additional roughness for near-bank vegetation and steep banks would improve hydraulic prediction, but without a systematic, objective approach to distributing near-bank roughness at the time this study was conducted, it is beyond the scope of this investigation. Observed WSE values obtained during the gap-fill survey were compared against model-predicted WSE values for different simulations using n=0.035, 0.04, and 0.045 for all model reaches. Statistical error and histogram analyses comparing the different values found that 0.04 yielded the most accurate predictions, so that value was used in the model validation investigations reported in this study.

Aside from channel roughness, turbulence closure and the model's boundary conditions must also be assigned. Turbulence closure was achieved using the parabolic/zero equation model (Lai, 2008), with eddy viscosity varying as a function of depth and shear velocity, modified by an eddy viscosity coefficient of 0.6 based on the long-standing representative value taken from historical dye studies (Fischer et al., 1979). The eddy viscosity term is a practical strategy for calculation that ignores the small-scale vortices in the motion and calculates a large-scale motion with eddy viscosity which characterizes the transport and dissipation of energy in the smaller-scale flow. This approach and coefficient was previously validated for use in a pool-riffle-run sequence on this same river (Moir and Pasternack, 2008; Sawyer et al., 2010). Eddy viscosity should not be confused with eddy diffusivity, which is used with any dependent variable (a scalar or a component of a vector) in the discretization of the governing equations to represent the process by which substances are mixed.

The boundary conditions required for SRH-2D are input flows and corresponding exit WSE. Input flows were obtained from the Yuba River at Smartville (#11418000), Deer Creek at Mooney Flat Road (#11418500), and Yuba River at Marysville (#11421000) United States Geological Survey (USGS) gaging stations. Corresponding downstream WSE values were collected in two ways. Since the DGR downstream boundary was located at the Marysville gauging station, WSE values were readily available from the professional rating relation for that gage. For the other model reaches, downstream WSE values were observed using a suitably mounted Level TROLL® 500 water level sensor (In-Situ Inc., Fort Collins, Colorado). The lag time between gage-recorded discharges and local WSE was optimized for by statistically matching fluctuations in corresponding records to yield the highest correlation between lagged records.

Model simulations were comprehensively validated for flows ranging over an order of magnitude of discharge (0.1 to 1.0 times bankfull) using three approaches: (i) traditional cross-sectional validation methods, (ii) comparison of LiDAR-derived water surface returns against modeled water surface elevations and (iii) Lagrangian particle tracking with RTK GPS to assess the velocity vectors. Model set-up and performance details are reported in the box below:

Attribute	Description
Model domains	For this study, there were 3 modeling reaches to make the computational process more efficient. They are given the abbreviations, HR, DGR and FR below.
Computation mesh type	All model domains use an unstructured mesh with triangular and polygonal elements.
Computational Mesh Resolution	HR: For flows 0-36.81 m^3/s, 0.91 m internodal spacing. For flows 36.81-212.4 m^3/s, 1.5 m internodal spacing. For flows >283.2 m^3/s, 3 m internodal spacing. DGR: For flows 0-36.81 m^3/s, 1.5 m internodal spacing. For flows 36.81-212.4 m^3/s, 1.5 m internodal spacing. For flows >283.2 m^3/s, 3 m internodal spacing. FR: For flows 0-36.81 m^3/s, 1.5 m internodal spacing. For flows 36.81-212.4 m^3/s, 1.5 m internodal spacing. For flows >283.2 m, 3 m internodal spacing.
Discharge Range of Model	8.495 to 3126 m^3/s.
Downstream WSE data/model source	HR: Continuous direct observation of WSE at flows <~623.0 m^3/s. For higher flows the downstream WSE was taken as the upstream WSE from the HR model at that flow. DGR: Reach ends exactly at Marysville gaging station, so the WSE data is of the highest quality and abundance. Continuous WSE data for all flows ~14.16 - 3126 m^3/s. FR: Continuous direct observation of WSE at flows <~623.0 m^3/s. For higher flows the downstream WSE was set to yield an upstream WSE equal to that at the Marysville gage.
River roughness	Because the scientific literature reports no consistent variation of Manning's n as a function of stage-dependent relative roughness or the whole wetted area of a

specification	river (i.e., roughness/depth), a constant value was used for all unvegetated sediment as follows: 0.04 for the HR, DGR, and FR models (based on validation testing of 0.03, 0.035, 0.04, 0.045 and 0.05 as possible options). This study did not use spatially distributed vegetated roughness, because the majority of flows were in-channel and it was not warranted.
Eddy viscosity specification	Parabolic turbulence closure with an eddy velocity that scales with depth, shear velocity and a coefficient (e_0) that can be selected between ~0.05 to 0.8 based on expert knowledge and local data indicators. $Q<283.2\ m^3/s$: $e_0 = 0.6$ $Q\geq283.2\ m^3/s$: $e_0 = 0.1$
Hydraulic Validation Range	Point observations of WSE were primarily collected at 24.92 m^3/s, with some observations during higher flows, but not systematically analyzed. Velocity observations were collected for flows ranging from 15.01-141.9 m^3/s. Cross-sectional validation data collected at 22.65 m^3/s above DPD and 15.29 m^3/s below DPD.
Model mass conservation (Calculated vs Given Q)	0.001 to 1.98 %
WSE prediction accuracy	At 24.92 m^3/s there are 197 observations. Mean raw deviation is -0.0018 m. 27% of deviations within 0.03 m, 49% of deviations within 0.076 m, 70% within 1.5 m, 94% within 0.3 m. These results are better than the inherent uncertainty in LiDAR obtained topographic and water surface elevations.
Depth prediction accuracy	From cross-sectional surveys, predicted vs observed depths yielded a correlation (r) of 0.81.
Velocity magnitude prediction accuracy	5780 observations yielding a scatter plot correlation (r) of 0.887. Median error of 16%. Percent error metrics include all velocities (including V <0.3 m/s, which tends to have high error percentages) yielding a rigorous standard of reporting.
Velocity direction prediction accuracy	5780 observations yielding a scatter plot correlation (r) of 0.892. Median error of 4%. Mean error of 6%. 61% of deviations within 5 degrees and 86% of deviations within 10 degrees.

Using the workflow of Pasternack (2011), SRH-2D model outputs were processed to produce rasters of depth and velocity within the wetted area for each discharge. The first task involved creating the wetted area polygon for each discharge. To do this, point files of depth results were first converted to triangular irregular networks (TIN) and then to a series of 0.9144-m hydraulic raster files. Depth cells greater than zero were used to create a wetted area boundary applied to all subsequent hydraulic rasters. Next, the SRH-2D hydraulic outputs for depth and depth-averaged velocity were converted from point to TIN to raster files within ArcGIS 10.1 staying within the wetted area for each discharge. The complete dataset was a series of 0.9144-m resolution hydraulics rasters derived from SRH-2D hydrodynamic flow simulations at the following discharges: 8.5, 9.9, 11.3, 12.7, 15.0, 17.0, 17.6, 19.8, 22.7, 24.9, 26.3, 28.3, 36.8, 42.5, 48.1, 56.6, 70.8, 85.0, 113.3, 141.6, 212.4, 283.2, 424.8, 597.5, 849.5, 1195.0, 2389.9 and 3126.2 m^3/s.

Despite best efforts with modern technology and scientific methods, the 2D models used in this study have uncertainties and errors. Previously it has been reported that 2D models tend to underrepresent the range of hydraulic heterogeneity that likely exists due to insufficient topographic detail and overly efficient lateral transfer of momentum (Pasternack et al., 2004; MacWilliams et al., 2006). For this study those deficiencies result in a conservative outcome, such that there could be more fine details to the sizes and shapes of peak velocity patches than what is revealed herein. Overall, this study involves model-based scientific exploration with every effort made to match reality at near-census resolution over tens of kilometers of river length using current technology, but recognizing that current models do have uncertainties.

3.2 Fixed-Point Hydraulic Data

WSE observations were collected along the LYR between the Highway 20 Bridge and the Feather River confluence using a Leica® System 1200 RTK-GPS (ΔH= 1 cm, ΔZ=2 cm). In total, 199 points were used to compare observed and modeled WSE values. Measurements were collected over the course of two weeks in November 2009. During this period flows remained constant at 24.9 m^3/s above DPD and 15.0 m^3/s below.

A total of 199 point-based field observations of depth and velocity were made at ~ 2 m intervals along 17 cross sections (Fig. S1) on December 8-10, 2009. Seventeen cross-sections constitute a large number relative to published journal articles that report 2D model validation (see citations in the article's section 1.1), but a small number to assess a 36-km-long model domain. Discharge above Daguerre Point Dam was 22.521 and 22.57 m^3/s, but 15.26 m^3/s below it due to irrigation diversions. Cross-sections were chosen based on whether or not they were wadable and also if the given cross-section spanned a wide range of velocities. Wherever possible, measurements spanned the full channel, but in some cases, cross-sections became un-wadable and measurements were only made as far out into the current as possible. This is a common problem limiting 2D model validation. The water surface elevation was also measured at the water's edge on either side of the river. Point-velocity measurements were made using either a Marsh-McBirney® Flo-Mate (±33 mm s^{-1} root mean square error) electromagnetic current meter sampling at 30 Hz or a Price AA mechanical impellor current meter (Fulford, 2001). Both methods averaged velocity measurements over 40 s with sensors positioned at 0.6 of the depth to obtain a measure of the depth-averaged velocity (Buchanan and Somers, 1969; Rantz, 1982; Smart, 1999; Pasternack et al., 2006). Depths for all points were measured using the depth-setting wading rods equipped with the velocity sensors. For the Marsh-McBirney Flo-Mate wading rod, depths were measured to a resolution of ±1 cm, while the Price AA flow meter wading rod depths were measured to resolution of ±3 cm.

3.3 Kayak Velocity Measurement

Kayak velocity data was collected at six discharges, including 17.61, 23.13, 30.95, 105.3, 114.9, and 141.9 m^3/s. Water surface velocity (V_t) at time *t* was calculated by determining the total horizontal displacement from one point to the next and then dividing that displacement by the change in time (Δt) between position measurements:

$$\Delta \mathrm{N_t} = \mathrm{N_t} - \mathrm{N_{t-1}} \qquad \Delta \mathrm{E_t} = \mathrm{E_t} - \mathrm{E_{t-1}} \tag{5, 6}$$

$$\mathrm{dH_t} = \sqrt{(\Delta \mathrm{N_t})^2 + (\Delta \mathrm{E_t})^2} \tag{7}$$

$$\mathrm{V_t} = \frac{\mathrm{dH_t}}{\Delta \mathrm{t}} \tag{8}$$

where N_t is the northing, E_t is the easting, and dH_t is the total horizontal displacement at time *t*. This velocity was then assigned to a horizontal coordinate (X_t, Y_t) that is located at the midpoint of the observed positions used to calculate the velocity.

$$\mathrm{X_t} = \mathrm{E_{t\text{-}1}} + \frac{\Delta \mathrm{E_t}}{2} \tag{9}$$

$$\mathrm{Y_t} = \mathrm{N_{t\text{-}1}} + \frac{\Delta \mathrm{N_t}}{2} \tag{10}$$

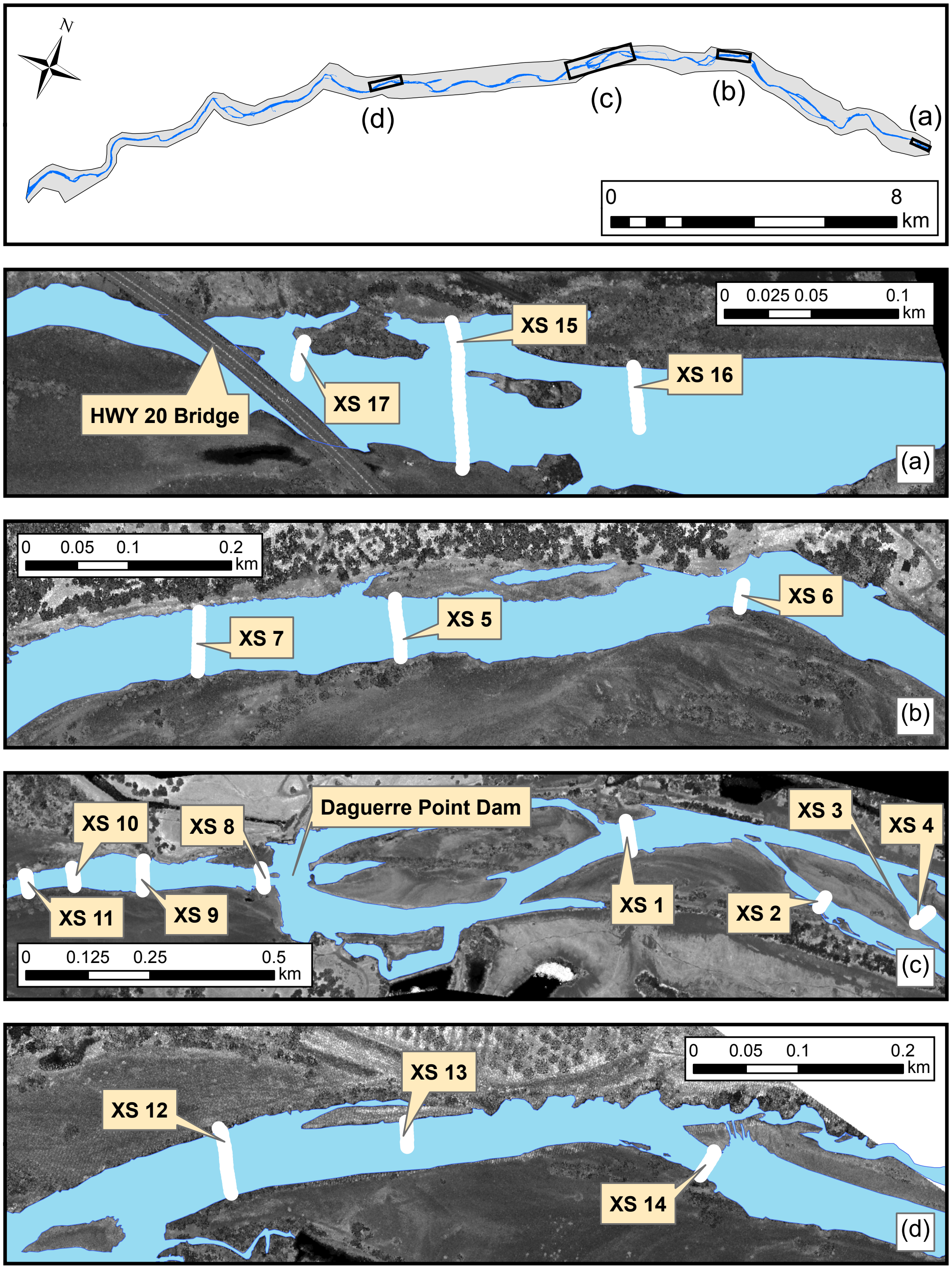
Finally, in order to create a velocity vector, a direction value (θ_t) was assigned to each coordinate using the differences in the northing and easting of consecutive position measurements:

$$\theta_t = (\tan^{-1}\frac{\Delta N_t}{\Delta E_t}) * \frac{180°}{\pi} \qquad \Delta E_t > 0,\ \Delta N_t > 0 \tag{11a}$$

$$\theta_t = \left\{(\tan^{-1}\frac{\Delta N_t}{\Delta E_t}) * \frac{180°}{\pi}\right\} + 360° \qquad \Delta E_t > 0,\ \Delta N_t < 0 \tag{11b}$$

$$\theta_t = \left\{(\tan^{-1}\frac{\Delta N_t}{\Delta E_t}) * \frac{180°}{\pi}\right\} + 180° \qquad \Delta E_t < 0,\ \Delta N_t < 0 \tag{11c}$$

where the direction is in degrees. When $\theta_t = 0$, that means the vector is directed due east.

371 

372 **Figure S1. Locations of the 17 cross-sections where wading velocity measurements were**

373 **made.**

3.4 Data Analysis

3.4.1 Velocity Data Preparation

None.

3.4.2 Model Performance Indicators

Among 2D hydrodynamic modeling studies, there is a common suite of validation metrics used in the peer-reviewed literature. Typical metrics include (i) basic statistical measures (i.e. mean, median, and standard deviation) for signed and unsigned deviations as well as those for signed and unsigned percent differences; (ii) regression analysis results in terms of slope, y-intercept, r^2, and p-value; and (iii) the same as the previous two, but segregating velocity data into small versus large values, as values $< \sim 0.5$ m/s tend to have disproportionately higher error percentages. There is no systematic agreement as to the performance indicators for these metrics for 2D modeling, except that Ballard et al. (2010) proposed that r^2 should be > 0.36 (i.e. $r > 0.6$). Pasternack (2011) presented methodologies for 2D model validation and recommended a few more performance metrics. First, the mean and median of the unsigned velocity percent error should be < 30%. Second, the slope of the regression should be > 0.9, though this is a quite strict standard compared to most peer reviewed 2D models. Third, the y-intercept should be < 5% of the maximum velocity. Finally the mean of unsigned velocity direction error should be within 10%.

Given the lack of velocity direction validation studies, no performance benchmarks existed *a priori*. Direction varies over 360°, so a 5% error corresponds with 18°. However, for the mainstem thalweg where flow is directed downstream, then a more stringent criterion might be to limit the range to 180°, yielding a 5% error metric of 9°, which was used in this study. This is different from and more stringent than the 10% suggestion of Pasternack (2011).

To go deeper into error analysis than is commonly done with standard statistical metrics for 2D model validation, this study also computed the metrics of regression slope standard error and regression intercept standard error. The application of these metrics in this study involved testing the 1:1 expectation of predictions versus observations. By definition, a 1:1 relationship is linear (x=y). Therefore, the regression function is properly expected to be linear and the computation of standard error metrics assuming a linear regression is also proper for this application.

In contrast to the 2D hydrodynamic modeling community, the hydrological modeling community typically uses different metrics as performance indicators for discharge prediction. The choice of metrics ultimately reflects the different nature of the data and different expectations for model performance. Although there is no official standardization of validation in hydrology either, there are a few widely used metrics. For example, Moriasi et al. (2007) present a review of such metrics and that article has been cited more than 1,800 times. Moriasi et al. (2007) does affirm the use of some common 2D modeling validation metrics, but it also describes three common hydrological metrics: Nash-Sutcliffe efficiency (NSE), percent bias (PBIAS), and the root mean square error-observations standard deviation ratio (RSR). The performance standards for these metrics, as reported in that article, often rest on a single reference, so they are not necessarily robust standards. Nevertheless, the standards for discharge prediction are NSE > 0.5, PBIAS within 25%, and RSR < 0.7. How these apply to depth and velocity vector prediction is unknown, as almost no one uses these in 2D model validation studies. As a new direction for 2D modeling validation, these metrics were applied in this study.

Ballard et al. (2010a) suggested that an R-value of 0.6 (R^2 =0.36) constitutes a validated 2D model. Peer-reviewed reports and journal articles have reported R^2 values and slopes ranged from 0.25-0.92 and 0.66-0.86, respectively, in similar comparisons of model-predicted versus field-measured velocities (e.g. Lane et al., 1999; Pasternack et al., 2006; Harrison and Keller, 2007; May et al., 2009; Ballard et al., 2010a,b; Pasternack and Senter, 2011). These values therefore were used as performance indicators, though the range of previously accepted benchmark values for “validation” is so broad that no effort was made to judge the appropriateness of a specific threshold value for use this study; it is left up to the reader to decide that relative to the peer-reviewed literature values cited above. Histogram analysis and descriptive statistics of differences between model-predicted and field-measured velocities were also used to test model performance. Absolute velocity errors were compared with observed velocities to determine how the model performed across the range of velocities and determine if there were any noticeable trends. Furthermore, to help explain the cause of velocity errors, model-predicted depths were evaluated as an indicator of topographic uncertainty and Manning’s roughness parameterization. Cross-sectional analysis of model performance relative to DEM structure was also evaluated to help visualize and explain model errors. The above analyses constitute the typical suite of tests performed for 2D models in the peer-reviewed journal literature.2D Model Validation Using Fixed-Point Velocities

Wading-based depth and velocity data were directly compared with model results as one test of model performance. The data also served as a benchmark for kayak-based velocity performance. Observed velocities were graphed against model-predicted velocities and a linear best-fit trendline was added to the data (likewise for depths). The coefficient of determination (R^2) value of the best-fit line and slope were used as indicators of model performance.

3.4.3 Direct 2D Model Validation

None.

3.4.4 Adjusting Kayak Surface Velocity To Mean Velocity

According to Rantz (1982) using the data from the study conducted by Hulsing et al. (1966) natural channels have a DAVC of ~0.85-0.86 for surface velocities. The raw data was collected at cross-sections in channels classified into five types, phrased as: (1) natural trapezoidal-shaped channel without overbank flow and no bridge piers or other manmade obstructions, (2) natural channel with bridge piers, abutments or manmade obstructions that may affect the flow pattern, (3) canal or manmade channel without overbank flow, (4) and (5) same as (1) and (3) above, but with overbank-flow sections. They also span discharges of 0.031 to 18,000 m^3/s (Hulsing et al., 1966). As a result, the measurements generally reflect settings with a low roughness:depth ratio. Meanwhile, Pasternack et al. (2006) collected full vertical velocity profiles over an artificially constructed gravel bed in the Mokelumne River (similar to the bed present in the lower Yuba River, except for the absence of any sand in the artificial Mokelumne bed) and found that the slope of the least squared regression equation between the observed mean column velocity and observed near-surface velocity was 0.71, with a sample size of 23 and an R^2 of 0.672. Although the magnitude of the slope coefficient may change with different depth:bed material ratios, the existence of this fundamental relation was persistent and reliable. Furthermore, vertical velocity profiles collected in Timbuctoo Bend on the LYR in an unpublished study (co-author Greg Pasternack, UC Davis) yielded a minimum

DAVC of 0.65, indicating that there is a range of possible DAVC's across rivers of diverse flow:geometry ratios (though that was in 2004 before the big floods that stripped off the armor layer causing the bed to become less rough).. Thus, in addition to testing Rantz's published value of 0.85, the DAVC was varied to optimize three performance indicators with model-predicted results. These indicators were (1) the mean velocity difference between observed and predicted velocities (i.e. difference of ~0 m/s), (2) the mean error (i.e. mean error ~ 0%), and (3) best-fit trendline slope of $V_{modeled}$ versus $V_{observed}$ (i.e. slope ~ 1). The goal of this test was to see how the three DAVC values compared with each other and with the value published by Rantz (1982).

3.4.5 Testing Discharge Dependence of Model Error

None.

3.4.6 Comparing Validation Outcomes

None.

4 Results

4.1 2D Model Validation Using Fixed-Point Velocities

Velocities were measured in depths < 1.23 m. The mean velocity was 0.67 m/s, with a range of 0.00–1.62 m/s. The average coefficient of variation of velocity was 0.39, with a range of 0.12–0.70. Regression and correlation analyses of observed versus model-predicted depths and velocities yielded statistically significant trendlines and correlations above the 95% confidence level. For depth, the best-fit trendline had slope and R^2 values of 0.73 and 0.64, respectively (Fig. S2a). For velocity, it had corresponding values of 0.78 and 0.57, respectively, indicating a better 1:1 performance, but worse correlation performance than depth (Fig. S2b). Both relations show a bias in which low values are over predicted and high values are under predicted. Both indicators of performance are right in the middle of the range of peer-reviewed 2D-model validations. When modeled results were regressed on observations, the standard error of the regression slope and intercept for velocity were 0.0455 and 0.035, respectively. Both of these values are low and they help to appreciate that a regression plot with thousands of points often hides the true precision of the results, which in this case is quite high. Using hydrological performance indicators, the values of NSE, PBIAS, and RSR were 0.52, -4.4%, and 0.69. These values are within the thresholds accepted by the hydrological community.

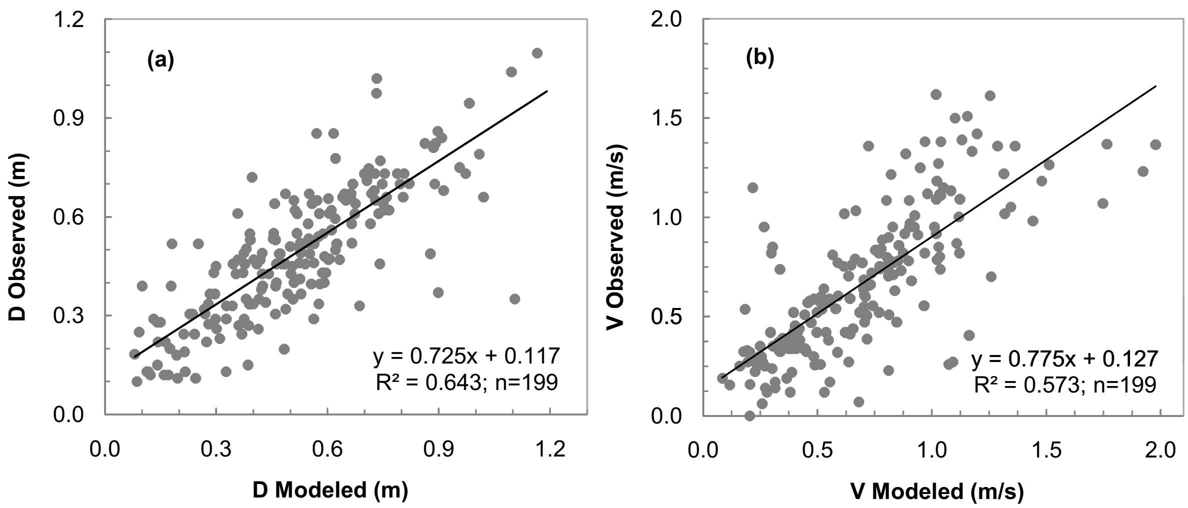


Figure S2. Predicted model results versus measured, cross-sectional wading values with calculated best-fit trendlines and linear regression for: (a) depths; (b) velocities.

The mean value of raw differences between modeled and observed depths was 0.025 m with a standard deviation (SD) of 0.140 m (Table S1). Histogram analysis showed that 20.1% of modeled depths were within 0.03 m. of observed values and 77.4% were within 0.15 m. For reference, the median bed material size in the vicinity of the cross-sections is ~0.06 m, so one would not expect most deviations to be smaller than that. The mean absolute error was 25.5% with a SD of 30.3%.

The mean raw difference of modeled and observed velocities for all the points was 0.029 m/s with a SD of 0.25 m/s (Table S1). The mean absolute error was 40.0% with a SD of 81.6%. The largest velocity errors (200-800%) correlated with some of the smallest observed velocities (Fig. S3). Histogram analysis found that 16.1% of the modeled values were within 0.03 m/s of the observed values, 60.3% were within 0.15 m/s, and 90% were within 0.41 m/s.

Table S1. Descriptive statistics for modeled versus wading-based hydraulics.

	Depth (m)			
	D_{Diff}	ABS(D_{Diff})	D_{Error}	ABS(D_{Error})
Mean	0.025	0.102	9.3%	25.5%
Standard Error	0.010	0.007	2.7%	2.1%
Median	0.027	0.075	5.0%	17.2%
Standard Deviation	0.140	0.099	38.5%	30.3%
Range	1.092	0.755	289.6%	215.3%
Minimum	-0.337	0.000	-74.3%	0.0%
Maximum	0.755	0.755	215.4%	215.4%
	Velocity (m/s)			
	V_{Diff}	ABS(V_{Diff})	V_{Error}	ABS(V_{Error})
Mean	0.029	0.176	22.0%	40.0%
Standard Error	0.018	0.013	6.3%	5.8%
Median	0.014	0.111	2.3%	18.2%
Standard Deviation	0.253	0.183	88.2%	81.6%
Range	1.753	0.931	952.2%	870.9%
Minimum	-0.932	0.001	-81.1%	0.2%
Maximum	0.821	0.932	871.1%	871.1%

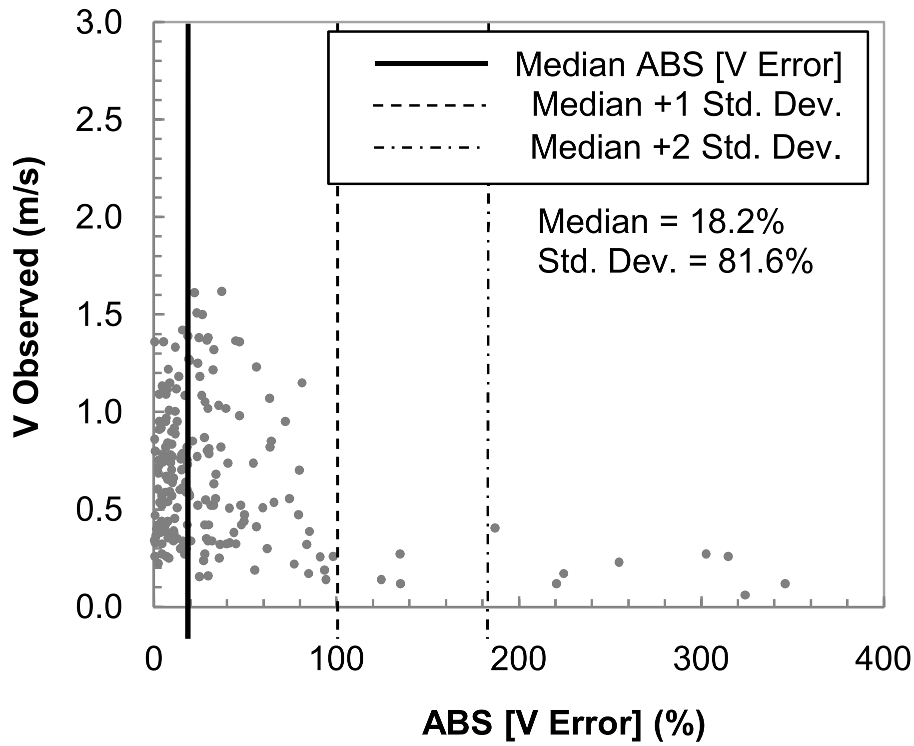


Figure S3. Absolute velocity error versus wading velocities.

Cross-sectional comparisons of the lateral patterns of observed and modeled depths and velocities help explain the significant causes of model error and bias (Fig. S4). As is typical, cross-sections have low depths and velocities along the banks, except where there is forced scour. However, for some cross-sections, the model over-predicted both variables there (e.g. Fig. S4a,i,t,u). A poor specification of roughness would yield an inverse bias between depth and velocity, so that is unlikely to be the main cause. The choice to not use a different roughness value for banks than the ambient bed could still be a secondary factor causing bank velocities to be too high. An excessive eddy-viscosity coefficient disproportionately affects velocity, so adjusting that down in future studies might improve model performance (but that was not evaluated as part of this study). Finally, given a near-flat water surface across the cross-sections, the deviations in depth are a surrogate for topographic error (Pasternack et al., 2006). Such error is attributable to the gap between the airborne LiDAR survey of terrestrial topography and the boat-based survey of bathymetry; wherever the boat could not get within a meter of the bank, meaningful landform variability may have been missed (Fig. S5). Although there was an extensive effort to avoid inaccuracies in the DEM due to the large scale of the project, inability to wade certain areas, along with time and monetary restrictions, a couple of locations were not as detailed as required for accurate model predictions. TIN-based interpolation during DEM production would then just cut off those features (e.g. Fig. S4l,n,r). In other instances, boat-based bathymetric measurements simply appear to be faulty (e.g. Fig. S4f,w). Observations made within 2.5 m of the model predicted wetted channel perimeter averaged a difference in velocity of 0.072 m/s which is ~2.5 times larger than the mean for all points. Furthermore, mean velocity error for the same points (45.8%) was ~24% higher than the total mean.

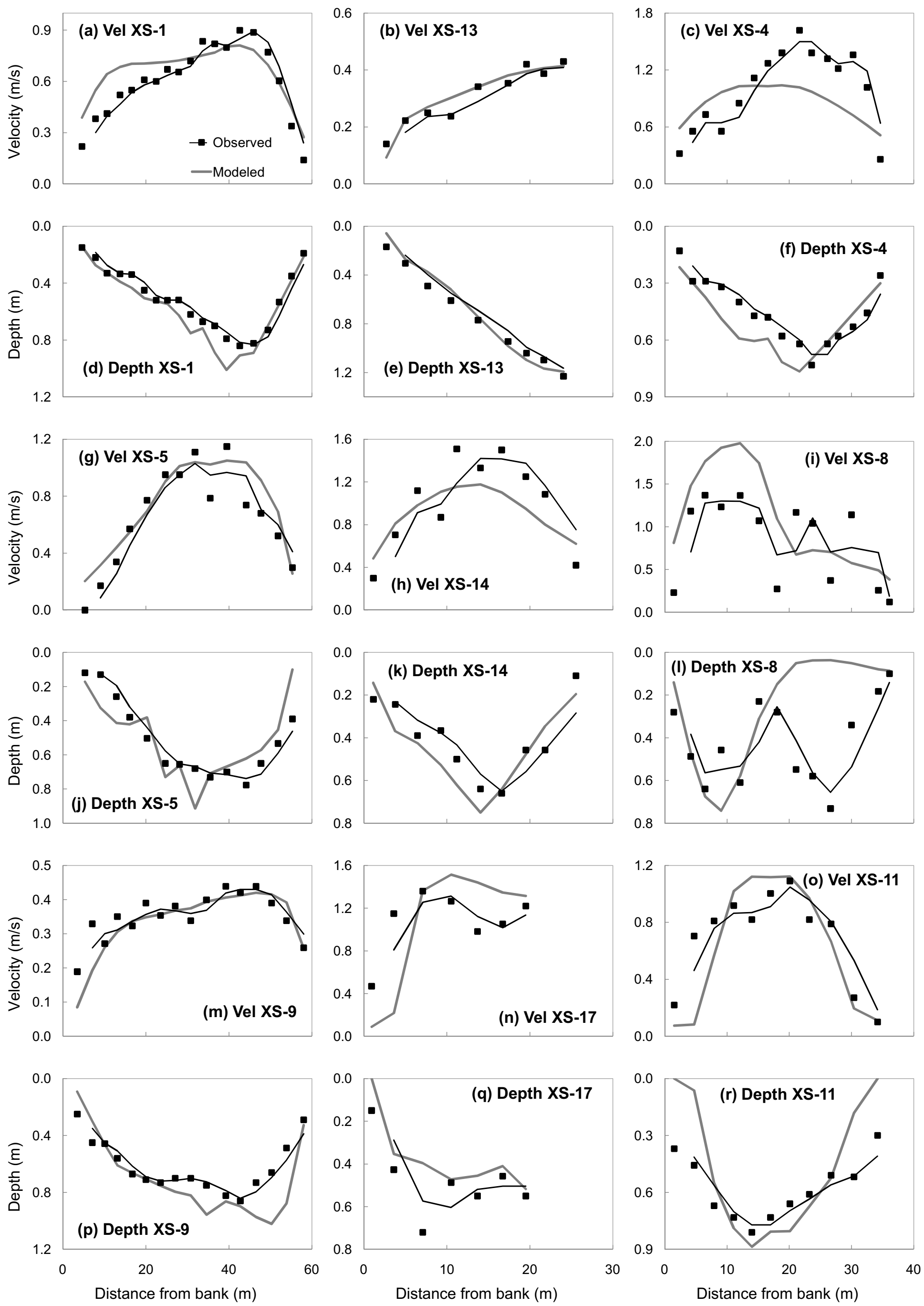


Figure S4. Comparisons of observed versus predicted depths and velocities at all cross-sections. Field observations were fit with a curve using the local average to reduce measurement noise.

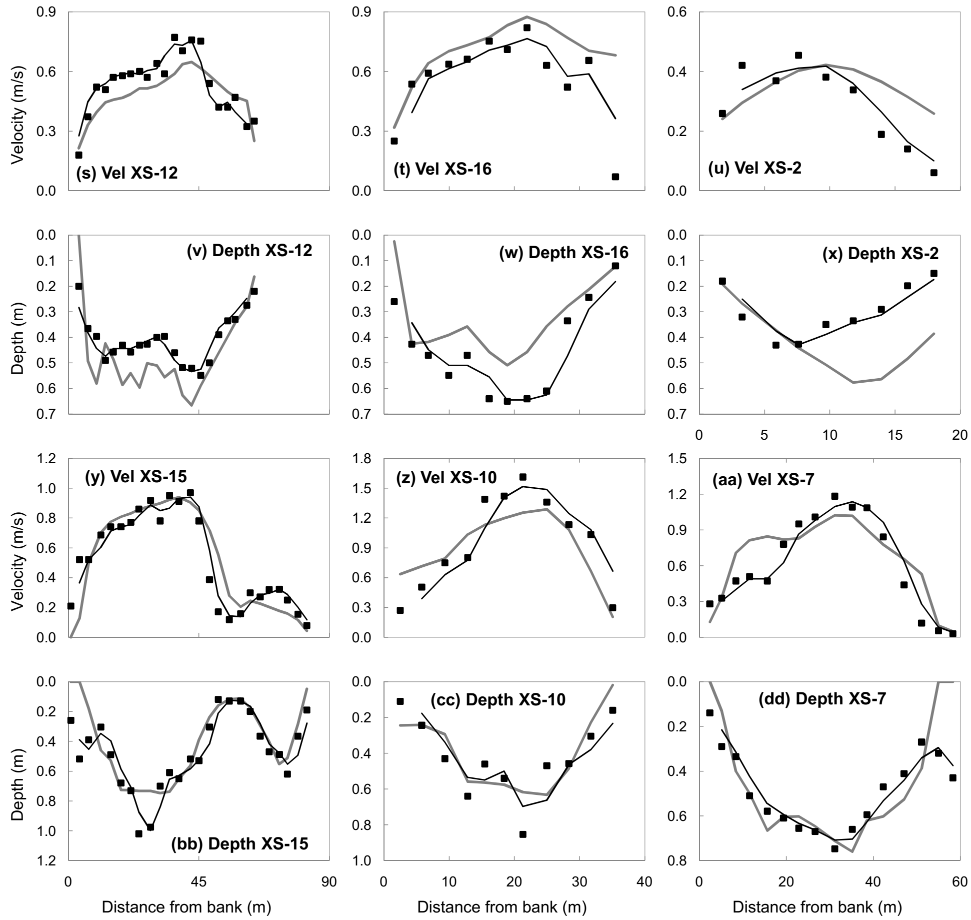


Figure S4. Continued.

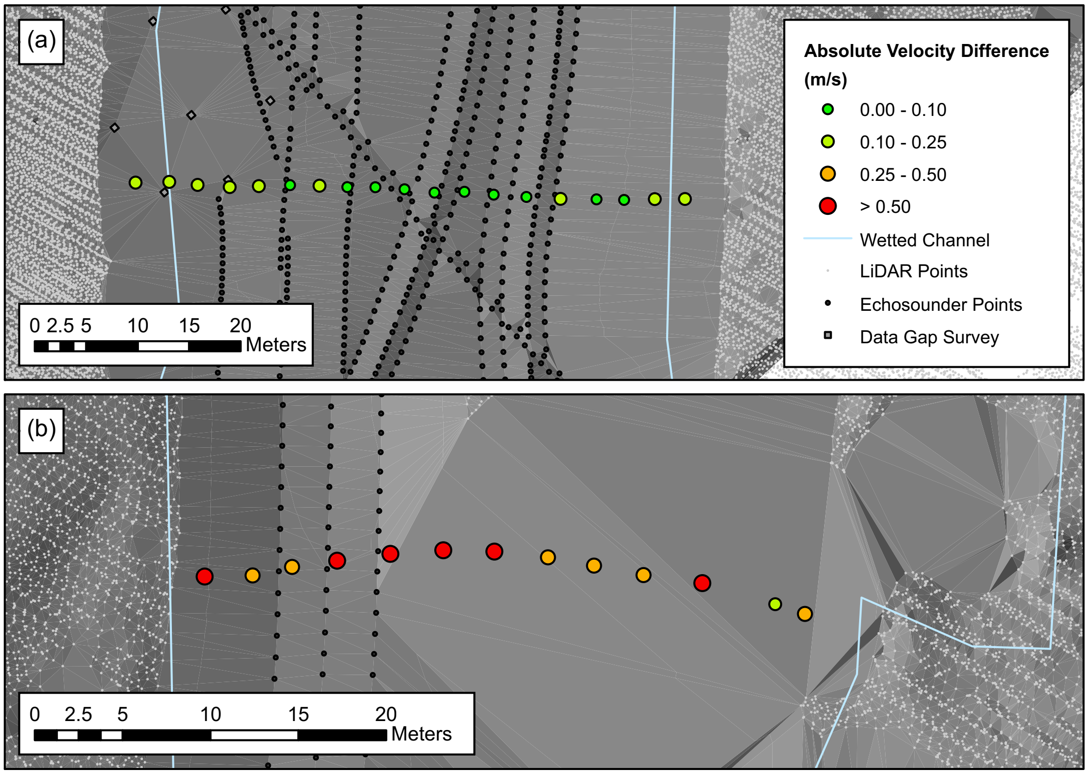


Figure S5. Maps of the DEM showing the absolute difference between model velocities and measured, wading velocities: (a) small differences in velocity because of accurate representation of the land surface due to good survey coverage [XS 1] and (b) large differences in velocities due to poor, in channel survey coverage [XS 8].

4.2 2D Model Validation Using Kayak Observations

Mean observed surface velocity was 0.87 m/s, with a range of 0.15–3.44 m/s respectively. Kayak velocities were made at model-predicted depths of up to 4.37 m. Mean, minimum, and maximum kayak directional values were 208.0°, 134.7°, and 303.5° respectively.

4.2.1 Statistical Velocity Validation

Differencing results and velocity error analysis for Rantz's published DAVC yielded the largest dissimilarities, of the four DAVCs tested, with the modeled results. The mean difference was -0.285 m/s and SD was 0.286 m/s (Tables S2-S3). Mean error and SD of the error were 20.1% and 24.3% respectively. The best-fit trendline slope was 1.195 with a y-intercept of 0.115. Mean velocity-difference optimization obtained a DAVC equal to 0.64, which resulted in a mean difference of 0.00 m/s, an absolute mean difference of 0.15 m/s, and a standard deviation of 0.21 m/s. DAVC equal to 0.64 resulted in the largest mean, standard deviation, and range for velocity error which was -6.1%, 32.3% and 388.4% respectively. Mean error optimization resulted in a DAVC of 0.68. Mean error came out to be 0.1% with an absolute error of 20.6% and a SD of 30.3%. Finally, trendline slope optimization resulted in a DAVC equal to 0.71, the same as the value used for the 2D model validation. A description of those results can be found in the previous section.

The analysis of varying the DAVC to optimize different parameters showed that DAVC values were low compared to published values, most of which come from deeper, slower channels with fine riverbed sediment. Although each of the DAVC metrics has its own benefits, there is no obvious best choice. The DAVC of 0.64 with an optimized mean velocity difference also has the best velocity difference distribution and error distribution. As mentioned in the section 3.6.3, unpublished vertical velocity profiles collected by co-author Pasternack and associates in Timbuctoo Bend on the LYR yielded a minimum DAVC of 0.65, which is close to this value (though that was in 2004 before the big floods that stripped off the armor layer causing the bed to become less rough). However, 0.65 was the lowest value found in the published literature, which raises the question: Should observed values be adjusted to match predicted values? While 0.64 seems like the best option based on the numbers, 0.71 is likely to be the more realistic value. First, this value is the same that was calculated by

Pasternack et al. (2006) for 24 vertical profiles where full vertical velocity profiles were measured. Second, although the mean velocity difference was -0.1 m/s for DAVC=0.71, this can be explained by the fact that there were slightly more values at low velocities where the model tended to over-predict velocity. After further experimentation with 2D models in diverse channel settings, it is now thought that this can be improved in the future by significantly reducing the eddy-viscosity coefficient for shallow gravel-bed rivers from the generic value of ~0.6-0.7 to ~0.075-0.1, thereby enhancing the lateral gradient in velocity magnitude. It is recommended that studies adjust this value specifically for improving the lateral gradient as well as the pattern of eddies (Wheaton et al., 2004). Finally, this study was performed in a channel with a high roughness to depth ratio, whereas classic studies yielding DAVCs of >0.8 were done in channels with low ratios. Extra roughness means a stronger vertical velocity gradient and a lower DAVC. This likely is the reason for the large dissimilarity between model and observed results when applying Rantz's DAVC of 0.85.

4.2.2 No Discharge Dependence of Model Error Found

Probability density functions of the velocity deviations for each discharge are plotted, so interested readers who attempt this approach can compare their distributions (Figure S6). Mean, Standard deviation, skewness, and kurtosis values are 0.237, 0.263, 1.234, and 2.988 for the lowest flow and 0.314, 0.240, -0.026, and 0.766 for the highest flow.

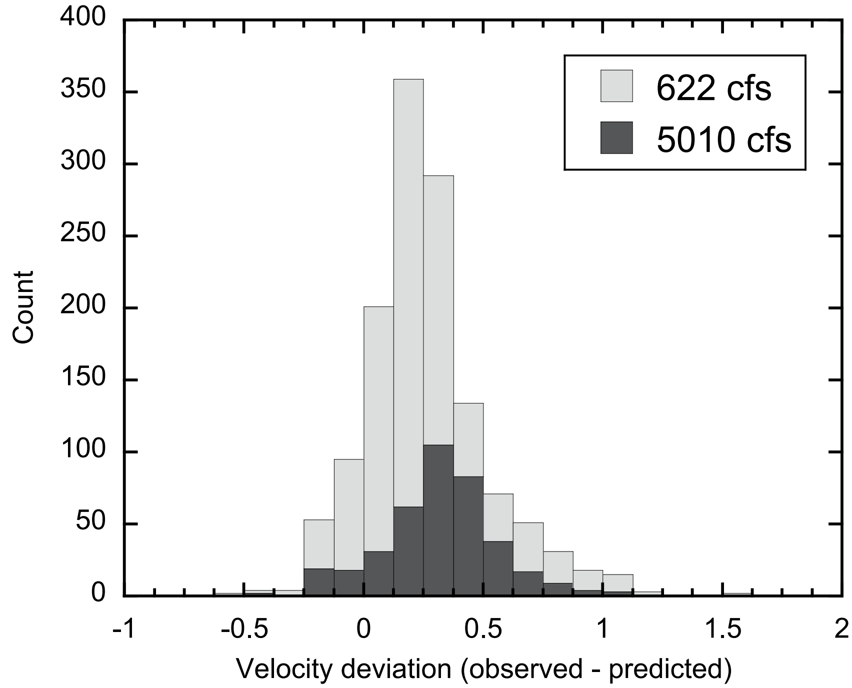


Figure S6. Probability density functions of velocity deviations between observed and predicted by the 2D model.

4.2.3 Spatial Velocity Validation

Despite efforts to avoid it, there were localized occurrences of poor topographic interpolation between in-channel single-beam echosounder points and out-of-water airborne LiDAR points, and these caused locally poor model performance (e.g. see Figure S7). In this study, there was an abundance of data available and breaklines were used in places where triangulation issues were evident at a coarser scale, but there were still localized data gaps. With ~ 36 km of river to map and model at one-meter resolution, it is costly and time-consuming to inspect every locality within the map, despite reasonable efforts to do so. In similar fashion for hydrological modeling, hardly any effort is made to thoroughly inspect landscape DEMs to represent every nook and cranny. When creating a TIN through the gaps in the point data, a triangle sometimes interpolated laterally from the bank to a point in the channel creating a ridge extending into the channel (features circled in red Figure 76). The model would interpret these features as constrictions or barriers, and would either accelerate flow through the main channel (Figure S7a) or block flow (Figure S7b). Where available and suitable today, bathymetric LiDAR and multibeam echosounding are replacing single-beam echosounding, avoiding such problems. Another solution, though time-consuming, is to run a coarse-resolution 2D model quickly, test the model against data to find the worst performing localities, and fix them iteratively.

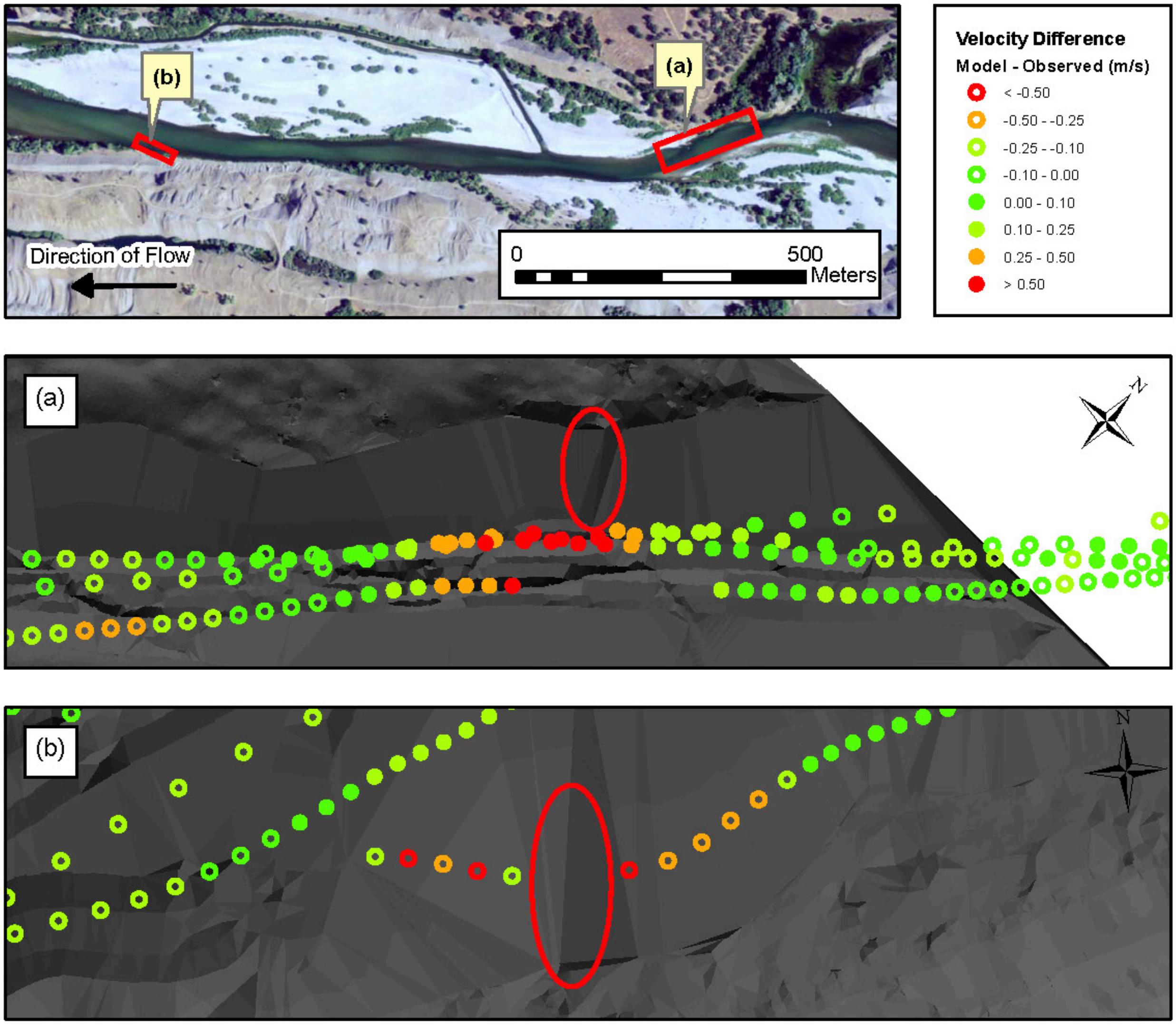


Figure S7. Differences between modeled and observed kayak velocities due to locally inaccurate DEM interpolations.

4.2.4 Directional Validation

This study led to the realization that validation of eddies important to stakeholders requires focused kayak data collection in which the boat recirculates several times around the eddy on different paths. Slow eddies require slow GPS time interval sampling of 10-30 s.

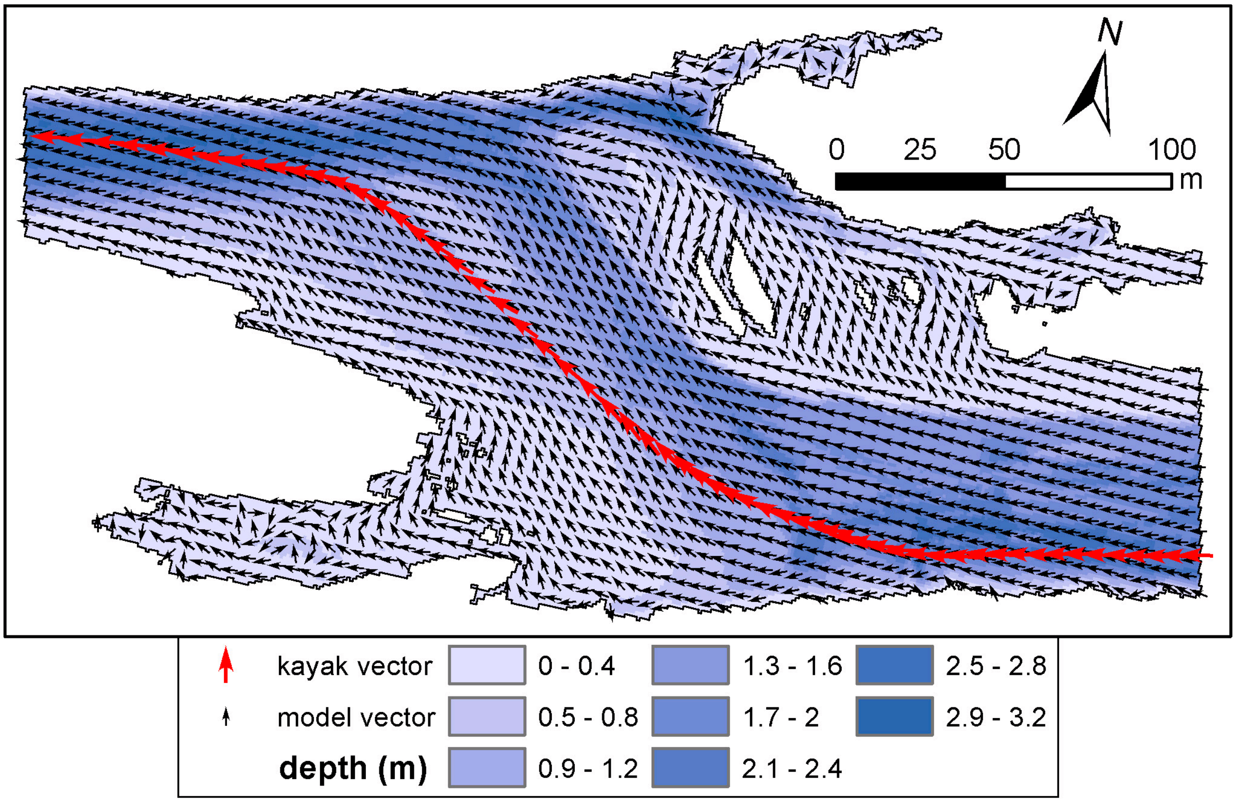


Figure S8. Vector fields on a small portion of the Daguerre Point Dam reach at 105 m^3/s.

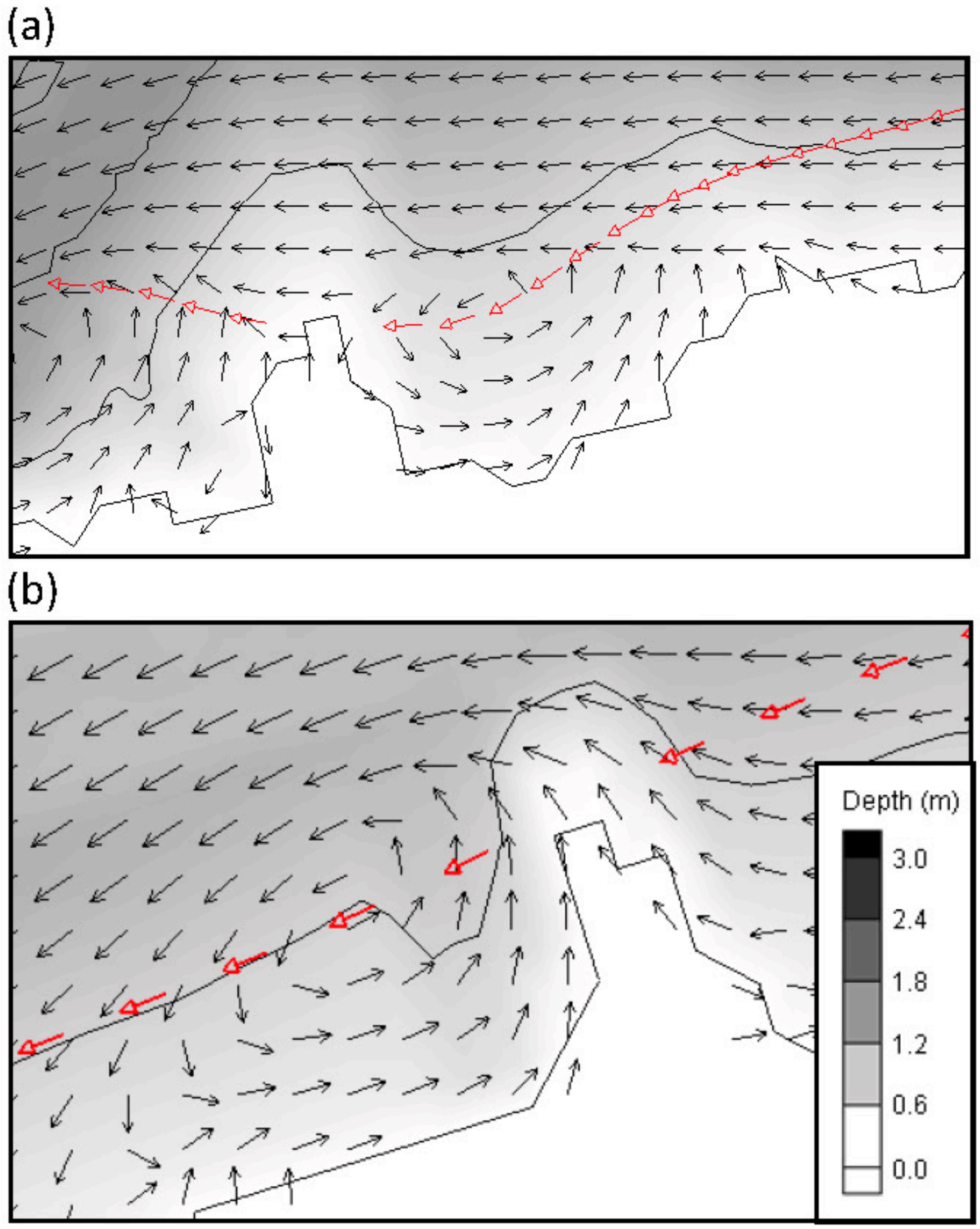


Figure S9. Areas where the model predicted flow vectors differ greatly from the kayak measured velocity vectors due to model predicted eddies.

4.3 Comparing Validation Outcomes

Distributions of observed velocities with an ABS(V_{error}) > 50% for both fixed-point and kayak velocity datasets showed important differences (Figure S10). The fixed-point dataset had a lognormal distribution, while the kayak dataset had an exponential distribution. The peak of these errors was higher with the fixed-point data than with the kayak data. The latter also showed a longer upper tail.

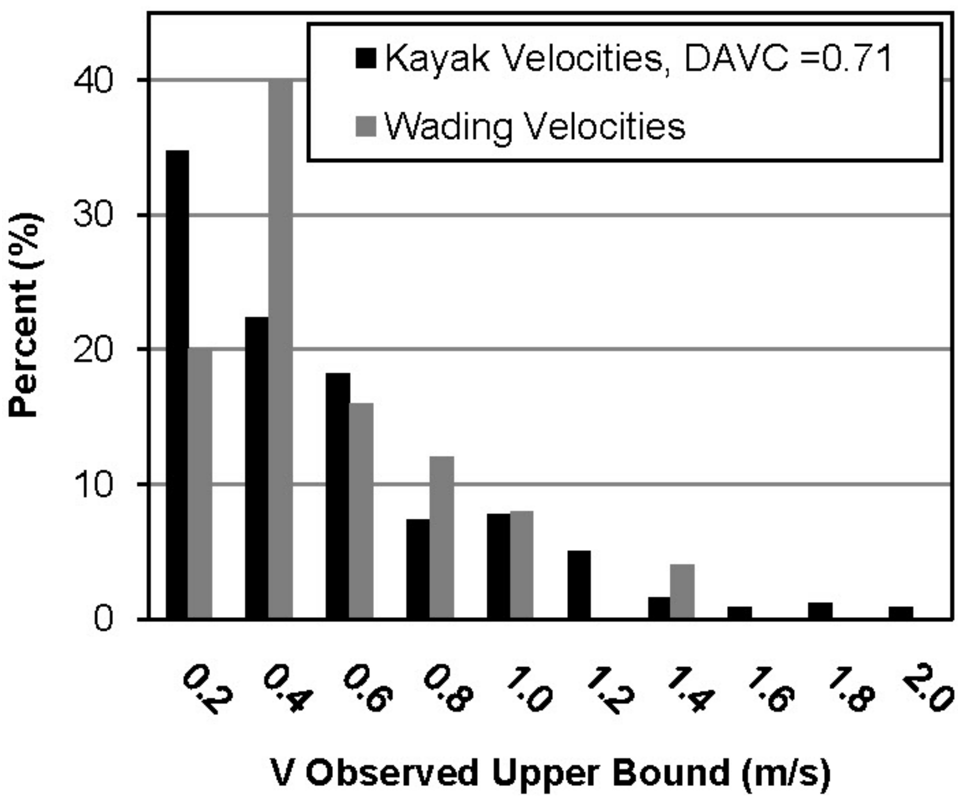


Figure S10. Distribution of observed fixed-point and kayak velocities with V Error > 50%.

Table S2. Descriptive statistics for modeled versus kayak-based velocity deviations and percent errors for all DAVC values tested.

(A) Velocity Differences, Modeled - Observed (m/s)

	DAVC = 0.64		**DAVC = 0.68**	
Statistic	V_{Diff}	ABS(V_{Diff})	V_{Diff}	ABS(V_{Diff})
Mean	***0.001***	0.150	-0.054	0.159
Std. Error	0.009	0.006	0.009	0.007
Median	0.013	0.109	-0.036	0.111
Std. Dev.	0.210	0.147	0.219	0.160
Range	2.042	1.154	2.087	1.269
Min	-1.154	0.000	-1.269	0.000
Max	0.888	1.154	0.818	1.269

	DAVC = 0.71		**DAVC = 0.85**	
Statistic	V_{Diff}	ABS(V_{Diff})	V_{Diff}	ABS(V_{Diff})
Mean	-0.095	0.175	-0.285	0.314
Std. Error	0.010	0.007	0.004	0.003
Median	-0.072	0.120	-0.247	0.254
Std. Dev.	0.228	0.174	0.286	0.254
Range	2.127	1.355	2.377	1.756
Min	-1.355	0.000	-1.757	0.000
Max	0.772	1.355	0.620	1.757

(B) Velocity Error (%)

	DAVC = 0.64		**DAVC = 0.68**	
Statistic	V_{Error}	ABS(V_{Error})	V_{Error}	ABS(V_{Error})
Mean	-6.1	21.5	***0.1***	20.6
Std. Error	0.42	0.33	0.40	0.29
Median	-1.6	14.2	4.3	14.2
Std. Dev.	32.2	24.7	30.3	22.2
Range	388	292	366	269
Min	-292	0.00	-269	0.00
Max	96.5	292	96.7	269

	DAVC = 0.71		**DAVC = 0.85**	
Statistic	V_{Error}	ABS(V_{Error})	V_{Error}	ABS(V_{Error})
Mean	4.4	20.8	20.1	27.0
Std. Error	0.38	0.27	0.32	0.21
Median	8.4	15.6	23.5	25.3
Std. Dev.	29.0	20.7	24.3	16.2
Range	350	253	292	195
Min	-253	0.00	-195	0.04
Max	96.8	253	97.3	195

Table S3. Histogram analysis of kayak observed velocity.

Velocity Difference Distribution	**DAVC = 0.64**	**DAVC = 0.68**	**DAVC = 0.71**	**DAVC = 0.85**
Within 0.03 m/s	15.74%	13.91%	12.73%	4.88%
Within 0.15 m/s	64.62%	63.03%	58.67%	28.13%
Error Distribution				
Within 5%	19.88%	19.13%	17.37%	5.85%
Within 10%	37.13%	37.35%	34.26%	12.40%
Within 25%	73.04%	72.65%	70.48%	48.82%
Best-Fit Line				
Slope	0.900	0.956	***0.998***	1.195
Intercept (m/s)	0.087	0.092	0.096	0.115

5 Discussion

Using the same traditional methods of 2D-model validation as commonly reported in the peer-reviewed literature, the LYR 2D model performed on par with past peer-reviewed studies (Figs. S2, S4). Observed velocities for cross-sections 1, 5, 9, 11, 12, 13, 15, and 16 match model-predicted velocity's trends, while those for cross-section 2, 4, and 8 showed major discrepancies, which is typical in the literature. Depth and velocity test R^2 values of 0.64 and 0.57 were right in the middle to upper range of the values previously reported. Note that some studies only report R (e.g. Ballard et al., 2010a,b), not R^2, so care must be used in making comparisons. Further, lateral patterns of deviations in depth and velocity showed smoothing across the channel with lows too high and highs too low. This is a typical occurrence of errors due to topographic uncertainty and errors due to an inadequate rate of velocity change associated with excessively efficient turbulent mixing (MacWilliams et al., 2006) and perhaps insufficient bank roughness (Abu-Aly et al., 2013).

5.1 Improved Sampling of Deep And Fast Areas

Boat-based ADCP can go into deep current, but it normally requires holding position to get a good measurement and it is extremely difficult to collect boat-based ADCP data in fast, hazardous conditions, locations with obstructions, and shallow water impacted by the ADCP's blanking depth, all of which the kayak method can handle.

Similar results were obtained by only using the highest and lowest flows, indicating that cost savings are possible and that model performance is likely insensitive to specific discharges. This is promising, because it indicates that it may be unnecessary to do validations at every discharge for which scientific analysis would be performed, and validating models for hazardous floods is rarely feasible. It is really only necessary to collect new model calibration and validation data when the hydraulic roughness structure changes a lot, such as when vegetation becomes submerged (Abu-Aly et al., 2013).

5.2 Improved Sampling Of Statistical Structure

Testing 2D-model performance with the kayak observations yielded a 39% higher value for r^2 (0.79 versus 0.57), and the value of DAVC did not matter for this metric. This high correlation between modeled and observed velocity is among the best in peer-reviewed reports and journal articles evaluating a shallow gravel-bed river. After adjustment with the DAVC, the comparative performance between velocity tracking versus fixed-point observations, as measured with the standard error of the regression slope and intercept metrics, was ~6-9 times better with the former approach. In terms of the hydrological performance metrics, both NSE and RSR were significantly better in the velocity tracking results, while PBIS was better in the fixed-point results.

5.3 Improved Spatial Testing

None.

5.4 Putting The 2D In 2D Model Validation

None.

5.5 Important Kayak Limitations

Sections 1.3 and 3.4 of the main article and supplementary file previously addressed the uncertainties in kayak RTK GPS surface velocity measurements, so the focus here is on the limitations of using such data in 2D model validation. Although the kayak RTK GPS method of measuring surface velocity was valuable for fast, low-cost 2D model validation, it did have some problems. First, there were major discrepancies between modeled and measured results at very low velocities (Figure 4). The best solution would be to adjust the GPS time interval for different flow speeds, so 5 s would be adequate for velocities of > 0.5 m/s, while 10–30 s might be needed for 0–0.5 m/s speeds.

Second, GPS units only record time to the nearest second, so it is plausible that more error is arising from inaccurate time stamps than spatial precisions. If the fixed time interval sampling algorithm in the Trimble software uses the internal clock only to identify the time interval, then the exact time does not matter and the error in change in time is likely quite low. However, if the time interval is not truly fixed and hinges on the absolutely clock time, recorded only to the nearest 1 s, then that would introduce uncertain error. For a 5 s time interval, an error of 0.5 s (the worst possible for a 1 s clock) would yield a 10% error in velocity, so that is significant. However, for any instant in time with a 5 s sampling interval, the likelihood of 10% error is equal to that of 0% error, so one cannot account for this quantitatively. Third, no surface particle tracking method is viable in high winds. High winds can be more of a problem at slow base flows and over slow, shallow embayments than in the thalweg of a large flood in the middle of a big storm. Finally, for a large enough flood, it may be sensible to switch from a kayak to a larger boat, possibly motorized for safety and effectiveness in that regime. Initial testing with a motorboat found that it was more difficult to match the timing of speed adjustments and the boat was more sensitive to wind, but for large floods it would be viable. Conversely, if a stream was only 1-20 m wide, then a manned kayak would likely be too big of a drifter to be viable. Further discussion about future improvements to kayak velocity mapping is provided in the supplementary materials file in section 5.6.

5.6 Future Improvements

In this study, velocities were largely measured in the thalweg and observations did not accurately represent the whole range of velocities at each individual flow, especially with insufficient sampling along channel banks and in very slow conditions. This caused a clustering of observations and a velocity gap between high and low flows. Pasternack and Senter (2011) built on this pilot effort and tested 2D model performance in a steep mountain river over two orders of magnitude in the range in flow to assess validation sensitivity to discharge, among other things. They took the approach of being sure to sample the full range of safely accessible velocity bins by wading and kayak at every discharge, with less emphasis placed on spatial coverage in light of the remote setting. Therefore, each day's dataset could be used to produce a scatter plot validation test, yielding an even more robust statistical analysis when all days were combined. This approach of mindful sampling to achieve a more equal representation of each velocity bin each time out guarantees that a scatter plot can be made from the available data even if higher flows never come, since one cannot usually preordain what flows a river will deliver during the validation period. Thus, the best procedure involves focusing on sampling a diversity of velocity vector conditions instead of staying on one streamline. People should spend time mapping recirculating eddies, observing zero-velocity locations, and going over the most aggressive chutes. This is done by starting and stopping the GPS data collection as needed, as well as using a path identifier in the data description field. Being on the drifter is the key to mindful mapping, and this in turn may lead to more robust validation testing, because it enables more sampling in the periphery where survey data and DEM representation are often worse making validation tests more useful to find such problems. Also, 2D models typically better represent deeper flows than very shallow flows, so more data collection in shallow, outlying areas of the channel would help strengthen the validation process.

In the future, it may be beneficial to attach an echosounder and/or an ADCP to the kayak to measure depths and velocity profiles to get more data for validation and to aid interpretation of validation problems. Topographic error is the dominant cause of velocity error (Figure S5), even with the high-quality topography collected for this study, so dual observation is sensible.

6 CONCLUSIONS

None.

7 GEOLOCATION

39°13′13″ N, 121°20′7″ W

8 ACKNOWLEDGMENTS

We would like to thank Casey Campos, Derek Givens, and Uminchu Naoaki at Pacific States Marine Fisheries Commission for their help on the river making velocity measurements. We thank Chris Hammersmark and Sam Diaz with CBEC Eco-Engineering as well as Jimmy

Kulpa of Environmental Data Solutions for their assistance with the surveying and mapping effort. Professor Fabian Bombardelli (UC Davis) and Robert Hilldale (U.S. Bureau of Reclamation) provided constructive reviews of the manuscript prior to journal submission. We thank anonymous reviewers for comments to improve the manuscript.

9 SUPPLEMENTAL REFERENCES

Abu-Aly TR, Pasternack GB, Wyrick JR, Barker R, Massa DA, Johnson TR. (2013). Effects of LiDAR–derived, spatially distributed vegetation roughness on two–dimensional hydraulics in a gravel–cobble river at flows of 0.2 to 20 times bankfull. Geomorphology 206:468–482. doi:10.1016/j.geomorph.2013.10.017.

Anderson MG, Bates PD. (1994). Evaluating data constraints on two-dimensional finite element models of floodplain flow. Catena 22: 1–15.

Ariens Specialty Brands LLC. (2014). WAAS and GPS accuracy. Document Number 300. Ben Meadows, Janesville, WI.

Ballard E, Gard M, Pelle E. (2010a). Flow–habitat relationships for spring and fall–run Chinook salmon and steelhead/rainbow trout spawning in the Yuba River. Sacramento (CA): United States Fish and Wildlife Service, Energy Planning and Instream Flow Branch.

Ballard E, Gard M, Pelle E. (2010b). Flow-habitat relationships for juvenile fall/spring-run Chinook salmon and steelhead/rainbow trout rearing in the Yuba River. Sacramento (CA): United States Fish and Wildlife Service, Energy Planning and Instream Flow Branch.

Barker JR. (2010). Lower Yuba River QAQC Comparison: LIDAR Data, Boat Echo Sounder Data, NGS Benchmarks, Total Station and RTK-GPS Survey Points. Marysville (CA): Yuba Accord River Management Team.

Barker JR. (2011). Rapid, abundant velocity observation to validate million-element 2D hydrodynamic models. [M.S. Thesis]. [Davis (CA)]: University of California at Davis.

Brown, R. A., & Pasternack, G. B. (2008). Engineered channel controls limiting spawning habitat rehabilitation success on regulated gravel–bed rivers. *Geomorphology*, 97, 631–654.

Brown, R. A., Pasternack, G. B, & Lin, T. (2016). The topographic design of river channels for form–process linkages. *Environmental Management*, 57, 929–942.

Buchanan TJ, Somers WP. (1969). Discharge measurements at gaging stations. Washington (DC): United States Geological Survey. Techniques of Water-Resources Investigations, Book 3. Chapter A8. http://pubs.usgs.gov/twri/twri3a8/.

Carley JK, Pasternack GB, Wyrick JR, Barker JR, Bratovich PM, Massa DA, Reedy GD, Johnson TR. (2012). Significant decadal channel change 58–67 years post–dam accounting for uncertainty in topographic change detection between contour maps and point cloud models. Geomorphology 179:71–88.

Clark, J. S, Rizzo, D. M, Watzin, M. C, & Hession, W. C. (2008). Spatial distribution and geomorphic condition of fish habitat in streams: an analysis using hydraulic modelling and geostatistics. *River Research and Applications*, 24, 885–899.

Creëlle, S., Roldan, R., Herremans, A., Meire, D., Buis, K., Meire, P., Van Oyen, T., De Mulder, T., Troch, P. (2016). Validation of large–scale particle image velocimetry to acquire free–surface flow fields in vegetated rivers. *Journal of Applied Water Engineering and Research*. DOI: 10.1080/23249676.2016.1251856.

Creutin, J. D., Muste, M., Bradley, A. A., Kim, S. C., & Kruger, A. (2003). River gauging using PIV techniques: a proof of concept experiment on the Iowa River. *Journal of Hydrogeology*, 277, 182–194.

Detert, M., & Weitbrecht, V. (2015). A low–cost airborne velocimetry system: proof of concept. *Journal of Hydraulic Research*, 53, 532–539.

Elkins, E. M., Pasternack, G. B., & Merz, J. E. (2007). Use of slope creation for rehabilitating incised, regulated, gravel bed rivers. *Water Resources Research*, 43, W05432. DOI: 10.1029/2006WR005159.

Fischer HB, List EJ, Koh RCY, Imberger J, Brooks NH. (1979). Mixing in Inland and Coastal Waters. New York (NY): Academic Press, Inc.

French JR, Clifford NJ. (2000). Hydrodynamic modeling as a basis for explaining estuarine environmental dynamics: some computational and methodological issues. Hydrological Processes 14:2089–2108.

Fujita, I., & Kunita, Y. (2011). Application of aerial LSPIV to the 2002 flood of the Yodo River using a helicopter mounted high density video camera. *Journal of Hydro–environment Research*, 5, 323–331.

Fulford J.M. (2001). Accuracy and consistency of water-current meters. Journal of the American Water Resources Association 37: 1215–1224.

Ghanem, A., & Hicks, F. E. (1992). A Review of Hydraulic Models used in Instream Flow Needs Assessment Methods (Water Resources Engineering Report No. 92–4). Edmonton, Alberta: University of Alberta.

Harrison LR, Keller EA. (2007). Modeling forced pool–riffle hydraulics in a boulder-bed stream, southern California. Geomorphology 83:232–248.

Hulsing H, Smith W, Cobb ED. (1966). Velocity–head coefficients in open channels. Washington (DC): United States Geological Survey Water–Supply Paper 1869–C.

Kammel, L. E., Pasternack, G. B., Massa, D. A., & Bratovich, P. M. (2016). Near–census ecohydraulics bioverification of Oncorhynchus mykiss spawning microhabitat preferences. *Journal of Ecohydraulics*, 1(1-2), 62-78. DOI: 10.1080/24705357.2016.1237264.

Lai YG. (2008). SRH–2D version 2: theory and user's manual. Denver (CO): U.S. Department of the Interior, Bureau of Reclamation, Technical Service Center.

Lane SN, Bradbrook KF, Richards KS, Biron PA, Roy AG. (1999). The application of computational fluid dynamics to natural river channels: three–dimensional versus two–dimensional approaches. Geomorphology 29:1–20.

Lee, J. H., Kil, J. T., & Jeong, S. (2010). Evaluation of physical fish habitat quality enhancement designs in urban streams using a 2D hydrodynamic model. *Ecological Engineering*, 36, 1251–1259.

Lee, H. C., Lin, C. Y., Lin, C. H., Hsu, S. W., & King, C. T. (2011). A Low-Cost Method for Measuring Surface Currents and Modeling Drifting Objects. *IEEE Transactions on Instrumentation and Measurement*, 60, 980-989. DOI: 10.1109/TIM.2010.2062730.

Lewis, Q., & Rhoads, B. (2015). Resolving two–dimensional flow structure in rivers using large–scale particle image velocimetry: An example from a stream confluence. *Water Resources Research*. DOI:10.1002/2015WR017783.

Ludwig D. (2001). The Era of Management Is Over. Ecosystems 4: 758-764.

MacWilliams ML, Wheaton JM, Pasternack GB, Street RL, Kitanidis PK. (2006). Flow convergence routing hypothesis for pool-riffle maintenance in alluvial rivers. Water Resources Research 42. DOI: 10.1029/2005WR004391

Marks K, Bates P. (2000). Integration of high-resolution topographic data with floodplain flow models. Hydrological Processes 14:2109–2122.

Mason, D. C., Horritt, M. S., Hunter, N. M., & Bates, P. B. (2007). Use of fused airborne scanning laser altimetry and digital map data for urban flood modeling. *Hydrologic Processes*, 21, 1436–1447.

May CL, Pryor BS, Lisle T, Lang M. (2009). Coupling hydrodynamic modeling and empirical measures of bed mobility to predict the risk of scour and fill of salmon redds in a large regulated river. Water Resources Research 45:W05402. doi:10.1029/2007WR006498.

Milanesi L, Pilotti M, Ranzi R. (2015). A conceptual model of people's vulnerability to floods. Water Resources Research 51:182-197. doi:10.1002/2014WR016172

Mueller DS, Abad JD, García CM, Gartner JW, García MH, Oberg KA. (2007). Errors in acoustic Doppler profiler velocity measurements caused by flow disturbance. Journal of Hydraulic Engineering 133:1411-1420.

Pasternack GB. (2008). SHIRA-Based River analysis and field-based manipulative sediment transport experiments to balance habitat and geomorphic goals on the lower Yuba River. Davis (CA): University of California at Davis. Cooperative Ecosystems Studies Unit 81332 6 J002 Final Report.

Pasternack GB. (2009). Specific Sampling Protocols and Procedures for Topographic Mapping. Marysville (CA). The Lower Yuba River Accord Planning Team

Pasternack GB, Senter AE. (2011). 21st Century instream flow assessment framework for mountain streams. Sacramento (CA): California Energy Commission. PIER. CEC-500-2013-059.

Pasternack GB, Tu D, Wyrick JR. (2014). Chinook adult spawning physical habitat of the lower Yuba River. Davis (CA): University of California at Davis. Prepared for the Yuba Accord River Management Team.

Pasternack GB, Wang CL, Merz JE. (2004). Application of a 2D hydrodynamic model to design of reach-scale spawning gravel replenishment on the Mokelumne River, California. River Research and Applications 20:205–225. doi: 10.1002/rra.748.

Pasternack, G. B., Bounrisavong, M. K., & Parikh, K. K. (2008). Backwater control on riffle–pool hydraulics, fish habitat quality, and sediment transport regime in gravel–bed rivers. *Journal of Hydrology*, 357, 125–139.

Pasternack GB, Gilbert AT, Wheaton JM, Buckland EM. (2006). Error propagation for velocity and shear stress prediction using 2D models for environmental management. Journal of Hydrology 328:227–241.

Rantz SE. (1982). Measurement and computation of streamflow, Volume 1, Measurement of stage and discharge, U.S. Geol. Surv. Water Supply Pap. 2175, 1982.

Reinfelds, I., Lincoln–Smith, M., Haeusler, T., Ryan, D., & Growns, I. (2010). Hydraulic assessment of environmental flow regimes to facilitate fish passage through natural riffles: Shoalhaven River below Tallowa Dam, New South Wales, Australia. *River Research and Applications*, 26, 589–604.

Sawyer AM, Pasternack GB, Moir HJ, Fulton AA. (2010). Riffle–pool maintenance and flow convergence routing observed on a large gravel–bed river. Geomorphology. 114:143–160.

Smart, G.M. (1999). Turbulent velocity profiles and boundary shear in gravel bed rivers. Journal of Hydraulic Engineering 125:106–116.

Sontek/YSI Inc. (2007). RiverSurveyor system manual software version 4.60. San Diego: Sontek/YSI Inc.

Stewart, G. B. (2000). Two–dimensional hydraulic modeling for making instream–flow recommendations. Colorado State University (unpublished MS thesis).

Stockdale, R. J., McLelland, S. J., Middleton, R., & Coulthard, T. J. (2007). Measuring river velocities using GPS River Flow Tracers (GRiFTers). *Earth Surface Processes and Landforms*, 33, 1315-1322, DOI: 10.1002/esp. 1614.

Strom, M. A., Pasternack, G. B., & Wyrick, J. R. (2016). Reenvisioning velocity reversal as a diversity of hydraulic patch behaviours. *Hydrological Processes*, 30, 2348–2365.

Tayefi, V., Lane, S. N., Hardy, R. J., & Yu, D. (2007). A comparison of 1–D and 2–D approaches to modelling flood inundation over complex upland floodplains. *Hydrological Processes*, 21, 3190–3202.

Tiffan KF, Garland RD, Rondorf DW. (2002). Quantifying flow–dependent changes in subyearling fall chinook salmon rearing habitat using two–dimensional spatially explicit modeling. North American Journal of Fisheries Management 22:713–726.

Van Asselt MBA, Rotmans J. 2002. Uncertainty in integrated assessment modelling. Climatic Change 54: 75-105.

Waddle, T. (2010). Field evaluation of a two–dimensional hydrodynamic model near boulders for habitat calculation. *River Research and Applications*, 26, 730–741.

Wheaton JM, Pasternack GB, Merz JE. (2004). Spawning habitat rehabilitation–II. Using hypothesis development and testing in design, Mokelumne River, California, U.S.A. International Journal of River Basin Management 2:21–37.

White JQ, Pasternack GB, Moir HJ. (2010). Valley width variation influences riffle–pool location and persistence on a rapidly incising gravel-bed river. Geomorphology 121:206–221. doi: 10.1016/j.geomorph.2010.04.012.

Yu, D., & Lane, S. N. (2006). Urban fluvial flood modelling using a two–dimensional diffusion–wave treatment, part 1: mesh resolution effects. *Hydrological Processes*, 20, 1541–1565.

Zhang Y, Unger DR, Hung I-K, Kulhavy DL. (2014). WAAS, GLONASS, and GPS accuracy. The Forestry Source. December 2014.