https://doi.org/10.5194/egusphere-2025-6190
Preprint. Discussion started: 5 March 2026

An ADCP-Based Data-Driven Framework for Proxy Sediment Transport Monitoring: From Controlled Flumes to Natural Rivers

Mohammd Tanvir Haque Tuhin[1,2], Reinhard Hinkelmann[2], Christoph Mudersbach[1]

[1]Institute of Hydraulic Engineering and Hydromechanics, Department of Civil and Environmental Engineering, Bochum University of Applied Sciences, Bochum, 44801, Germany
[2]Chair of Water Resources Management and Modeling of Hydrosystems, Institute of Civil Engineering, Technische Universität Berlin, Berlin, 13355, Germany

Correspondence to: Mohammd Tanvir Haque Tuhin (mohammd.tuhin@hs-bochum.de)

Abstract.

Acoustic Doppler Current Profilers (ADCPs) provide a rich yet underutilized source for monitoring hydrodynamics and sediment transport. Accurate prediction of sediment-related variables is critical for river engineering, morphological studies, and environmental management. Among these, Bottom-Track Velocity (BT_Vel) serves as a robust proxy for near-bed sediment dynamics and bedload activity. This study develops a machine learning (ML) and deep learning (DL) framework to predict BT_Vel from ADCP-derived hydrodynamic and acoustic features, enabling proxy estimation of sediment transport processes in both controlled flume and natural riverine environments. Two datasets were analyzed: (i) a laboratory dataset of 22,650 ensemble samples obtained under controlled flow regimes, and (ii) a field dataset of 5,900 ensemble samples collected across seven campaigns at a fixed river cross-section. A consistent benchmarking strategy was applied across Random Forest, Gradient Boosting, LightGBM, CatBoost, XGBoost, LSTM, GRU, CNN, RNN, ANN, and a hybrid LSTM+CNN, with evaluation based on both an 80/20 split and a stratified 5-fold cross-validation (CV). SHAP analysis was conducted for model interpretability. In the laboratory, Random Forest (R^2 = 0.804 split / 0.783 CV) and Gradient Boosting (0.787 / 0.757) achieved the best generalization, while LSTM+CNN (0.770 / 0.730) and LSTM (0.775 / 0.718) remained competitive. In the field, Random Forest again delivered the strongest results (0.573 / 0.603), followed closely by CatBoost, LightGBM, and XGBoost. Notably, LSTM improved under cross-validation (0.468 → 0.529), suggesting fold-wise diversity stabilized training under noisy, heterogeneous river data. By contrast, the Stacking Regressor consistently showed the weakest generalization across both environments. SHAP revealed a shift in feature relevance: in the laboratory, Mean water velocity (Mean_Speed) dominated predictions, while in the field, Depth and signal-to-noise ratio (SNR) emerged as stronger drivers, reflecting the influence of stage variability and acoustic quality. Overall, the study demonstrates that ADCP-derived features, coupled with explainable ML/DL models, provide robust potential for proxy sediment transport modeling. Conversion to absolute transport rates requires paired sediment measurements, while future work should expand field campaigns and explore hybrid physics–data frameworks toward operational forecasting.

https://doi.org/10.5194/egusphere-2025-6190
Preprint. Discussion started: 5 March 2026

1 Introduction and Objectives

Hydrodynamics and sediment transport processes are fundamental to the evolution, stability, and management of riverine and coastal environments. These intertwined processes control channel morphology, influence flood risk, affect infrastructure, and sustain aquatic ecosystems (Garcia, 2008; Frings and Vollmer, 2017; Latosinski et al., 2017). Accurate monitoring of bedload and suspended sediment flux is notoriously challenging due to the need for in-situ sampling during dynamic flow events. Traditional methods, such as physical bedload samplers (e.g. Helley-Smith traps) or sediment traps are labor-intensive, costly, and often yield sparse, high-uncertainty data – particularly during high flows when field measurements become challenging and temporally limited (Mir et al., 2024; Tabesh et al., 2024; Emmett, 1980; Hubbell, 1964). In recent years, hydroacoustic methods have become increasingly important for hydrodynamics and sediment transport monitoring. Acoustic Doppler Current Profilers (ADCPs), originally developed for flow measurements, now also enable non-intrusive, continuous observation of sediment-related processes (Gartner, 2004; Kostaschuk et al., 2005). Through the emission of acoustic pulses and analysis of Doppler-shifted backscatter signals, ADCPs generate high-resolution velocity profiles (Simpson, 2001). Additionally, their bottom-tracking capability measures motion relative to the riverbed, offering valuable insight into bed dynamics (Rennie and Rainville, 2006). These features allow ADCPs to function as dual-purpose instruments, providing real-time measurements of both hydrodynamic behavior and sediment transport processes, including insight into bedload and suspended load dynamics. As such, they have become indispensable tools in hydraulic engineering investigations, both in controlled laboratory settings and complex natural environments.

A substantial body of research has established that ADCPs, through their bottom-tracking functionality, can detect the apparent movement of the riverbed – a phenomenon directly correlated with active bedload transport. Foundational studies by Rennie et al., (2002) and Rennie and Villard (2004) demonstrated that bottom-track velocity (BT_Vel) can serve as a reliable proxy for bedload flux in rivers, a conclusion later reinforced by Gaeuman and Jacobson, (2006) in large sand-bed rivers. Controlled flume studies by Ramooz and Rennie (2010) have shown a strong statistical agreement (R^2 up to 0.93) between ADCP-derived velocities and physically measured bedload rates, particularly under well-defined sediment and flow conditions. In large-scale riverine environments, Latosinski et al., (2017) and Le Guern et al., (2021) confirmed the feasibility of using ADCP-DGPS (Acoustic Doppler Current Profiler integrated with Differential GPS) systems to estimate bedload transport, reporting deviations within 15–30% of traditional dune-tracking methods when proper calibration was applied. These advances underscore the ADCP's robustness as a non-intrusive surrogate for bedload assessment, provided site-specific tuning is conducted to account for factors such as frequency, grain size, bedform dynamics, and flow depth (Gaeuman and Jacobson, 2006; Conevski et al., 2021; Conevski et al., 2023). In parallel, ADCP acoustic backscatter intensity – a measure of returned signal strength – has been extensively used as an indirect indicator of suspended sediment concentration (SSC). Guerrero et al., (2016) demonstrated that backscatter data, once empirically calibrated, correlate well with measured SSC in large rivers such as the Paraná River (Argentina) and the Danube River (Hungary). Recent

https://doi.org/10.5194/egusphere-2025-6190
Preprint. Discussion started: 5 March 2026

65 developments have begun integrating these two modalities: bottom-track velocity and backscatter intensity, allowing multi-dimensional characterization of sediment transport regimes. For instance, Conevski et al., (2023) showed that combining bed movement velocity with riverbed backscatter improved bedload estimation across varying sediment and flow conditions.

Despite these innovations, significant challenges remain. The relationship between acoustic signals and sediment
70 concentration is complex – affected by particle size, composition, and sound attenuation – requiring site-specific calibration (Thorne et al., 1991; Guerrero et al., 2016; Latosinski et al., 2017). Similarly, inferring bedload from bottom-track velocities is complicated by spatially variable bed mobility and sensor limitations. While flume studies confirm that apparent bedload velocities correlate with transport under controlled conditions, field settings introduce hydraulic and sediment heterogeneity (Conevski et al., 2020; Conevski et al., 2023). Thus, although ADCP reduces dependence on empirical rating curves,
75 reliability improves when multiple signals are fused using advanced analytics to capture nonlinear sediment-flow dynamics.

Machine learning (ML) and deep learning (DL) have emerged as powerful approaches to tackle such complex, multivariate problems in hydrology and geomorphology (Gacu et al., 2025). In particular, data-driven models can learn the relationship between observable signals (e.g. flow velocity, turbulence, acoustic intensity) and sediment responses, without requiring
80 explicit physical assumptions (Lund et al., 2022; Allawi et al., 2023). This is especially useful given the nonlinear and hysteretic nature of sediment transport, where sediment concentration for a given flow can differ between a rising and falling hydrograph (Jing et al., 2025; Achite et al., 2025). By training on historical or experimental data, ML algorithms can capture these details. Recent studies have applied data-driven models to sediment transport prediction with impressive success. For example, Lund et al., (2022) developed boosted decision-tree models (XGBoost) to predict suspended sediment and bedload
85 across Minnesota rivers, US; the models explained approximately 70 % of the variance in sediment datasets. Likewise, DL models such as Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNNs) have demonstrated superior performance in forecasting suspended sediment concentrations (SSC) by capturing temporal dynamics and spatial patterns beyond the reach of simpler methods (Pham Van et al., 2023; Chen et al., 2022). Allawi et al., (2023) demonstrated that an LSTM model can achieve correlations of ~0.97 for daily suspended sediment load prediction in a
90 tropical river, outperforming linear regression and other ML models. Similarly, Achite et al., (2025) found that a DL network consistently surpassed traditional rating curves and other ML/DL approaches, with NSE values near 0.99 in an Algerian catchment. These results underscore both the predictive power and the importance of interpretability and generalizability for deploying ML/DL in sediment transport applications.

95 To date, studies utilizing complete ADCP datasets for ML applications remain largely unexplored. Only recently have researchers begun to experiment with ML and DL to analyze and extract insights from ADCP output. For example, Ozsahin et al., (2025) applied a suite of ML models – including Random Forest, LightGBM, and MLP (Multilayer Perceptron) – to ADCP sensor outputs for river discharge prediction, demonstrating that ensemble models can achieve near-perfect accuracy

https://doi.org/10.5194/egusphere-2025-6190
Preprint. Discussion started: 5 March 2026

(AUC ROC up to 0.995) in estimating flow volumes from raw ADCP velocity profiles. In contrast, Kakarla et al., (2023) focused on improving the analysis and interpretation of ADCP data itself, employing hybrid ML approaches – including CNN, Principal Component Analysis (PCA), and clustering methods – to reconstruct and classify velocity patterns, thereby enhancing the information content and reliability of ADCP-based measurements. Conevski et al., (2025) showed that neural networks – including LSTM models – can predict bedload transport rates and particle sizes from stationary ADCP measurements with R^2 above 0.8. These studies collectively underscore the growing potential of integrating advanced data analytics with ADCP observations to enable more accurate and automated predictions of flow and sediment transport. However, the application of ML and especially deep learning to full ADCP datasets remains novel. While ADCPs produce rich, high-frequency multidimensional data such as velocities, turbulence metrics, and acoustic backscatter, traditional analyses often condense these to simple discharge calculations. By leveraging ML/DL, we can now extract subtle patterns from these signals to infer sediment transport proxies continuously and across broader spatial and temporal scales than is possible with physical sampling alone. This multivariate, data-driven approach – especially when combined with ML filtering – holds significant promise for advancing ADCP-based sediment transport monitoring in both experimental and real-world river settings.

In light of these developments, this study aims to integrate machine learning with ADCP data to enhance sediment transport and hydrodynamics monitoring. We leverage the rich outputs of ADCPs and feed them into state-of-the-art data-driven models to create an intelligent monitoring framework. Specifically, we focus on predicting the ADCP's Bottom-Track Velocity (BT_Vel) – a proxy indicator of bedload movement – using a broad suite of ML and DL models. We evaluate these models on two complementary datasets: a large-scale laboratory flume dataset with 22,650 ensemble samples collected under different flow rates, and a real-world field dataset with 5,900 ensemble samples from a single cross-section of the River Stever (North Rhine–Westphalia, Germany), gathered across seven campaigns in 2023-2024 to cover a range of natural flow conditions. By applying both a standard train–test split validation and a rigorous cross-validation, we compare model performance, generalization ability, and robustness to overfitting. We benchmark BT_Vel prediction on a laboratory dataset and a field dataset, evaluate models with both an 80/20 split and stratified 5-fold cross-validation, and report R^2 (dimensionless) and MSE (Mean Squared Error) in $(m\ s^{-1})^2$. The intent of this work is to systematically identify which data-driven models are most effective for ADCP-based hydrodynamics and sediment transport proxy prediction, and to extract practical insights into model behavior across both controlled laboratory and complex field environments. Through this approach, we fuse modern ML with ADCP hydroacoustic sensing to enable continuous, operationally useful monitoring of hydrodynamics and sediment transport. All code and datasets are publicly available to support transparency and reproducibility.

https://doi.org/10.5194/egusphere-2025-6190
Preprint. Discussion started: 5 March 2026

2 Methodology

2.1 Study Sites and Data Collection

2.1.1 Laboratory Flume Dataset Collection

Experiments were conducted in a 16-meter-long, 0.6-meter-wide laboratory flume at the Hydraulic Engineering and Hydromechanics Laboratory, Bochum University of Applied Sciences, Germany. The flume was operated with a fixed bed slope of 0.5 %. A sand bed with a median grain diameter of D_{50} = 0.45 mm was prepared, filling the first 11 m of the flume. To regulate flow and prevent unwanted sediment transport beyond the test section, a 4 cm high block was installed at the downstream end of the sand bed.

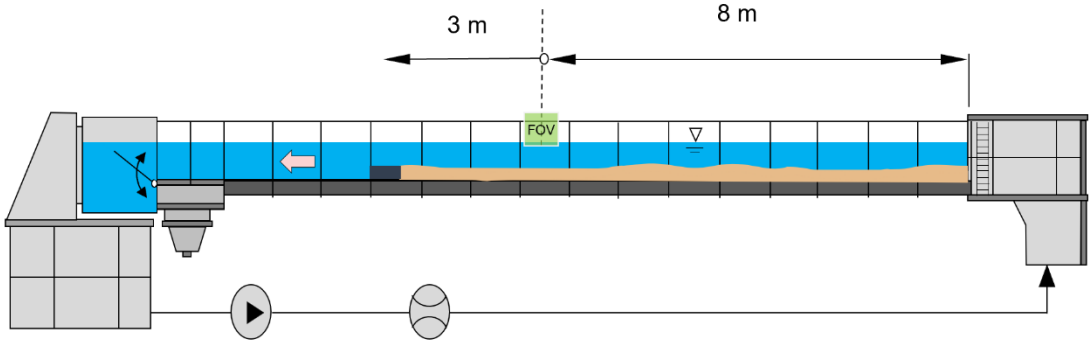


Fig. 1. Schematic of the 16 m laboratory flume showing the 11 m sand-bed section and the fixed ADCP Field of View (FOV) location.

The thickness of the sand bed was maintained at 3 to 4 cm throughout the experimental section. At 8 m downstream from the flume inlet, a SonTek RS5 ADCP (3 MHz broadband pulse-coherent) was installed in a stationary position, near the midpoint of the flume, to provide continuous flow profiling. The RS5 ADCP employs a four-beam Janus configuration (25° beam tilt, 3° beam width), supplemented by a vertical depth beam. The default profiling setup generates a series of closely spaced vertical cells with approximately 15 % overlap, ensuring seamless spatial coverage and robust acoustic signal returns. Once the flow was stabilized, data acquisition was initiated using SonTek RSQ software. The RS5 collected data at a 1 Hz sampling rate, with 30-ping ensembles used to form a single averaged measurement sample. For each run, data were logged for up to 1,620 time steps. Across fourteen different flow rates, this resulted in approximately 22,650 ensemble samples collected throughout the experimental campaign. Each ensemble sample thus corresponds to one time step - i.e., a full, vertically resolved velocity profile acquired by the ADCP under steady-state flow conditions.

https://doi.org/10.5194/egusphere-2025-6190
Preprint. Discussion started: 5 March 2026

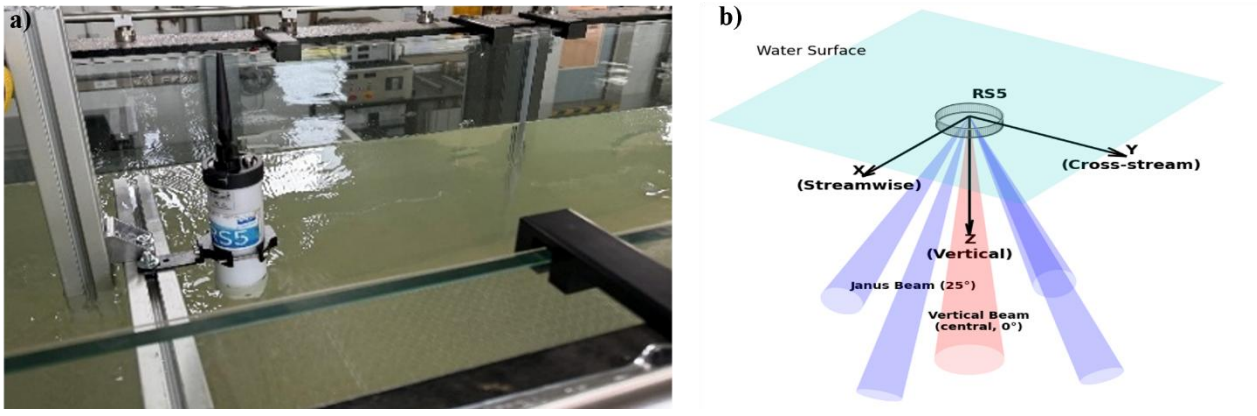


Fig. 2. Experimental setup and ADCP beam geometry. (a) RS5 acoustic Doppler current profiler (ADCP) mounted in an aluminum frame in the 16 m recirculating flume. (b) Schematic of the RS5 beam configuration showing the vertical beam (0°) and Janus beams (25°) along streamwise (X), cross-stream (Y), and vertical (Z) axes.

The flume discharge and flow depth were varied systematically across multiple runs to generate a range of hydraulic conditions. Under certain flows, migrating sand dunes and bedforms developed, providing varied sediment transport scenarios. The ADCP's Bottom-Track velocity readings were considered to detect bed movement during each run.

2.1.2 Field Dataset Collection

Field measurements were conducted on the River Stever, a mid-sized river in North Rhine-Westphalia, Germany. Data collection was performed at a fixed site across seven measurement campaigns between January 2023 and January 2024 (specifically: January 2023, June 2023, twice in August 2023, October 2023, December 2023, and January 2024). During each campaign, the survey focused on a single cross-section, where five to eight transects were performed to capture spatial variability across the section.

A remote-controlled rQPOD platform equipped with a SonTek RS5 ADCP was used to acquire the data. Each ensemble sample corresponds to a single vertical water velocity profile, and measurements were collected under a range of flow and seasonal conditions to represent the river's hydrodynamic variability throughout the year. In total, 5,900 ensemble samples were obtained from these campaigns. Compared to the laboratory data, the field dataset is characterized by higher uncertainty and greater heterogeneity and noise – attributable to factors such as uneven bathymetry, variable bed conditions, and instrument noise from the moving platform. Additionally, field measurements are temporally sparse, as data were acquired during discrete surveys rather than continuous monitoring.

https://doi.org/10.5194/egusphere-2025-6190
Preprint. Discussion started: 5 March 2026

Fig. 3. Field deployment of the RS5 on a remote-controlled survey boat (rQPOD) during measurements on the River Stever.

2.2 Pre-Processing and Feature Extraction

Raw ADCP data from both laboratory and field campaigns were stored as MATLAB .mat files, with each file representing a distinct flow rate (lab) or transect (field). For each dataset, matrix-type features (such as velocity standard deviation and SNR) were averaged across bins to produce one value per ensemble (time step), while scalar features (Mean_Speed, depth, etc.) were extracted directly. Bottom-Track velocities were pre-processed by setting negative values to zero and averaging across beams to remove non-physical measurements caused by signal dropouts. All pre-processed data were compiled into structured data frames, ensuring each sample corresponded to a single vertical velocity profile per ensemble. This harmonized dataset formed the basis for subsequent feature selection and ML analysis.

2.2.1 Physical Interpretation of ADCP-Derived Predictors

In both datasets, we derived a set of input features from the ADCP outputs to serve as predictors for Bottom-Track velocity (BT_Vel). Taking guidance from prior analyses of ADCP data (Tuhin and Mudersbach, 2025), we focused on seven features that characterize the flow's intensity, variability, and acoustic signal properties: Mean water velocity (Mean_Speed), the depth-averaged velocity magnitude and direct proxy for transport capacity (Ali et al., 2012; U.S. Army Corps of Engineers); Depth, as a measure of stage affecting both sediment mobility and signal behavior; Velocity Standard Deviation (Vel_StdDev), reflecting turbulence and bedform-induced fluctuations; Signal-to-Noise Ratio (SNR), a proxy for data quality and sediment-induced attenuation; Bin Distance, the transducer-to-cell distance, retained to capture range-dependent effects and allow future normalization strategies, though Relative Acoustic Backscatter (Relative_ABS) was only used for quality control; Expected Velocity Standard Deviation (Vel_Expected_StdDev), a diagnostic of instrument noise versus true

https://doi.org/10.5194/egusphere-2025-6190
Preprint. Discussion started: 5 March 2026

variability; and Correlation, an ADCP quality index indicating stability of acoustic returns and potential flow–sediment interactions. Together, these features capture the intensity, variability, and acoustic properties of flow, enabling proxy modeling of near-bed sediment transport dynamics.

The statistical ranges of the ADCP-derived variables differ markedly between the controlled flume environment and the natural river setting. These contrasts illustrate the very different hydraulic and acoustic conditions under which the proxy models were trained. In the laboratory dataset, flow conditions were tightly regulated, which produced narrow and predictable distributions. Mean_Speed remained between 0.04 and 0.54 m/s (mean 0.17 m/s), Depth was confined to 0.34–0.90 m, and turbulence levels were extremely low (Vel_StdDev mean about 0.0009 m/s). Acoustic signal quality was consistently high (SNR mean about 45.1), reflecting the short measurement range and the absence of suspended sediment. As expected, BT_Vel values were close to zero for most ensembles (mean about 0.0011 m/s, maximum 0.085 m/s), indicating that the sand bed remained essentially stationary during the controlled runs. The field dataset shows a very different pattern, capturing a much wider range of hydraulic conditions and a higher level of measurement uncertainty. Depth varied from 0 to 3.03 m (mean 1.45 m), and Mean_Speed covered more than an order of magnitude from 0 to 1.21 m/s (mean 0.25 m/s). These variations reflect seasonal changes in discharge and the spatial heterogeneity of the river cross-section. Turbulence intensity was also much higher (Vel_StdDev mean about 0.0088 m/s, maximum 0.074 m/s), and SNR values ranged broadly from 5.86 to 56.7 due to the combined effects of sediment concentration, beam range, and bed roughness. BT_Vel magnitudes were also larger in the field (mean 0.0211 m/s, maximum 0.1813 m/s), consistent with intermittent bed mobility that does not occur under controlled flume conditions.

Taken together, these contrasting distributions highlight the physical relevance of the selected predictors. Mean_Speed and Depth represent the hydraulic forcing that controls the onset of sediment motion. Vel_StdDev and Vel_Expected_StdDev help separate turbulence-driven variability from instrument noise. SNR, Correlation and Bin_Distance capture acoustic attenuation, signal quality, and range effects that often co-vary with sediment presence and flow structure. Although BT_Vel is an indirect proxy rather than a direct measure of sediment transport, it responds systematically to increases in flow strength and turbulence in both datasets. Overall, these observations establish a clear physical foundation for the proxy modelling approach and show how ADCP-derived hydraulic and acoustic variables relate to near-bed sediment dynamics.

https://doi.org/10.5194/egusphere-2025-6190
Preprint. Discussion started: 5 March 2026

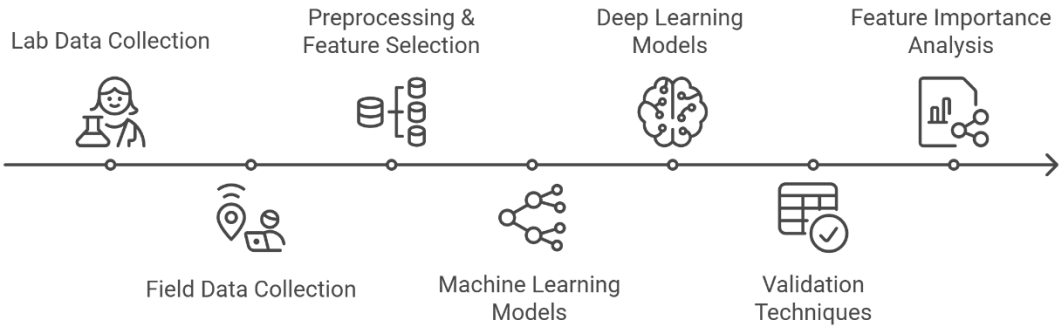


Fig. 4. Schematic workflow for Bottom-Track Velocity (BT_Vel) prediction.

2.3 Machine Learning Models

To evaluate the predictive capability of ADCP-derived features for sediment transport proxy estimation, several ensemble-based machine learning (ML) models were applied. These models were chosen for their proven performance in capturing nonlinear relationships and robustness under variable data conditions. All models were trained and evaluated using both a traditional train–test split (80/20) and 5-fold stratified cross-validation to assess generalization. Hyperparameters were unified across both laboratory and field experiments to ensure consistent comparative analysis, with parameters such as the number of estimators (n = 100) and learning rate (0.1, where applicable) selected based on domain knowledge and empirical validation. Features were standardized with StandardScaler fitted on the training portion only (and within each fold for cross-validation) to prevent data leakage. For regression stratification, the continuous target was discretized into 10 quantile-based bins. Performance was summarized using R^2, MSE, and MAE (Mean Absolute Error). To maintain a like-for-like comparison with the DL baselines, we report results from fixed baseline hyperparameters (no per-model tuning for ranking). Group-blocked validation (run-blocked in the laboratory, campaign-blocked in the field) was conceptually considered but not applied in the main benchmarking due to the small number of runs/campaigns; its implications are discussed in Section 4.5 (Limitations).

2.3.1 Ensemble Tree-Based Models

The study utilized five ensemble tree-based algorithms that highlight both bagging and boosting approaches. Random Forest is a bagging-based ensemble method that constructs multiple decision trees using bootstrapped datasets and random feature selection. It is known for its robustness to overfitting, ability to handle noisy or high-dimensional data, and strong performance across a range of regression tasks (Cutler et al., 2007). Gradient Boosting builds decision trees sequentially by minimizing prediction errors through gradient descent. It emphasizes residual fitting and is effective in capturing complex feature interactions, though it may be sensitive to noise and requires careful regularization (Friedman, 2001). Extreme

https://doi.org/10.5194/egusphere-2025-6190
Preprint. Discussion started: 5 March 2026

Gradient Boosting (XGBoost) enhances gradient boosting by incorporating L1/L2 regularization, column subsampling, and parallelization. Its computational efficiency and scalability make it a preferred choice for structured regression tasks in environmental and hydrological domains (Chen and Guestrin, 2016). LightGBM is a leaf-wise gradient boosting algorithm based on histogram binning that improves computational speed and memory efficiency, and its ability to handle large datasets and support categorical features make it well suited for environmental prediction applications (Ke et al., 2017). CatBoost is a boosting framework designed to handle categorical variables natively and mitigate prediction shift via ordered boosting. It is particularly advantageous when feature importance and generalization are prioritized under noisy conditions (Prokhorenkova et al., 2018). Together, these algorithms provide a comprehensive ensemble framework balancing accuracy, interpretability, and robustness across both laboratory and field datasets.

2.3.2 Stacking Ensemble

Stacking Regressor was employed to enhance generalization and integrate the complementary strengths of individual learners. The two-level ensemble architecture consisted of a base layer combining Random Forest and Gradient Boosting models, with a Gradient Boosting Regressor acting as the meta-learner. This architecture aims to balance bias-variance trade-offs by combining predictions from diverse model families. Such stacked approaches have shown improved robustness in hybrid modeling contexts, particularly where no single algorithm dominates across validation folds.

All models followed the preprocessing and validation protocol described above, including leak-free scaling (fit on train / per fold) and identical random seeds for reproducibility. Comparative performance across the laboratory and field datasets is reported in Sections 3.1.2 and 3.2.2.

2.4 Deep Learning Models

Six deep learning (DL) architectures were implemented identically for the laboratory and field datasets. These included a feed-forward ANN, a 1D CNN, three recurrent models (LSTM, GRU, RNN), and a hybrid, selected to represent common DL families used for structured and sequential hydroacoustic data. To ensure like-for-like comparison with the ML baselines and across DL families, we fixed one baseline per model and used the same preprocessing, loss, and evaluation regime. Features were standardized with a StandardScaler fitted on the training portion only (and per fold in cross-validation) to avoid leakage.

2.4.1 Evaluation protocols

Models were assessed under two regimes: an 80/20 hold-out split and 5-fold stratified cross-validation (CV) with 10 quantile bins on the continuous target. For the hold-out split, we swept batch size over {6, 8, 10, 12, 16, 20, 24, 32, 40, 64} using train-only validation; the batch size with the lowest validation loss was selected, after which the model was retrained on the training set with early stopping (patience 5–7; best weights restored) and evaluated once on the untouched test set. For CV, scaling was performed per fold, with batch_size = 32, ≈30 epochs, and the same early-stopping criterion. Evaluation metrics

https://doi.org/10.5194/egusphere-2025-6190
Preprint. Discussion started: 5 March 2026

included R^2 and MSE in $(m\ s^{-1})^2$ as primary indicators, while MAE in $(m\ s^{-1})$ was calculated after training for supplementary error analysis.

During development, two optimization strategies were tested. First, manual architecture refinements (e.g., a deeper ANN and heavily tuned LSTM) were applied, and relative model rankings were altered under specific conditions. Second, Bayesian hyperparameter optimization across learning rate, dropout, and batch size was conducted, which yielded only marginal improvements or slight performance losses, with overall rankings remaining unchanged. To maintain fairness and comparability across model families, standardized baseline configurations are reported in the main analysis, while it is acknowledged that architecture-specific refinements can further optimize individual models at the expense of cross-family comparability. Full baseline architectural details, including input layouts, layer configurations, and optimizer settings, are provided in Table S1 (Supplement) and are fully available on GitHub.

2.5 Feature Importance Analysis

To enhance interpretability, feature contributions were analyzed using SHAP (SHapley Additive exPlanations) values. Since tree-based models (Random Forest, Gradient Boosting, XGBoost, LightGBM, and CatBoost) provide efficient compatibility with SHAP's TreeExplainer, mean absolute SHAP values were computed directly for these models. For non-tree learners, KernelExplainer was applied on a subset of the training data to approximate local contributions. SHAP values were derived from the hold-out 80/20 split, rather than across all cross-validation folds, to avoid redundant fold-level computation. This choice was validated by confirming that feature importance rankings remained consistent between split and cross-validation, ensuring that the explanations are both stable and representative. The final SHAP analysis thus reflects model behavior in a manner consistent with the broader evaluation framework, while prioritizing interpretability and computational efficiency.

2.6 Classical Baselines and Justification

Classical sediment transport formulas such as Meyer–Peter–Müller, Engelund–Hansen and van Rijn rely on quantities like bed shear stress, shear velocity, the critical Shields parameter and actual sediment load. None of these variables are available in our ADCP-only dataset. Since the aim of this study is to model Bottom-Track velocity as a proxy signal, rather than compute sediment transport rates directly, applying these classical formulas would not match the nature of the measurements. To provide a simple baseline using the variables we do have, we tested a standard Linear Regression model. Its performance was poor (mean R^2 approximately 0.25), showing that BT_Vel cannot be explained through linear combinations of the ADCP predictors. This reinforces that near-bed hydrodynamics behave in a nonlinear way and supports the use of more flexible machine-learning and deep-learning models, which achieved substantially higher predictive skill throughout our experiments.

https://doi.org/10.5194/egusphere-2025-6190
Preprint. Discussion started: 5 March 2026

3 Results

3.1 Laboratory Dataset

3.1.1 Feature Correlation Analysis

To better understand the internal relationships among variables in the laboratory dataset, a correlation heatmap was produced (Fig. 5). The target variable, Bottom-Track Velocity (BT_Vel), displays a moderate association with Mean_Speed ($r \approx 0.43$) and Bin Distance ($r \approx 0.26$), suggesting that higher bulk flow and increased measurement range are linked to elevated near-bed velocities. Strong mutual correlations among Depth, Mean_Speed, and Velocity Standard Deviation ($r > 0.85$) reflect their joint response to increasing discharge levels imposed in the flume, confirming their coherence as indicators of energetic flow conditions. Bin Distance also exhibits high correlation with these variables, though this arises from its dependence on measurement geometry (i.e., increasing depth extends the sampling range), rather than representing a dynamic flow variable.

In contrast, Correlation, a quality control metric derived from acoustic return consistency, exhibits noticeable negative correlations with flow-dependent features, especially Vel_StdDev ($r \approx -0.64$). This pattern suggests a reduction in signal coherence during turbulent or mobile-bed phases. SNR, while less strongly correlated overall, shows inverse tendencies with variables such as Mean_Speed and Vel_StdDev, hinting at a potential influence of acoustic scattering due to sediment mobilization or depth-induced signal loss. Together, these relationships provide valuable context for downstream model performance and justify the use of these features in the predictive framework, while also highlighting areas of potential redundancy and complementarity.

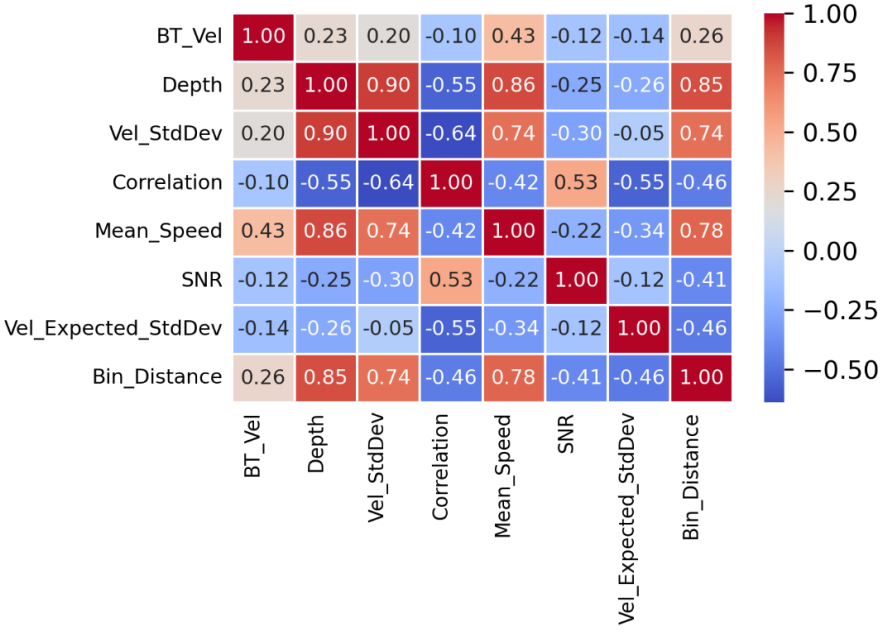


Fig. 5. Feature correlation heatmap for the laboratory dataset.

https://doi.org/10.5194/egusphere-2025-6190
Preprint. Discussion started: 5 March 2026

3.1.2 Model Performance

To evaluate predictive accuracy and generalization, R^2 was computed for each model using both an 80/20 train–test split and stratified 5-fold cross-validation (Fig. 6). Among machine-learning models, Random Forest (R^2 = 0.804 split / CV = 0.783) and Gradient Boosting (0.787 / 0.757) consistently outperformed others, demonstrating strong generalization across validation regimes. Ensemble tree models were also competitive: LightGBM (0.752 / 0.743), CatBoost (0.741 / 0.741), and XGBoost (0.747 / 0.731) captured the nonlinear structure of the laboratory dataset effectively.

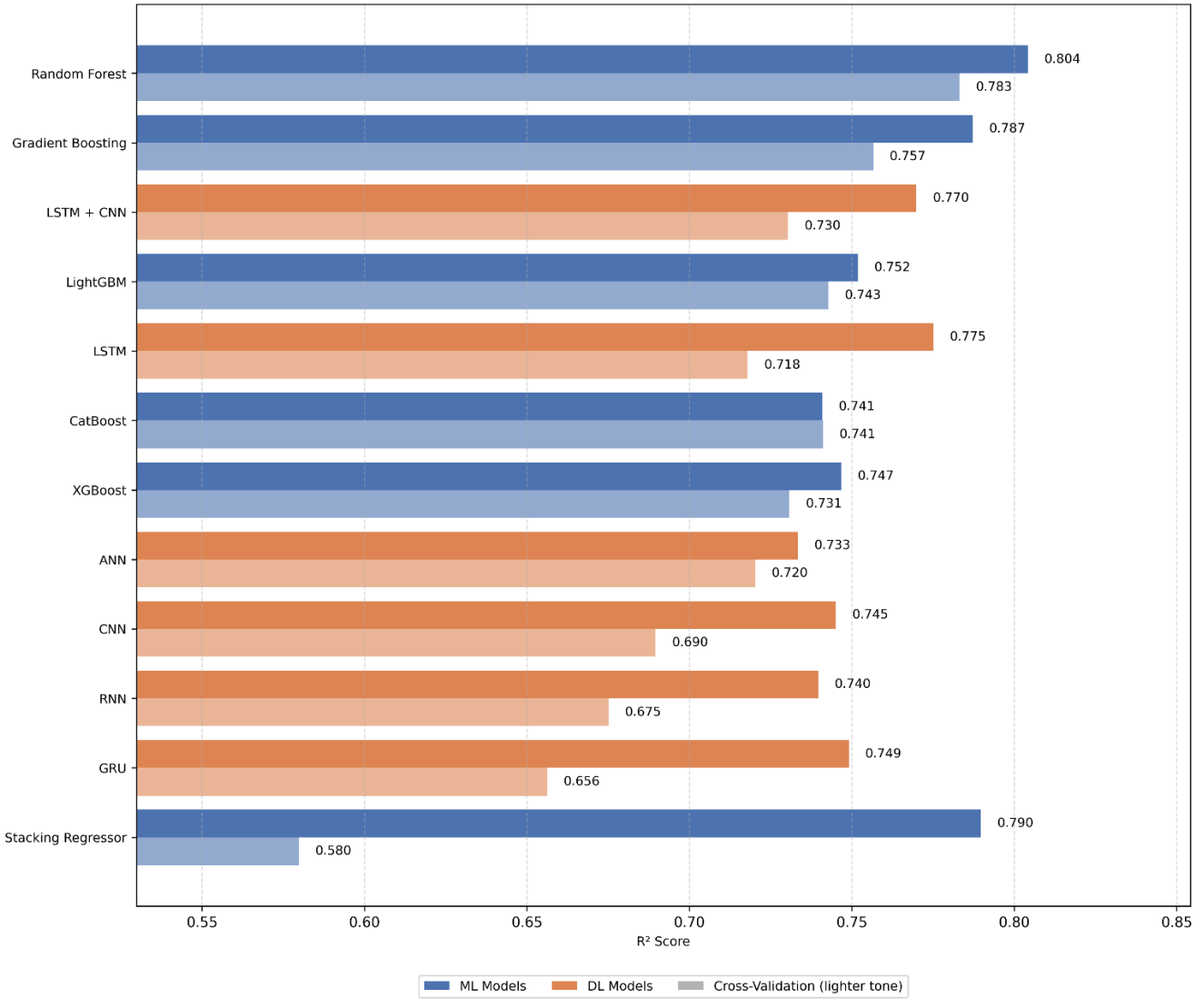


Fig. 6. R^2 comparison for Bottom-Track Velocity (BT_Vel) prediction using the laboratory dataset. Results are shown for ML (blue) and DL (orange) models, with darker bars for train–test split (80/20) and lighter bars for stratified cross-validation (5-fold).

https://doi.org/10.5194/egusphere-2025-6190
Preprint. Discussion started: 5 March 2026

Deep learning models showed nuanced behavior. The strongest baselines were LSTM+CNN hybrid (0.770 / 0.730) and the LSTM (0.775 / 0.718), trailing the top ML ensembles but remaining competitive. ANN (0.733 / 0.720), CNN (0.745 / 0.690), RNN (0.740 / 0.675), and GRU (0.749 / 0.656) exhibited larger drops from split to CV, indicating mild overfitting under the modest lab sample size.

In contrast, the Stacking Regressor showed the weakest generalization (0.790 / 0.580), suggesting instability of the meta-learner across folds. Overall, models with embedded memory (LSTM+CNN/ LSTM) or ensemble learning capacity (Random Forest/Gradient Boosting) provided the best balance between fit and generalization, while deeper recurrent variants benefited less without additional regularization or data.

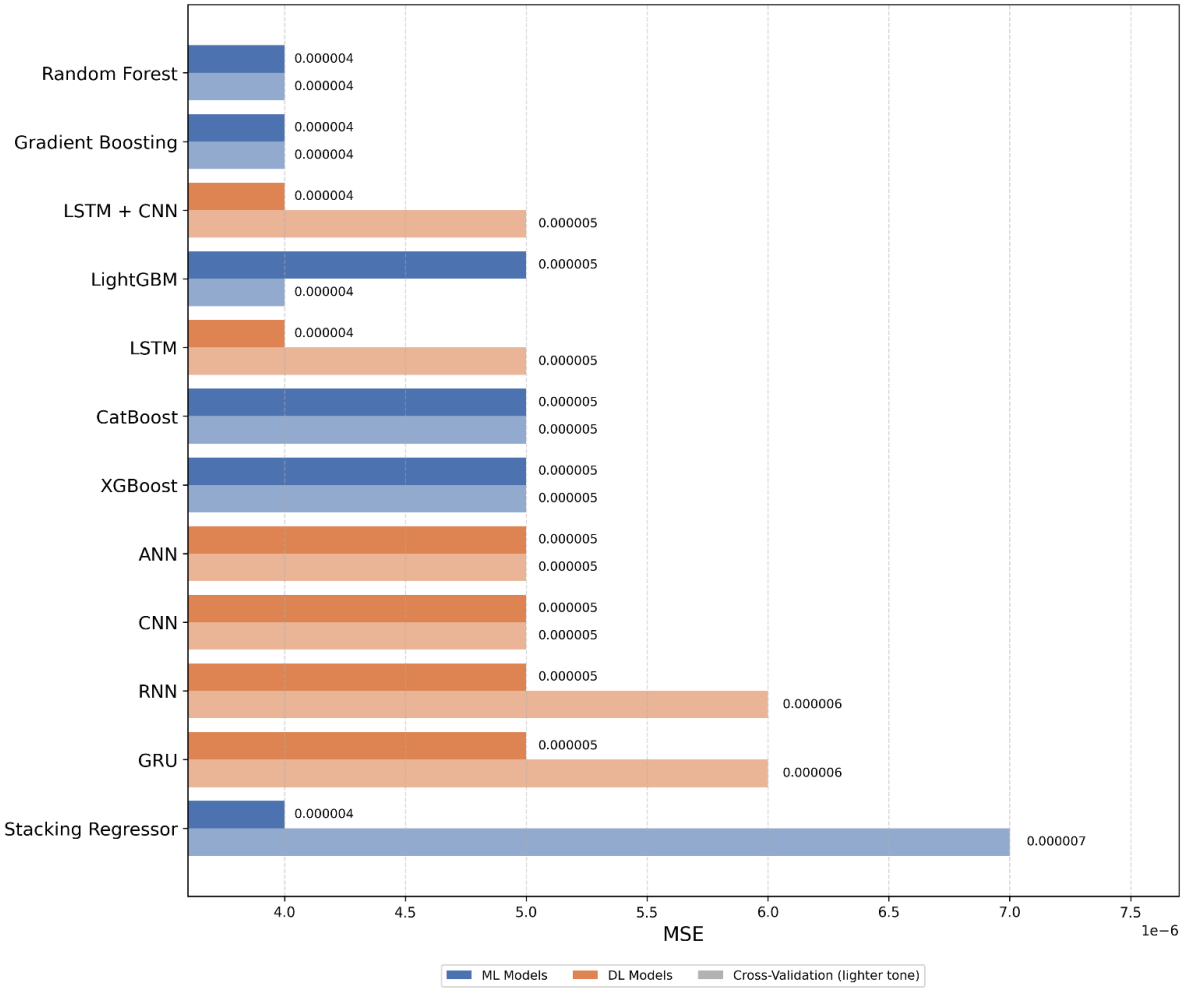


Fig. 7. Mean Squared Error (MSE, in $(m\ s^{-1})^2$) comparison for Bottom-Track Velocity (BT_Vel) prediction using the laboratory dataset. Results are shown for ML (blue) and DL (orange) models, with darker bars for train–test split (80/20) and lighter bars for stratified cross-validation (5-fold).

https://doi.org/10.5194/egusphere-2025-6190
Preprint. Discussion started: 5 March 2026

To complement the R^2 analysis, model errors were also assessed using Mean Squared Error (MSE, in $(m\ s^{-1})^2$; Fig. 7). Consistent with the R^2 trends, Random Forest and Gradient Boosting achieved the lowest errors ($\approx 4\times10^{-6}$ in both split and cross-validation), indicating strong stability. LSTM+CNN and LSTM showed comparably low split error ($\approx 4\times10^{-6}$) with a modest rise under cross-validation ($\approx 5\times10^{-6}$). A middle tier – LightGBM, XGBoost, CatBoost, ANN, and CNN – clustered around 5×10^{-6}, with only minor split–CV differences. In contrast, GRU and RNN exhibited higher errors ($\approx 6\times10^{-6}$), and the Stacking Regressor showed the largest degradation, increasing from $\approx 4\times10^{-6}$ (split) to $\approx 7\times10^{-6}$ (CV). These results mirror the R^2 ranking and reinforce the robustness of ensemble tree methods (and the best recurrent baseline) for structured flume data, while highlighting greater variance sensitivity in simpler recurrent and stacked ensembles under cross-validation.

3.1.3 Uncertainty under Cross-Validation

The fold-to-fold variability of cross-validation results is summarized in Table S2 (Supplement). Ensemble tree models exhibited the narrowest uncertainty ranges, with Random Forest ($R^2 = 0.783 \pm 0.011$, MSE (in $(m\ s^{-1})^2$) $= (4.0 \pm 0.2)\times10^{-6}$) and Gradient Boosting (0.757 ± 0.019, $(4.0 \pm 0.1)\times10^{-6}$) showing stable generalization. Other boosting approaches (LightGBM, CatBoost, XGBoost) also maintained low dispersion (±0.01–0.02 R^2). In contrast, deep learning models were more sensitive to fold composition: LSTM+CNN (±0.04 R^2), LSTM (±0.05 R^2), and CNN (±0.06 R^2) showed higher variability, while GRU was the most unstable (±0.11 R^2, $\pm1.7\times10^{-6}$ MSE). The Stacking Regressor also displayed elevated variance (±0.07 R^2). These findings confirm that tree-based ensemble ML models were both accurate and robust under limited laboratory data, while DL models exhibited greater fold sensitivity.

3.1.4 Feature Importance

To interpret the contribution of individual input variables in predicting BT_Vel, SHAP values were aggregated across all applied ML models on the laboratory dataset (Fig. 8).

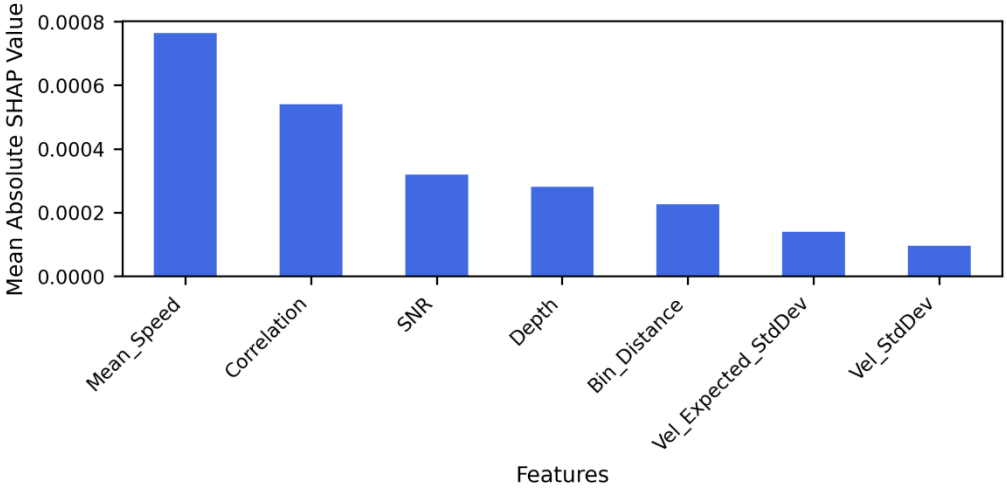


Fig. 8. Feature importance analysis using SHAP values for machine learning models applied to the laboratory dataset.

https://doi.org/10.5194/egusphere-2025-6190
Preprint. Discussion started: 5 March 2026

Mean_Speed emerged as the most influential feature, with a mean absolute SHAP value substantially higher than all others, affirming its dominant role in representing energetic flow conditions in the controlled flume environment. Correlation ranked second, followed by SNR, reflecting their combined ability to capture data quality and turbulence-induced variability. Depth and Bin Distance contributed moderately, consistent with their representation of hydraulic setting and sensor geometry. In contrast, velocity-based uncertainty metrics (Vel_Expected_StdDev and Vel_StdDev) showed the lowest SHAP magnitudes, aligning with their weaker direct linkage to near-bed dynamics in this setting. Overall, the feature ranking supports earlier correlation analyses and highlights the primacy of flow-driven bulk properties in driving accurate sediment proxy predictions under structured laboratory conditions.

3.2 Field Dataset

3.2.1 Feature Correlation Analysis

To evaluate feature interdependencies within the field dataset, a correlation heatmap was constructed (Fig. 9). Bottom-Track Velocity (BT_Vel) exhibits its strongest correlation with Mean_Speed ($r \approx 0.64$), followed by Depth ($r \approx 0.55$), indicating that faster, deeper flows tend to coincide with enhanced near-bed movement. Unlike the lab dataset, where certain variables were tightly coupled due to controlled discharge conditions, the field data display more moderate and dispersed correlations, reflecting the natural variability in hydrodynamics and sediment conditions across measurement campaigns.

Mean_Speed and Depth retain a moderate correlation ($r \approx 0.66$), with Velocity Standard Deviation also aligning somewhat with these energetic indicators ($r \approx 0.52$ with Mean_Speed). Bin Distance remains loosely associated with Depth ($r \approx 0.36$), but the relationship is weaker than in the lab, due to less systematic variation in ADCP deployment depth across surveys.

Correlation again emerges as a negatively correlated quality control feature, particularly with Velocity Standard Deviation ($r \approx -0.45$) and Mean_Speed ($r \approx -0.48$), consistent with signal degradation in turbulent or sediment-laden flows. SNR shows minimal correlation with most features, apart from a moderate inverse relationship with Vel_Expected_StdDev ($r \approx -0.64$), likely due to acoustic interference or signal attenuation in noisy flow conditions. Overall, the field heatmap reveals a more heterogeneous and less structured set of relationships, underscoring the increased complexity of modeling sediment transport proxies under real-world flow dynamics.

https://doi.org/10.5194/egusphere-2025-6190
Preprint. Discussion started: 5 March 2026

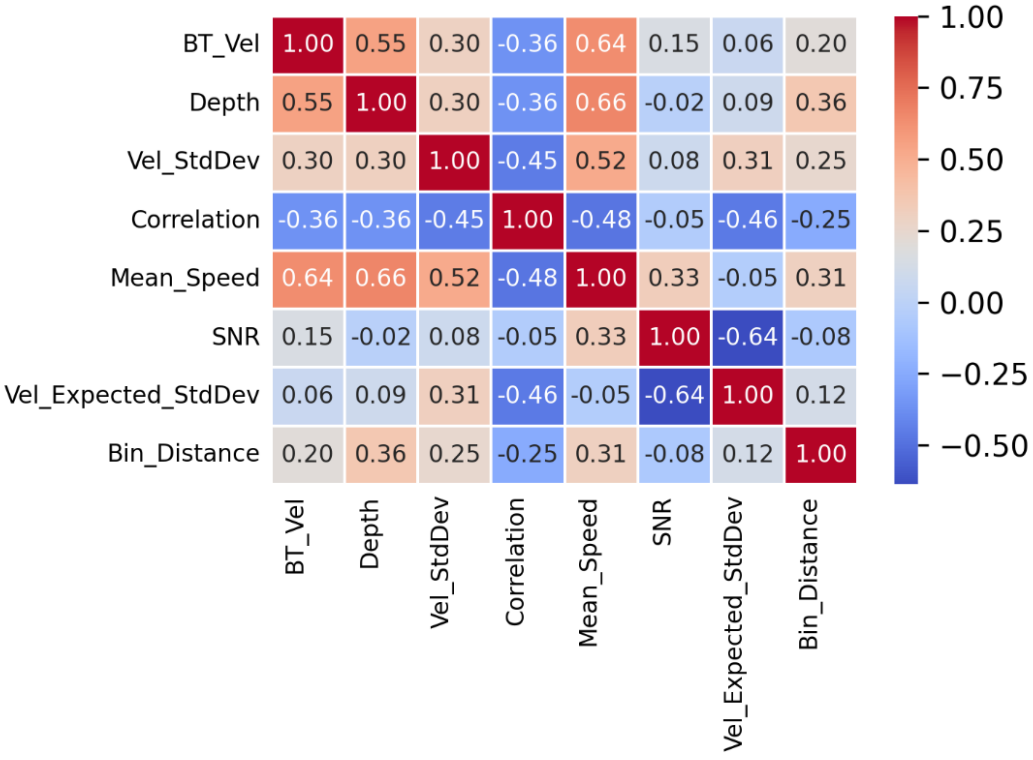
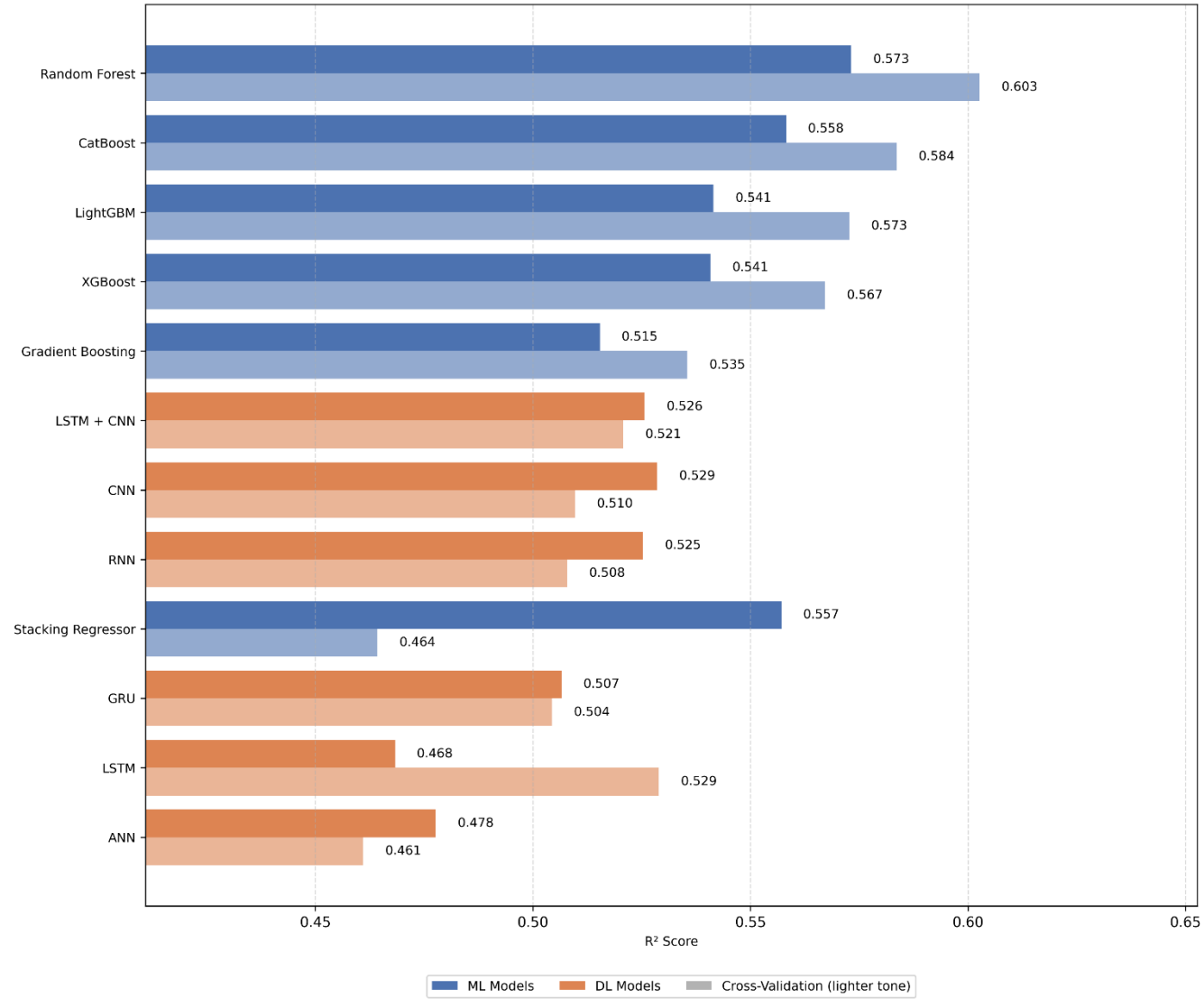


Fig. 9. Feature correlation heatmap for the field dataset.

3.2.2 Model Performance

To evaluate predictive reliability under real-river conditions, R^2 was computed for each model using both an 80/20 train–test split and stratified 5-fold cross-validation (Fig. 10). Tree-based ML models remained the most reliable. Cross-validation slightly exceeded the single split for all ensembles – suggesting stable generalization and that the held-out split was marginally harder: Random Forest (0.573 split / 0.603 CV), CatBoost (0.558 / 0.584), LightGBM (0.541 / 0.573), XGBoost (0.541 / 0.567), and Gradient Boosting (0.515 / 0.535). In contrast, the Stacking Regressor dropped from 0.557 (split) to 0.464 (CV), indicating overfitting to the base-learner outputs.

https://doi.org/10.5194/egusphere-2025-6190
Preprint. Discussion started: 5 March 2026



405

Fig. 10. R^2 comparison for Bottom-Track Velocity (BT_Vel) prediction using the field dataset. Results are shown for ML (blue) and DL (orange) models, with darker bars for train–test split (80/20) and lighter bars for stratified cross-validation (5-fold).

Among DL models, performance trailed the top ML methods and showed modest sensitivity to folds: LSTM+CNN (0.526 410 split / 0.521 CV), CNN (0.529 / 0.510), RNN (0.525 / 0.508), GRU (0.507 / 0.504), LSTM (0.468 / 0.529) and ANN (0.478 / 0.461). Notably, LSTM improves from 0.468 (split) to 0.529 (CV), indicating a more favorable performance under cross-validated folds.

https://doi.org/10.5194/egusphere-2025-6190
Preprint. Discussion started: 5 March 2026

Overall, ensemble trees delivered the best and most consistent generalization on the field dataset, while DL models remained competitive but more fold-sensitive; the Stacking Regressor showed the least robustness under cross-validation.

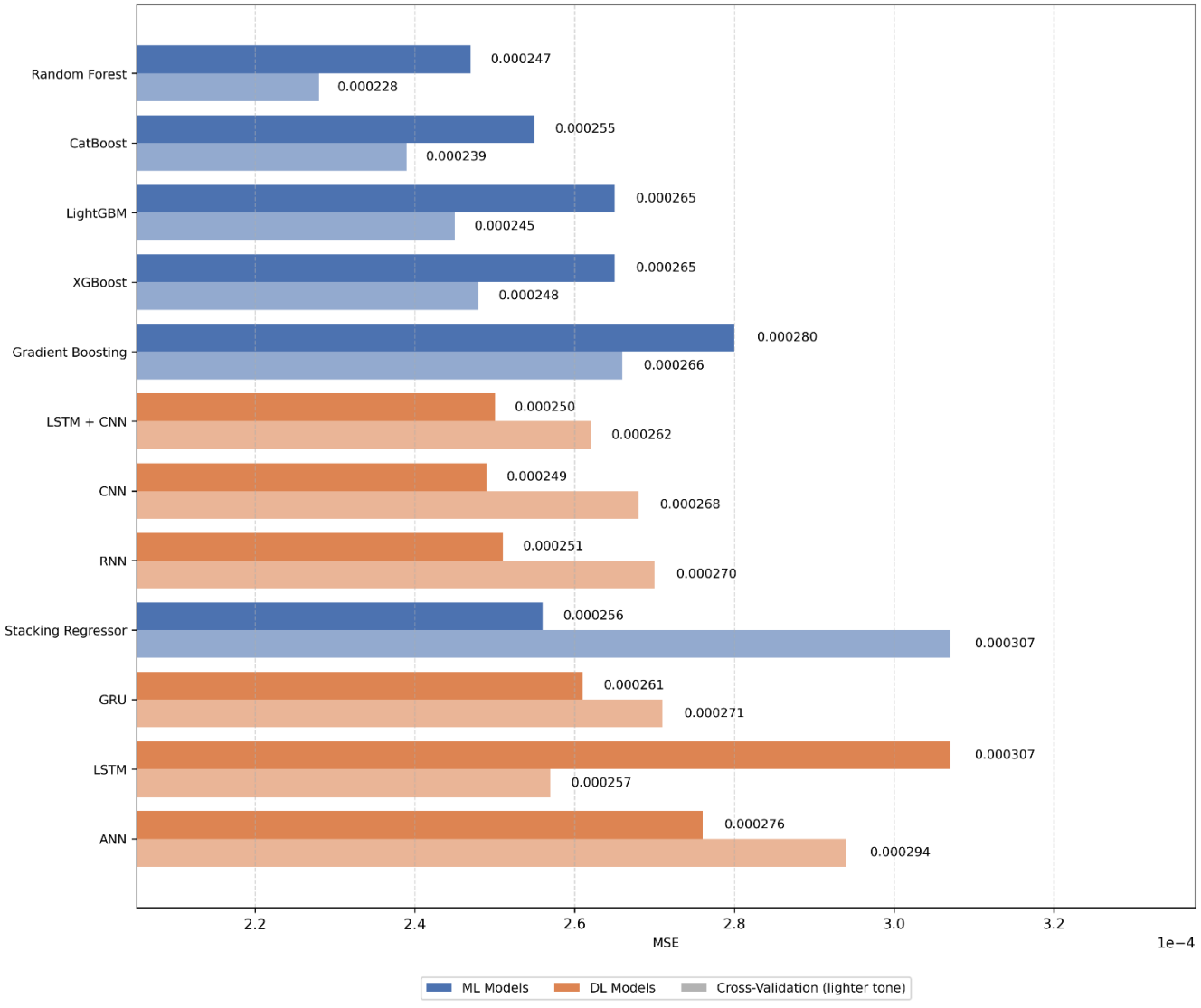


Fig. 11. Mean Squared Error (MSE, in $(m\ s^{-1})^2$) comparison for Bottom-Track Velocity (BT_Vel) prediction using the field dataset. Results are shown for ML (blue) and DL (orange) models, with darker bars for train–test split (80/20) and lighter bars for stratified cross-validation (5-fold).

To complement the R^2 analysis, we compared Mean Squared Error (MSE, in $(m\ s^{-1})^2$); Fig. 11). Consistent with the R^2 ranking, Random Forest achieved the lowest errors – 2.47×10^{-4} (split) and 2.28×10^{-4} (CV) – with CatBoost (2.55×10^{-4} split / 2.39×10^{-4} CV), LightGBM (2.45×10^{-4} / 2.65×10^{-4}), XGBoost (2.65×10^{-4} / 2.48×10^{-4}), and Gradient Boosting (2.80×10^{-4} / 2.66×10^{-4}) close behind. In contrast, the Stacking Regressor deteriorated most ($2.56\times10^{-4} \rightarrow 3.07\times10^{-4}$), confirming reduced

https://doi.org/10.5194/egusphere-2025-6190
Preprint. Discussion started: 5 March 2026

robustness across folds. Among DL, LSTM+CNN (2.50×10^{-4} split / 2.62×10^{-4} CV), CNN (2.49×10^{-4} / 2.68×10^{-4}), RNN (2.51×10^{-4} / 2.70×10^{-4}), and GRU (2.61×10^{-4} / 2.71×10^{-4}) exhibited slightly higher errors and modest CV increases. LSTM improved under CV ($3.07\times10^{-4} \rightarrow 2.57\times10^{-4}$), while ANN showed the opposite trend like other DL models ($2.76\times10^{-4} \rightarrow 2.94\times10^{-4}$). Overall, tree-based ensemble ML models delivered the smallest and most stable errors, while DL models were more fold-sensitive, with Stacking Regressor showing the weakest generalization.

3.2.3 Uncertainty under Cross-Validation

The variability of model skill across folds for the field dataset is summarized in Table S3 (Supplement). Ensemble models again showed the lowest dispersion, with Random Forest ($R^2 = 0.603 \pm 0.044$, MSE (in $(m\ s^{-1})^2$) = $(2.28 \pm 0.32)\times10^{-4}$) and CatBoost ($0.584 \pm 0.046$, $(2.39 \pm 0.33)\times10^{-4}$) demonstrating stable behavior. LightGBM and XGBoost exhibited similar uncertainty ranges (±0.04–0.05 R^2). Gradient Boosting and the LSTM baseline performed moderately, with variability near ±0.03–0.04 R^2. Among the other DL models, LSTM+CNN (±0.03 R^2), CNN (±0.03 R^2), RNN (±0.02 R^2), and GRU (±0.02 R^2) remained within comparable uncertainty ranges, though their mean accuracy was lower than ensemble trees. ANN showed the widest variation (±0.06 R^2), reflecting its less stable training on heterogeneous field data. The Stacking Regressor was again the least robust (0.464 ± 0.044 R^2, $(3.07 \pm 0.35)\times10^{-4}$ MSE). Overall, these patterns indicate that tree-based ensemble ML models provided the most consistent performance, while DL models displayed modestly higher fold sensitivity.

3.2.4 Feature Importance

SHAP-based interpretability applied to the field-trained ML models revealed a comparable but nuanced pattern of feature relevance (Fig. 12).

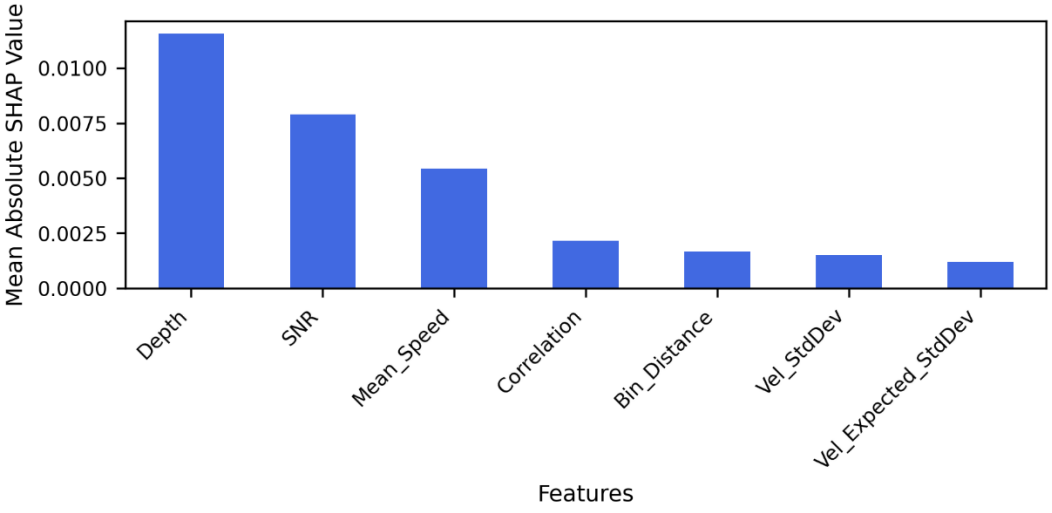


Fig. 12. Feature importance analysis using SHAP values for ML models applied to the field dataset.

https://doi.org/10.5194/egusphere-2025-6190
Preprint. Discussion started: 5 March 2026

Depth emerged as the most influential predictor, reflecting its strong association with stage-dependent transport capacity and its direct control on acoustic measurement geometry. SNR ranked second, capturing the quality of acoustic returns under varying sediment loads and flow turbulence, while Mean_Speed was the third most important driver, confirming its role as a fundamental proxy for transport energy in natural settings. Correlation contributed moderately, likely reflecting data stability and turbulence-induced variability. Bin Distance, Vel_StdDev, and Vel_Expected_StdDev showed comparatively minor importance, suggesting that under heterogeneous field conditions, range-dependent and noise-related metrics contributed less to predictive skill than bulk hydraulic indicators. Overall, the SHAP ranking highlights a shift from flume conditions, where Mean_Speed dominated, to field settings, where Depth and SNR provided stronger signals – emphasizing the influence of stage variability and acoustic quality in real-river measurements.

4 Discussion

4.1 Interpretation of Results

The results confirm that Acoustic Doppler Current Profiler (ADCP)–derived features – particularly Bottom-Track Velocity (BT_Vel), Mean_Speed, and Depth – are informative surrogates for sediment-transport dynamics in both controlled and natural settings. In the laboratory, where signals are comparatively consistent and low-noise, most models – especially tree-based ensembles and sequence-aware DL – achieved high predictive skill (R^2 frequently >0.75), with low MSE (~4×10^{-6}) for the best performers. In the field, the combination of smaller sample size, temporal sparsity, heterogeneous hydraulics, and instrument noise lowered absolute accuracy (typical $R^2 < 0.60$). Even so, multiple models retained stable generalization under cross-validation, indicating that the chosen feature set captures physically meaningful variability in river conditions.

A consistent pattern emerged across metrics: ensemble trees offered the most reliable accuracy–variance trade-off, while base (non-tuned) DL models showed greater fold sensitivity. Notably, LSTM behaved differently across contexts – strong in the lab split but softer under CV, yet improving under CV in the field (R^2 rising from the split to CV and MSE falling from ~3.07×10^{-4} to ~2.57×10^{-4}), suggesting that exposure to diverse folds stabilized its training on heterogeneous river profiles. By contrast, the Stacking Regressor repeatedly overfit – good split scores but pronounced CV degradation, especially in the field – highlighting its susceptibility to overfitting without stronger regularization or nested cross-validation.

It is also important to situate the data-driven framework relative to traditional sediment monitoring methods. Classical approaches such as Helley-Smith bedload samplers, dune-tracking, or sediment rating curves remain indispensable but are inherently episodic, labor-intensive, and prone to high uncertainty during peak flows. By contrast, ADCP-based proxies combined with ML/DL enable continuous, non-intrusive monitoring across entire hydrographs, capturing both rising and falling limbs where hysteresis effects dominate. While this framework cannot fully replace physical sampling, it

https://doi.org/10.5194/egusphere-2025-6190
Preprint. Discussion started: 5 March 2026

substantially reduces dependence on sparse measurements and enhances operational forecasting capacity, particularly in morphodynamically active rivers.

4.2 Model Performance Across Contexts

Cross-validation revealed critical insights that were not apparent in simpler split-validation approaches. In both datasets, split scores were generally optimistic relative to CV for complex models. This effect was most visible for deep networks in the lab (e.g., LSTM and LSTM+CNN: high split $R^2 \approx 0.77–0.78$, but lower CV), whereas tree ensembles retained strong generalization with only small differences between split and CV. In the field, most DL models still showed modest declines from split to CV (e.g., CNN, RNN, GRU), but LSTM was the exception, improving in CV – consistent with the model benefiting from fold-wise diversity in a noisy, sparse setting.

Random Forest consistently ranked highest across both laboratory and field datasets, with CatBoost, LightGBM, and XGBoost close behind. This group of ensemble models showed strong stability in both R^2 and MSE: performance under cross-validation generally matched or showed only slight declines relative to split evaluation, indicating robust generalization to varied fold compositions in the lab. Gradient Boosting, while the second-best performer in the lab ($R^2 \approx 0.79$ split), dropped to the lowest among the ensemble methods in the field ($R^2 \approx 0.52–0.54$). This suggests that its sequential residual-fitting mechanism is more sensitive to noise and sparsity in real-river data. Notably, in the field all five ensemble methods recorded lower MSE and higher R^2 under cross-validation than under split evaluation, underscoring their resilience when trained across more diverse fold partitions. Taken together, ensemble trees emerged as the most reliable class of models for both structured flume data and noisier field conditions, with Random Forest providing the best balance of accuracy and robustness.

Under the like-for-like design (shared baseline architectures, no model-specific tuning), DL models were competitive in the lab – particularly LSTM and the LSTM+CNN hybrid, which matched top ML models in split evaluation. However, their cross-validation results exposed greater fold sensitivity, with modest R^2 drops and higher MSE compared to tree ensembles. In the field, most DL models (CNN, LSTM+CNN, RNN, GRU) showed similar modest degradation in CV, consistent with the challenges of learning from sparse, noisy profiles. The key exception was LSTM, which improved under cross-validation in both R^2 and MSE, suggesting that recurrent memory benefits from the diversity of fold partitions. By contrast, ANN worsened under CV, reinforcing that shallow feed-forward networks are less resilient when confronted with noisy, heterogeneous river inputs. Taken together, DL models show promise but remain more data- and fold-sensitive than ensemble trees, with recurrent structures offering the most robustness.

The Stacking Regressor consistently showed the weakest generalization across both domains. In the laboratory, it produced relatively high split-validation scores but collapsed under cross-validation, exposing severe overfitting. In the field dataset, it

https://doi.org/10.5194/egusphere-2025-6190
Preprint. Discussion started: 5 March 2026

again exhibited notable R^2 and MSE deterioration under CV, underscoring its instability when exposed to heterogeneous folds. Without stronger regularization, more careful base-learner selection, or nested CV frameworks, stacking tends to overfit to single splits and underperform in generalized testing.

4.3 Uncertainty under Cross-Validation

The uncertainty analysis reinforces that robustness patterns closely followed mean performance trends in both datasets. Ensemble trees consistently showed narrow fold-to-fold variability in R^2 (typically within ± 0.01 – 0.05), confirming their stability under limited laboratory samples and heterogeneous field conditions. DL models, while often competitive in mean accuracy, exhibited higher variance across folds, particularly ANN and GRU, with CNN and LSTM also showing moderate sensitivity. These results suggest that ensemble methods provide the most reliable baseline under current data constraints, whereas DL models require larger and more diverse datasets to achieve consistent generalization.

4.4 Feature Relevance and Interpretability

The SHAP-based feature attribution aligned closely with physical expectations while revealing differences between controlled and natural environments. In the laboratory dataset, Mean_Speed was the most influential driver, reflecting its dominant role in representing energetic, uniform flow conditions, with Depth and Bin Distance contributing secondarily. In contrast, the field dataset highlighted Depth as the primary determinant, consistent with stage-dependent transport capacity, followed by SNR, which reflects the influence of acoustic signal quality under varying sediment concentrations and turbulence. Mean_Speed, while still important, ranked lower in the field than in the laboratory.

These shifts indicate that in controlled settings, flow intensity metrics dominate, whereas in natural rivers, variability in depth and acoustic quality become stronger indicators of sediment transport dynamics. Bin Distance, Correlation, and velocity-based uncertainty measures (Vel_StdDev, Vel_Expected_StdDev) consistently played minor roles in both contexts, reinforcing that bulk hydraulic indicators provide the most robust predictive signals. The coherence between SHAP interpretability and model performance underscores the value of integrating both hydrodynamic and acoustic features when developing proxy models for sediment transport.

4.5 Limitations

This study has several limitations. Most critically, the field dataset was restricted in spatial and temporal scope, comprising only seven campaigns at a single cross-section within one year. A campaign-blocked validation (i.e., holding out one entire campaign at a time) would therefore yield highly variable and statistically unreliable estimates; nonetheless, this highlights that each campaign represents a distinct hydrological state, where natural variability in discharge, stage, turbulence, sediment supply, and bedform roughness altered ADCP signatures in ways not fully captured by the chosen predictors. The low transferability across campaigns should therefore be interpreted not as model failure but as an expression of both aleatory

https://doi.org/10.5194/egusphere-2025-6190
Preprint. Discussion started: 5 March 2026

variability (true temporal changes in river hydraulics) and epistemic uncertainty (limited campaign count and missing covariates). Second, suspended sediment concentration (SSC) and bedload transport were not directly measured; instead, Bottom-Track Velocity (BT_Vel) was used as a proxy, which introduces representational uncertainty. Third, potential sensor-specific biases and variations in deployment geometry (e.g., changes in Bin Distance) were not corrected through independent calibration. Fourth, DL models were trained on relatively small datasets, constraining their generalization capacity. Fifth, external hydro-meteorological drivers such as rainfall, upstream sediment supply, and antecedent flow history were excluded, despite their influence on sediment transport. In addition, exploratory refinements (e.g., a deeper ANN and an optimized LSTM) improved their absolute performance and ranking relative to the baselines, as expected. However, to maintain fairness and a like-for-like comparison across the full set of ML/DL models, only standardized baseline configurations are reported. Similarly, Bayesian hyperparameter searches produced marginal gains without changing the broader performance patterns and were therefore excluded from the reported benchmarking.

4.6 Future Outlook

Future work should aim to expand the dataset, both temporally and spatially, through increased field campaigns and incorporation of more diverse hydraulic conditions. Direct measurements of SSC and bedload transport would enhance model calibration and validation. Hybrid modeling approaches that integrate physical modeling (e.g., Delft3D or CFD) with data-driven methods may improve both accuracy and interpretability. Real-time deployment of ADCPs combined with ML or data assimilation may support more continuous monitoring of sediment transport dynamics, though practical implementation will require addressing data volume, noise, and model updating. Furthermore, future efforts could explore robust data augmentation techniques, domain adaptation, and multi-target modeling to enhance the applicability of these models across different riverine systems.

Overall, this study demonstrates the potential of ADCP-derived features in proxy modeling of sediment dynamics, while highlighting the importance of validation strategy, model selection, and environmental context. For real-world applications, efforts should prioritize increasing training data volume, applying regularization or data augmentation for DL models, and exploring hybrid modeling frameworks that balance robustness and expressiveness. These strategies are likely to be important for enabling data-driven sediment transport tools at operational scales in morphodynamically active riverine systems and furthering the integration of hydroacoustic sensing with modern predictive analytics.

Conclusions

This study provided a systematic evaluation of Acoustic Doppler Current Profiler (ADCP)–derived features as proxies for sediment transport dynamics in both controlled flume experiments and natural riverine conditions. Using Bottom-Track

https://doi.org/10.5194/egusphere-2025-6190
Preprint. Discussion started: 5 March 2026

Velocity (BT_Vel) as the primary target, we benchmarked a suite of machine learning (ML) and deep learning (DL) models under standardized baseline configurations to enable direct comparison across approaches.

Tree-based ensembles emerged as the most reliable predictors across environments. Random Forest consistently achieved the highest R^2 values and lowest MSE in both datasets, while Gradient Boosting, LightGBM, CatBoost, and XGBoost also delivered robust performance with only modest variance between split and cross-validation. These results highlight the strength of ensemble methods in balancing accuracy and stability, even under heterogeneous field conditions.

DL baselines demonstrated competitive performance, particularly the LSTM and the LSTM+CNN hybrid, which performed comparably to ensemble trees in the flume and showed resilience under field cross-validation. The field dataset revealed a unique pattern: while most DL models experienced modest performance declines under CV, LSTM improved relative to its split performance, suggesting that recurrent memory structures may benefit from fold-level diversity in noisy, sparse environments. By contrast, simpler networks such as ANN showed limited generalization, and the Stacking Regressor consistently underperformed, reflecting its vulnerability to overfitting without stronger regularization.

SHAP-based interpretability reinforced the physical relevance of the selected features. In the laboratory, Mean_Speed dominated, consistent with flow-controlled hydraulic conditions. In the field, Depth and SNR emerged as primary drivers, underscoring the influence of stage variability and acoustic return quality under natural conditions. These complementary findings highlight the importance of integrating both hydrodynamic and acoustic indicators for reliable proxy modeling.

Overall, this work demonstrates that ADCP-derived features, when paired with robust data-driven models, hold strong potential for operational sediment monitoring. To advance toward deployment, future research should prioritize enlarging field datasets, incorporating direct sediment transport measurements for calibration, and extending baseline architectures through regularization, sequence-aware inputs, and hybrid physics–ML frameworks. Such developments will be essential for achieving resilient, transferable models capable of supporting sediment transport forecasting in dynamic riverine systems.

Data and code availability

All datasets used in this study (laboratory and field datasets) and all analysis scripts (preprocessing, ML/DL, cross-validation, and SHAP analysis) are permanently archived and openly available on Zenodo: https://doi.org/10.5281/zenodo.18716833 (Tuhin, 2026). For a quick interactive view of the full repository, it can also be opened directly in VS Code for the Web (https://vscode.dev/github/thtuhin488/ADCP_BT_Vel_ML_DL) or GitHub.dev (https://github.dev/thtuhin488/ADCP_BT_Vel_ML_DL).

https://doi.org/10.5194/egusphere-2025-6190
Preprint. Discussion started: 5 March 2026

Author contributions

M. T. H. Tuhin designed and carried out the study, including conceptualization, methodology, data curation, formal analysis, visualization, development, and writing – original draft. R. Hinkelmann provided supervision and contributed to writing – review and editing. C. Mudersbach provided supervision, resources, and funding acquisition, and contributed to writing – review and editing.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgments

The authors thank colleagues at Bochum University of Applied Sciences for their support during field and laboratory campaigns, as well as the Technische Universität Berlin for academic guidance. The authors also gratefully acknowledge using ChatGPT (OpenAI) for language polishing and take full responsibility for the scientific content. A preliminary version of this work was presented as a poster at the EGU General Assembly 2025, Vienna, Austria.

Financial support

This research was funded by the Deutsche Forschungsgemeinschaft (DFG) under the program Forschungsimpulse (Project KOENIG – A calibrated and virtual multi-scale wave measurement laboratory to improve and supplement flow measurements: From basic research to industrial application), grant identifier FIP 61/1 – 2024, project number 528740160.

References

Achite, M., Katipoğlu, O. M., Elshaboury, N., Tuğrul, T., and Pandey, K.: Intercomparison of sediment transport curve and novel deep learning techniques in simulating sediment transport in the Wadi Mina Basin, Algeria, Environ. Earth Sci., 84, 12051, https://doi.org/10.1007/s12665-024-12051-w, 2025.

Ali, M., Sterk, G., Seeger, M., Boersema, M., and Peters, P.: Effect of hydraulic parameters on sediment transport capacity in overland flow over erodible beds, Hydrol. Earth Syst. Sci., 16, 591–601, https://doi.org/10.5194/hess-16-591-2012, 2012.

Allawi, M. F., Sulaiman, S. O., Sayl, K. N., Sherif, M., and El-Shafie, A.: Suspended sediment load prediction modelling based on artificial intelligence methods: The tropical region as a case study, Heliyon, 9, e18506, https://doi.org/10.1016/j.heliyon.2023.e18506, 2023.

https://doi.org/10.5194/egusphere-2025-6190
Preprint. Discussion started: 5 March 2026

Chen, T. and Guestrin, C.: XGBoost: a scalable tree boosting system, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16), San Francisco, USA, 13–17 August 2016, ACM, New York, NY, USA, 785–794, https://doi.org/10.1145/2939672.2939785, 2016.

Chen, X., Hassan, M. A., and Fu, X.: Convolutional neural networks for image-based sediment detection applied to a large terrestrial and airborne dataset, Earth Surf. Dynam., 10, 349–366, https://doi.org/10.5194/esurf-10-349-2022, 2022.

Conevski, S., Aleixo, R., Guerrero, M., and Ruther, N.: Bedload velocity and backscattering strength from mobile sediment bed: A laboratory investigation comparing bistatic versus monostatic acoustic configuration, Water, 12, 3318, https://doi.org/10.3390/w12123318, 2020.

Conevski, S., Guerrero, M., Rennie, C. D., and Ruther, N.: Towards an evaluation of bedload transport characteristics by using Doppler and backscatter outputs from ADCPs, J. Hydraul. Res., 59, 703–723, https://doi.org/10.1080/00221686.2020.1818311, 2021.

Conevski, S., Guerrero, M., Winterscheid, A., Faltis, D., Rennie, C. D., and Ruther, N.: Analysis of the riverbed backscattered signal registered by ADCPs in different bedload transport conditions – field application, J. Hydraul. Res., 61, 532–551, https://doi.org/10.1080/00221686.2023.2224276, 2023.

Conevski, S., Guerrero, M., Winterscheid, A., Rennie, C., and Ruther, N.: Neural networks for predicting bedload transport rates by using stationary ADCP measurements, in: River Flow 2024: Proceedings of the 12th International Conference on Fluvial Hydraulics, Liverpool, UK, 2–6 September 2024, CRC Press, 768–776, https://doi.org/10.1201/9781003475378-112, 2025.

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., and Lawler, J. J.: Random forests for classification in ecology, Ecology, 88, 2783–2792, https://doi.org/10.1890/07-0539.1, 2007.

Emmett, W. W.: A field calibration of the sediment-trapping characteristics of the Helley–Smith bed-load sampler, Washington, D.C., U.S. Geological Survey Professional Paper 1139, 44 pp., 1980.

Friedman, J. H.: Greedy Function Approximation: A Gradient Boosting Machine, The Annals of Statistics, 29, 1189–1232, 2001.

Frings, R. M. and Vollmer, S.: Guidelines for sampling bedload transport with minimum uncertainty, Sedimentology, 64, 1630–1645, https://doi.org/10.1111/sed.12366, 2017.

Gacu, J. G., Monjardin, C. E. F., Mangulabnan, R. G. T., Pugat, G. C. E., and Solmerin, J. G.: Artificial intelligence in surface water management: a comprehensive review of methods, applications, and challenges, Water, 17, 1707, https://doi.org/10.3390/w17111707, 2025.

Gaeuman, D. and Jacobson, R. B.: Acoustic bed velocity and bed load dynamics in a large sand bed river, J. Geophys. Res., 111, F02005, https://doi.org/10.1029/2005JF000411, 2006.

Garcia, M. H. (Ed.): Sedimentation engineering: processes, measurements, modeling, and practice, ASCE Manuals and Reports on Engineering Practice, No. 110, American Society of Civil Engineers, Reston, VA, ISBN 978-0-7844-0814-8, 2008.

https://doi.org/10.5194/egusphere-2025-6190
Preprint. Discussion started: 5 March 2026

Gartner, J. W.: Estimating suspended solids concentrations from backscatter intensity measured by acoustic Doppler current profiler in San Francisco Bay, California, Mar. Geol., 211, 169–187, https://doi.org/10.1016/j.margeo.2004.07.001, 2004.

Guerrero, M., Rüther, N., Szupiany, R., Haun, S., Baranya, S., and Latosinski, F.: The acoustic properties of suspended sediment in large rivers: Consequences on ADCP methods applicability, Water, 8, 13, https://doi.org/10.3390/w8010013, 2016.

Hubbell, D. W.: Apparatus and techniques for measuring bedload, U.S. Geological Survey, Washington, D.C., U.S. Geological Survey Water-Supply Paper, 1748, 74 pp., 1964.

Jing, T., Zeng, Y., Fang, N., Dai, W., and Shi, Z.: A review of suspended sediment hysteresis, Water Resour. Res., 61, e2024WR037216, https://doi.org/10.1029/2024WR037216, 2025.

Kakarla, S. M., Abuomar, O. "., Abojaradeh, M., Shaqadan, A., Alshoulat, I., and Omari, A.: Acoustic Doppler current profiler data analytics using a hybrid of machine learning methods, in: 2nd International Engineering Conference on Electrical, Energy, and Artificial Intelligence (EICEEAI), IEEE, Zarqa, Jordan, 2023.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.-Y.: LightGBM: A highly efficient gradient boosting decision tree, in: Advances in Neural Information Processing Systems 30 (NeurIPS 2017), Long Beach, USA, 4–9 December 2017, Curran Associates, Inc., Red Hook, NY, USA, 3149–3157, 2017.

Kostaschuk, R., Best, J., Villard, P., Peakall, J., and Franklin, M.: Measuring flow velocity and sediment transport with an acoustic Doppler current profiler, Geomorphology, 68, 25–37, https://doi.org/10.1016/j.geomorph.2004.07.012, 2005.

Latosinski, F. G., Szupiany, R. N., Guerrero, M., Amsler, M. L., and Vionnet, C.: The ADCP's bottom track capability for bedload prediction: Evidence on method reliability from sandy river applications, Flow Meas. Instrum., 54, 124–135, https://doi.org/10.1016/j.flowmeasinst.2017.01.005, 2017.

Le Guern, J., Rodrigues, S., Geay, T., Zanker, S., Hauet, A., Tassi, P., Claude, N., Jugé, P., Duperray, A., and Vervynck, L.: Relevance of acoustic methods to quantify bedload transport and bedform dynamics in a large sandy-gravel-bed river, Earth Surf. Dynam., 9, 423–444, https://doi.org/10.5194/esurf-9-423-2021, 2021.

Lund, J. W., Groten, J. T., Karwan, D. L., and Babcock, C.: Using machine learning to improve predictions and provide insight into fluvial sediment transport, Hydrol. Process., 36, e14648, https://doi.org/10.1002/hyp.14648, 2022.

Mir, A. A., Patel, M., Albalawi, F., Bajaj, M., and Tuka, M. B.: A comparative ensemble approach to bedload prediction using metaheuristic machine learning, Sci. Rep., 14, 25725, https://doi.org/10.1038/s41598-024-75118-5, 2024.

Ozsahin, D. U., Emegano, D. I., Uzun, B., and Ozsahin, I.: Machine learning approaches for river discharge prediction using acoustic Doppler current profiler (ADCP) data, in: Climate change and water resources in Mediterranean countries, Gökçekuş, H., Kassem, Y. (Eds.), Springer Nature Switzerland, Cham, 175–188, 2025.

Pham Van, C., Le, H., and van Chin: Estimation of daily suspended sediment concentration in the Ca River Basin using a sediment rating curve, multiple regression, and long short-term memory model, J. Water Clim. Change, 14, 4356–4375, https://doi.org/10.2166/wcc.2023.229, 2023.

https://doi.org/10.5194/egusphere-2025-6190
Preprint. Discussion started: 5 March 2026

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., and Gulin, A.: CatBoost: Unbiased boosting with categorical features, in: Advances in Neural Information Processing Systems 31 (NeurIPS 2018), Montréal, Canada, 3–8 December 2018, Curran Associates Inc., Red Hook, NY, USA, 6638–6648, 2018.

Ramooz, R. and Rennie, C. D.: Laboratory measurement of bedload with an ADCP, U.S. Geological Survey, U.S. Geological Survey Scientific Investigations Report, 2010–5091, 20 pp., 2010.

Rennie, C. D., Millar, R. G., and Church, M. A.: Measurement of bed load velocity using an acoustic Doppler current profiler, J. Hydraul. Eng., 128, 473–483, https://doi.org/10.1061/(ASCE)0733-9429(2002)128:5(473), 2002.

Rennie, C. D. and Rainville, F.: Case study of precision of GPS differential correction strategies: Influence on ADCP velocity and discharge estimates, J. Hydraul. Eng., 132, 225–234, https://doi.org/10.1061/(ASCE)0733-9429(2006)132:3(225), 2006.

Rennie, C. D. and Villard, P. V.: Site specificity of bed load measurement using an acoustic Doppler current profiler, J. Geophys. Res., 109, F03003, https://doi.org/10.1029/2003JF000106, 2004.

Simpson, M. R.: Discharge measurements using a broad-band acoustic Doppler current profiler, U.S. Geological Survey, Water-Resources Investigations Report 01-4044, 123 pp., 2001.

Tabesh, M., Lorenz, F., Reich, J., Artz, T., and Winterscheid, A.: Detailed workflow for estimating bedload transport based on dune tracking method, Int. J. River Basin Manag., 1–21, https://doi.org/10.1080/15715124.2024.2412194, 2024.

Thorne, P., Vincent, C., Hardcastle, P., Rehman, S., and Pearson, N.: Measuring suspended sediment concentrations using acoustic backscatter devices, Mar. Geol., 98, 7–16, https://doi.org/10.1016/0025-3227(91)90031-X, 1991.

Tuhin, M. T. H. and Mudersbach, C.: Integrating machine learning with ADCP data for advanced sediment transport and hydrodynamics monitoring, in: EGU General Assembly 2025, Vienna, Austria & online, 27 April–2 May, Copernicus Meetings, Göttingen, Germany. EGU25-8296, https://doi.org/10.5194/egusphere-egu25-8296, 2025.

Tuhin, M. T. H.: ADCP-Based Proxy Sediment Transport Monitoring – Data and Code Archive [Data set], Zenodo, https://doi.org/10.5281/zenodo.18716833, 2026.

U.S. Army Corps of Engineers (USACE): Sediment transport functions, https://www.hec.usace.army.mil/confluence/rasdocs/ras1dtechref/6.0/stable-channel-design-functions/sediment-transport-capacity/sediment-transport-functions, last access: 17 September 2025.