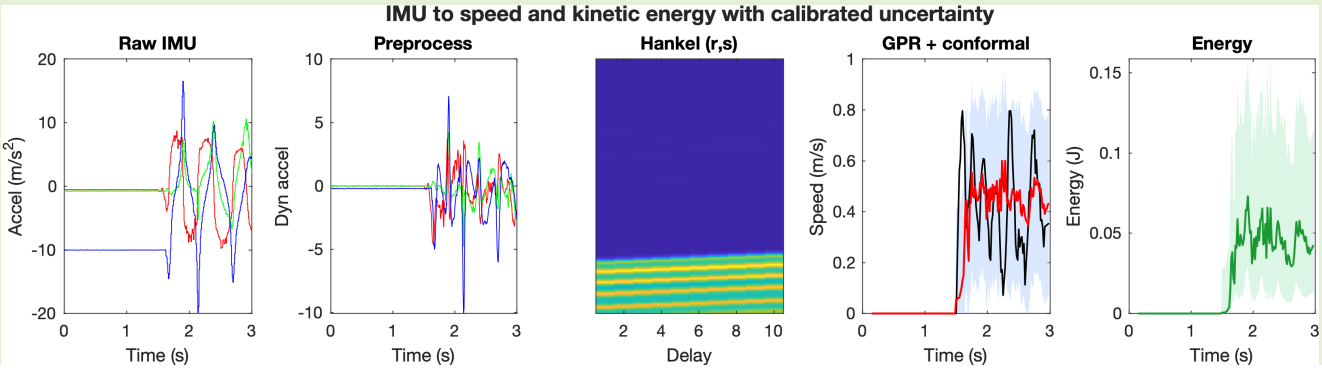
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/JIOT.2025.3641118,

Kinetic Energy Estimation of IMU-Equipped Sediment Particles with Gaussian Process Regression and Conformal Prediction

Georgios Maniatis, Jeffrey Tuhtan, Gert Toming, Edward Curley,
Charlie Gadd, Richard Williams, and Trevor Hoey

***Abstract*— Direct particle-scale sediment measurements remain difficult in turbid, high-energy rivers where optical methods fail. We present an integration-free IMU workflow that maps short windows to projected speed and kinetic energy using physics-aware preprocessing, orientation-invariant Hankel embeddings, Gaussian process regression (GPR), and split conformal prediction. On event-disjoint hold-out tests, the selected GPR model ($m = 10$) achieves $R^2 = 0.628$, $\mathrm{RMSE} = 0.168\,\mathrm{m\,s^{-1}}$, and $\mathrm{MAE} = 0.096\,\mathrm{m\,s^{-1}}$. A four-model benchmark on identical event-grouped folds (GPR, LSTM, SVR-RBF, LSBoost) gives the lowest RMSE for LSBoost ($0.158\,\mathrm{m\,s^{-1}}$); GPR is within $0.001\,\mathrm{m\,s^{-1}}$ of the strongest non-GPR comparator (LSBoost), and paired RMSE differences are non-significant ($p = 0.812$). Empirical conformal coverage is 87.6 %/93.7 %/97.9 % for nominal 90 %/95 %/99 % targets. River Calder deployments show peak kinetic energies up to $0.168\,\mathrm{J}$. The framework provides uncertainty-aware kinematics and energetics for autonomous sediment-transport monitoring.**

***Index Terms*— Gaussian Process Regression, Conformal Prediction, Inertial Measurement Units, Sediment Transport, Uncertainty Quantification, Smart Sensors, Geomorphology**



I. Introduction

SEDIMENT transport controls river morphology, flood hazards, and aquatic habitat, yet direct particle-scale measurement remains difficult in turbid, fast-changing flows in coarse-grained gravel-bed and bedrock rivers. Optical image velocimetry, acoustic Doppler instrumentation, and passive acoustic proxies are widely used but often require site-specific calibration and can degrade during high-flow events or in coarse-bed rivers [1]–[4]. These issues motivate sensing approaches that can operate without clear water, line-of-sight, or fixed infrastructure.

Instrumented particles with embedded inertial measurement units (IMUs) have been used to investigate initiation of motion and transport dynamics, offering an alternative to optical and acoustic sensing because they directly sense motion at the grain scale [5]–[8]. Modern smart pebbles typically log at 100 Hz–1000 Hz to resolve impacts and short-lived transitions [9]. However, IMU streams are noisy, biased, and only indirectly related to the translational kinematics of interest. Naive integration of accelerations drifts rapidly, while orientation changes and impacts introduce non-stationary artifacts; this motivates integration-free estimation discussed later. Reliable inference therefore requires physics-aware preprocessing, robust feature construction, and models that generalize across different motion regimes and deployment contexts [5]–[8].

Gaussian Process Regression (GPR) provides data-efficient nonlinear mappings with probabilistic outputs, making it attractive for short-window kinematic estimation. Yet GP posterior variance alone does not guarantee calibrated uncertainty, especially under distribution shift (changes between training and deployment data distributions). Distribution-free conformal prediction provides finite-sample coverage guarantees for

School of Applied Sciences, University of Brighton, United Kingdom
Tallinn University of Technology, Estonia
University of Glasgow, United Kingdom
Royal Holloway, University of London, United Kingdom
Earth Sciences New Zealand (ESNZ), New Zealand
Brunel University of London, United Kingdom

black-box regressors and offers an uncertainty layer suitable for decision-making in geomorphology [10]–[12].

This paper presents an integration-free pipeline that maps short IMU windows to one-dimensional projected speed and derived kinetic energy. The workflow combines static noise characterization to set gyro and accelerometer thresholds, a complementary filter with gated tilt correction, and orientation-invariant Hankel embeddings of dynamic acceleration and gyro norm. All splits are event-disjoint for hold-out testing, cross-validation kernel comparison, and conformal calibration. The selected GPR model is paired with split conformal prediction to yield calibrated speed intervals and propagated kinetic-energy bands, and the method is demonstrated on field episodes from the River Calder. These outputs support entrainment thresholds, transport stage classification, and energy budgeting where direct observation is impractical.

II. Related Work

Smart-pebble and instrumented-grain research has progressed from early feasibility demonstrations to more robust sensing platforms. Initial work validated embedded IMUs in flume settings and highlighted persistent challenges in waterproofing, bias/drift, and sensor–shape alignment [5], [6]. Subsequent studies refined calibration and used in-grain accelerometers to quantify hydrodynamic forces and motion transitions, with sampling rates increasing to resolve impacts and short-lived motion events [7]–[9].

Recent advances have pushed instrumented particles toward process interpretation and kinematic equivalence. Experiments have characterized vibration signatures near entrainment and their dependence on flow state [13], [14], introduced calibrated spherical inertial particles that preserve density and inertia [15], and applied IMU-equipped clasts to distinguish motion modes and energetic regimes [16], [17]. These efforts underline the need for short-window models that cope with impacts between the instrumented particle and the bed or neighboring particles, and with transitions between rolling, sliding, and saltation.

Field deployments have increasingly combined instrumented particles with external proxies to improve interpretability. Joint smart-particle and seismic sensing provides cross-validation during rapidly changing stages, while passive acoustic calibration studies demonstrate benefits for phase identification and quality control [18]–[20]. Such cross-modal strategies motivate uncertainty-aware inference pipelines that remain reliable under distribution shifts between laboratory training data and field deployments.

The literature on wearable and underwater IMUs emphasizes calibration protocols, bias compensation, and reliability assessment. These methodological lessons are directly transferable to fluvial sensing where waterproofing, alignment, and short-window inference are critical [21]–[24].

Delay embeddings provide a principled way to extract dynamics from short time series. Hankel structures connect to Koopman and Dynamic Mode Decomposition (DMD) viewpoints, where observable evolution is approximated in a linear subspace [25]. Applied studies across engineering show that delay embeddings can capture transient dynamics while controlling overfitting [10], [25]–[28].

Finally, coarse-grained fluvial transport is highly intermittent, and calibrated uncertainty is essential for interpreting entrainment and cessation thresholds, stage-dependent transport, and energy budgets [12]. Conformal prediction supplies distribution-free intervals with finite-sample guarantees and has been extended to time-series and dependent settings [11], [29]–[32]. Our work combines these ideas with a physics-aware IMU pipeline to provide risk-aware sediment-transport estimates.

III. Methodology

A. System overview and data acquisition

We use the *Kivi* smart pebble in its ellipsoidal configuration [9] to record tri-axial accelerometer and gyroscope data during laboratory and field transport. The particle has external dimensions $80 \times 63 \times 56$ mm (semi-axes $a = 0.0400\,\mathrm{m}$, $b = 0.0315\,\mathrm{m}$, $c = 0.0280\,\mathrm{m}$) with mass $M = 0.330\,\mathrm{kg}$. On-board electronics log inertial data and housekeeping variables to removable storage; the system supports $100\,\mathrm{Hz}$ IMU sampling, $1000\,\mathrm{Hz}$ high-g acceleration logging, and an $8\,\mathrm{GB}$ card with a battery rated for multi-hour deployment. A full hardware specification, engineering drawings and photographs are provided in the Kivi reference paper [9]; Fig. 1 summarizes the instrument layout and test geometries used throughout this study.

Desk experiments. To generate short bouts of unconstrained motion under controlled visibility, the ellipsoidal sensor was briefly flicked by hand on a flat, rigid surface while recording video at $60\,\mathrm{Hz}$. These dry sequences provide diverse initiation modes and onset transients and were used to validate time-stamping and to tune the static-window detector.

Laboratory flume. Wet experiments were conducted in a straight recirculating flume with bed slope $S = 0.0010$ and a $0.40\,\mathrm{m}$ wide test section. Flow rate was increased from $17\,\mathrm{L\,s^{-1}}$ to $20\,\mathrm{L\,s^{-1}}$ until the instrumented pebble was entrained (i.e., initiated sustained motion from rest), corresponding to $Re \approx (4.3\text{–}5.0) \times 10^4$ (using $B = 0.40\,\mathrm{m}$ and $\nu \approx 10^{-6}\,\mathrm{m^2\,s^{-1}}$). Depth was uniform along the $2\,\mathrm{m}$ recording section (four-point checks); pressure sensors on the pebble showed short-term fluctuations of $2.1\,\mathrm{mbar}$ ($2.1\,\mathrm{cm}$ water head, standard deviation) and a median inter-sensor head difference of $7.4\,\mathrm{mbar}$ ($7.5\,\mathrm{cm}$). For each run the pebble was placed on the prepared bed, allowed to interact with the flow, and—after motion ceased—left quiescent for at least $10\,\mathrm{s}$ to capture baseline segments. Selected runs were videoed at $60\,\mathrm{Hz}$ from the side for independent tracking.

Field deployment (River Calder, Scotland). The same ellipsoidal pebble was placed on a bedrock–gravel reach of the River Calder in western Scotland and left to be transported by ambient flow during typical seasonal stages [33]. Pressure sensors indicate median submergence of $0.16\,\mathrm{m}$ (5th–95th percentile $0.06\,\mathrm{m}$–$0.22\,\mathrm{m}$) when referenced to the lowest $1\,\%$ pressures across field events, and short-term fluctuations of $8.6\,\mathrm{mbar}$ ($8.8\,\mathrm{cm}$ water head, standard deviation). A DEM-based, Manning-derived median reach discharge is approximated at $5830\,\mathrm{L\,s^{-1}}$, which is consistent with records of

mean flow from the nearest flow station (*Calder at Muirshiel*, $6580\,\mathrm{L\,s^{-1}}$). After each transport episode the sensor was left static for at least $10\,\mathrm{s}$.

Training dataset and targets. The supervised dataset comprises 88 labelled $3\,\mathrm{s}$ events (60 desk flicks and 28 wet flume runs), with 56 events used for training, 14 for conformal calibration, and 18 for the held-out test. Video was analysed by manual digital tracking; the projected centre-of-mass speed along the dominant horizontal direction provides a one-dimensional reference projected speed at $60\,\mathrm{Hz}$ (cross-stream motion reduces the projection and thus underestimates the true 2-D speed). 5 short episodes ($12.5\,\mathrm{s}$–$23.0\,\mathrm{s}$) from the River Calder (no video) serve to demonstrate the estimator *in situ* together with uncertainty quantification and kinetic-energy reconstruction.

Split protocol (event-disjoint). Each event is a $3\,\mathrm{s}$ window resampled to $L = 180$ points; Hankel embedding of delay m yields $(L - m + 1)$ samples per event. All validation splits operate by event: a hold-out is drawn on unique event IDs; split conformal calibration divides the training events into disjoint model/calibration subsets; and cross-validation partitions event IDs into folds before expanding to Hankel rows. Kernel and delay selection are performed only on the development set (train+calibration); held-out test events are used once for final evaluation. Consequently, every Hankel row derived from a given event remains in exactly one split, eliminating leakage across train/calibration/test partitions. Figure 2 provides a compact workflow diagram across desk, flume, and field data; Figure 3 shows an example event and signal flow.

B. Signal preprocessing

Static noise characterization.: A static IMU recording provides noise scales for thresholding. Let $\boldsymbol{a}_t$ and $\boldsymbol{\omega}_t$ denote raw accelerometer and gyroscope samples. The per-axis noise scales are

$$\sigma_a = \tfrac{1}{3}\sum_{i=1}^{3}\mathrm{std}(a_i), \qquad \sigma_\omega = \tfrac{1}{3}\sum_{i=1}^{3}\mathrm{std}(\omega_i). \tag{1}$$

The three sensor axes are orthogonal in the IMU frame. Gyroscope units are detected ($^\circ\,\mathrm{s^{-1}}$ or $\mathrm{rad\,s^{-1}}$) and normalized to $^\circ\,\mathrm{s^{-1}}$ for thresholding. The gyro deadband is $\theta_{\omega,\mathrm{dead}} = 3\sigma_\omega$, the accelerometer tilt gate is $\theta_a = 3\sigma_a$, and the static-window detector uses the distinct threshold $\theta_{\mathrm{stat}} = 1.200\,^\circ\,\mathrm{s^{-1}}$. For reference, the dataset-level estimates are $\sigma_a = 0.0262\,\mathrm{m\,s^{-2}}$ and $\sigma_\omega = 0.0714\,^\circ\,\mathrm{s^{-1}}$, and the corresponding noise histograms are provided in Appendix Fig. 7. Gyroscope rates and σ_ω are converted to $\mathrm{rad\,s^{-1}}$ for kinetic-energy computation. The static threshold is set conservatively (tuned on desk trials) to avoid false motion detection and is intentionally higher than $3\sigma_\omega$; in this dataset, $\theta_{\mathrm{stat}} \approx 17\sigma_\omega$.

Per-episode bias and gravity reset.: Within each event, a static segment is identified where the gyro norm falls below θ_{stat} for at least $0.5\,\mathrm{s}$; if no such segment exists, the lowest $5\,\%$ of gyro magnitudes are used. The gyro bias is estimated from this segment and subtracted before filtering. The gravity magnitude g_0 is taken as the mean accelerometer norm over the same segment and is used for tilt gating and gravity compensation. This ensures that bias and gravity references are consistent within each episode while retaining global noise-derived thresholds.

Filtering and gravity compensation.: Gyroscope measurements are de-biased using the static mean and deadbanded when $\|\boldsymbol{\omega}_t\| < \theta_{\omega,\mathrm{dead}}$. Accelerometer and gyroscope signals are Savitzky–Golay smoothed component-wise with a cubic polynomial (order 3) and a 21-sample window [34]. A complementary filter propagates the quaternion using the gyro rate and applies an accelerometer tilt correction only when the acceleration magnitude is close to g_0 (note that gravity is *not* removed before this gate):

$$\hat{q}_t^- = q_{t-1} \otimes \exp\left(\tfrac{1}{2}\Delta t\,\hat{\omega}_t\right), \tag{2}$$

$$\hat{g}_t = R(\hat{q}_t^-)^\top e_z, \tag{3}$$

$$e_t = \hat{g}_t \times \tfrac{\hat{a}_t}{\|\hat{a}_t\|}, \tag{4}$$

$$q_t = \exp\left(\tfrac{1-\alpha}{2}e_t\right) \otimes \hat{q}_t^-, \tag{5}$$

$$\text{if } \big|\,\|\hat{a}_t\| - g_0\big| < \theta_a, \tag{6}$$

$$q_t = \hat{q}_t^-, \quad \text{otherwise.} \tag{7}$$

Here e_z is the unit vector aligned with gravity. The gravity vector is $g(t) = g_0 R(q_t)^\top e_z$ and dynamic acceleration is $\boldsymbol{a}_{\mathrm{dyn}}(t) = \hat{\boldsymbol{a}}_t - g(t)$. Orientation-invariant magnitudes (invariant to sensor orientation but responsive to motion dynamics) are then

$$r(t) = \|\boldsymbol{a}_{\mathrm{dyn}}(t)\|\,, \qquad s(t) = \|\hat{\boldsymbol{\omega}}_t\|\,. \tag{8}$$

Windowing and resampling.: Each supervised event is a $3\,\mathrm{s}$ window aligned to motion initiation and resampled to $L = 180$ samples at $60\,\mathrm{Hz}$; the video reference $v^{\mathrm{ref}}(t)$ is interpolated onto the same grid, and a normalized time coordinate $t^* \in [0, 1]$ is retained for alignment. Field episodes are processed in sliding windows with a stride of one sample ($0.0167\,\mathrm{s}$) for Fig. 6 to produce continuous traces.

C. Hankel time-delay embedding

To encode short-window dynamics without long sequences, we construct a Hankel delay embedding of the orientation-invariant scalars. For delay dimension m, the jth row stacks the last m samples of r and s:

$$\begin{aligned}&\boldsymbol{x}_j = [r_j, \ldots, r_{j+m-1},\ s_j, \ldots, s_{j+m-1}]^\top,\\ &\boldsymbol{x}_j \in \mathbb{R}^{2m}.\end{aligned} \tag{9}$$

$$y_j = v^{\mathrm{ref}}_{j+m-1}. \tag{10}$$

The target is aligned to the end of the window so that each feature vector predicts the current projected speed. We evaluate $m \in \{8, 10, 12, 15\}$ and select the best via event-grouped cross-validation. At $60\,\mathrm{Hz}$, these delays span roughly $0.13\,\mathrm{s}$–$0.25\,\mathrm{s}$ of recent history, providing short-term dynamics without long sequences.

Static gating.: Static detection is applied per Hankel row: if $\max_{k=0\ldots m-1} s_{j+k} < \theta_{\mathrm{stat}}$, then $\boldsymbol{x}_j = 0$ and the prediction is clamped to $\hat{y}_j = 0$. The same rule is enforced in training, calibration, test, and field inference to preserve consistency.

D. Gaussian process regression and kernel family

From the windowed samples $\{(\boldsymbol{x}_j, y_j)\}_{j=1}^{n}$, projected speed is modeled as

$$f(\boldsymbol{x}) \sim \mathcal{GP}\left(0,\, k(\boldsymbol{x}, \boldsymbol{x}')\right), \tag{11}$$

$$y = f(\boldsymbol{x}) + \varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, \sigma_n^2). \tag{12}$$

For a test input $\boldsymbol{x}_*$, the posterior mean and variance are

$$\mu_* = \boldsymbol{k}_*^\top (\boldsymbol{K} + \sigma_n^2 \boldsymbol{I})^{-1} \boldsymbol{y}, \tag{13}$$

$$\sigma_*^2 = k(\boldsymbol{x}_*, \boldsymbol{x}_*) - \boldsymbol{k}_*^\top (\boldsymbol{K} + \sigma_n^2 \boldsymbol{I})^{-1} \boldsymbol{k}_*. \tag{14}$$

Here $\boldsymbol{K}$ is the $n \times n$ kernel matrix with $K_{ij} = k(\boldsymbol{x}_i, \boldsymbol{x}_j)$, and $\boldsymbol{k}_*$ collects covariances between $\boldsymbol{x}_*$ and the training inputs.

Kernel comparison.: We evaluate squared-exponential and Matérn kernels alongside the MATLAB default GPR baseline:

$$k_{\mathrm{SE}}(\boldsymbol{x}, \boldsymbol{x}') = \sigma_f^2 \exp\!\left(-\frac{\|\boldsymbol{x} - \boldsymbol{x}'\|^2}{2\ell^2}\right), \tag{15}$$

$$k_{\mathrm{M32}}(\boldsymbol{x}, \boldsymbol{x}') = \sigma_f^2 \left(1 + \tfrac{\sqrt{3}}{\ell} r\right) \exp\!\left(-\tfrac{\sqrt{3}}{\ell} r\right), \tag{16}$$

$$k_{\mathrm{M52}}(\boldsymbol{x}, \boldsymbol{x}') = \sigma_f^2 \left(1 + \tfrac{\sqrt{5}}{\ell} r + \tfrac{5}{3}\tfrac{r^2}{\ell^2}\right) \exp\!\left(-\tfrac{\sqrt{5}}{\ell} r\right), \tag{17}$$

with $r = \|\boldsymbol{x} - \boldsymbol{x}'\|$. Event-grouped five-fold cross-validation evaluates RMSE, MAE, and the coefficient of determination. Embedding sizes are scanned over $m \in \{8, 10, 12, 15\}$. Model selection minimizes mean CV RMSE with tie-breakers on CV(RMSE), MAE, and compute time. Nonnegativity is enforced by clamping predictions to zero after static gating. Comparator benchmarking (GPR, LSTM, SVR-RBF, LSBoost) uses identical event-grouped outer folds, identical preprocessing/targets/metrics, and model-specific hyperparameter tuning in a nested workflow.

E. Split conformal prediction for coverage

Split conformal prediction is applied to the GPR point predictions using a calibration set disjoint from training. Nonconformity scores are

$$\alpha_j = \left|y_j - \hat{y}_j\right|. \tag{18}$$

For miscoverage ϵ, the finite-sample quantile is

$$q = \alpha_{(k)}, \quad k = \lceil (n+1)(1-\epsilon) \rceil, \tag{19}$$

with $\alpha_{(k)}$ denoting the kth order statistic. Prediction intervals are $[\max(0, \hat{y} - q), \hat{y} + q]$, and static windows map to $[0, 0]$.

F. Kinetic-energy reconstruction

Projected speed $v(t)$ and angular-rate magnitude $\omega(t)$ ($\mathrm{rad\,s^{-1}}$) yield kinetic energy

$$E_{\mathrm{tot}}(t) = \tfrac{1}{2} M v(t)^2 + \tfrac{1}{2} I_{\mathrm{eq}} \omega(t)^2. \tag{20}$$

For the ellipsoidal Kivi, principal moments are

$$\begin{aligned} I_{xx} &= \tfrac{1}{5} M (b^2 + c^2) = 0.000117 \ \mathrm{kg\,m^2}, \\ I_{yy} &= \tfrac{1}{5} M (a^2 + c^2) = 0.000157 \ \mathrm{kg\,m^2}, \\ I_{zz} &= \tfrac{1}{5} M (a^2 + b^2) = 0.000171 \ \mathrm{kg\,m^2}. \end{aligned} \tag{21}$$

An orientation-averaged effective inertia $I_{\mathrm{eq}} \approx 0.000149 \ \mathrm{kg\,m^2}$ is adopted when only $\|\boldsymbol{\omega}\|$ is available.

TABLE I

COMPARATIVE EXPERIMENT ON IDENTICAL EVENT-GROUPED FOLDS.

Model	RMSE (m/s)	MAE (m/s)	R^2	NSE
GPR	0.159	0.093	0.616	0.616
LSTM	0.161	0.090	0.602	0.602
SVR_RBF	0.163	0.092	0.599	0.599
LSBoost	0.158	0.092	0.620	0.620

Energy intervals from conformal speed bounds.: Let q denote the conformal half-width at the desired level and define $v_L = \max(0, \hat{v} - q)$ and $v_U = \hat{v} + q$. The translational energy bounds are

$$E_{\mathrm{trans}} \in \left[\tfrac{1}{2} M v_L^2,\ \tfrac{1}{2} M v_U^2\right]. \tag{22}$$

We form total-energy bounds by adding the measured rotational term, $E_{\mathrm{rot}} = \frac{1}{2} I_{\mathrm{eq}} \omega^2$, to obtain $E_{\mathrm{tot,low}} = \frac{1}{2} M v_L^2 + E_{\mathrm{rot}}$ and $E_{\mathrm{tot,high}} = \frac{1}{2} M v_U^2 + E_{\mathrm{rot}}$.

IV. RESULTS

A. Comparative benchmark across four models

To address baseline-comparison requirements, we benchmark the selected GPR pipeline against three alternatives on the same event-grouped folds: an LSTM sequence regressor [35], SVR with RBF kernel, and LSBoost trees. Table I reports fold-aggregated RMSE, MAE, R^2, and NSE for all four models. The lowest RMSE is achieved by LSBoost at $0.158\,\mathrm{m\,s^{-1}}$. GPR reports RMSE $= 0.159\,\mathrm{m\,s^{-1}}$; the strongest non-GPR comparator (LSBoost) reports RMSE $= 0.158\,\mathrm{m\,s^{-1}}$, giving an absolute gap of $0.001\,\mathrm{m\,s^{-1}}$. The paired nonparametric fold test for RMSE gives $p = 0.812$.

B. Model performance and kernel comparison

Using the IMU-equipped particle data, we estimate coarse-particle motion in river-relevant settings across desk, flume, and field trials. The optimized regressor uses Matérn 3/2 with delay dimension $m = 10$ and tracks the timing and relative magnitude of video-derived speeds while smoothing some sharp peaks. The event-level hold-out fit attains $R^2 = 0.628$, RMSE $= 0.168\,\mathrm{m\,s^{-1}}$, and MAE $= 0.096\,\mathrm{m\,s^{-1}}$ (Fig. 4a–c). Event-grouped cross-validation yields consistent scores (mean RMSE $= 0.161\,\mathrm{m\,s^{-1}}$, mean $R^2 \approx 0.598$, mean MAE $= 0.094\,\mathrm{m\,s^{-1}}$, and $\mathrm{CV_{RMSE}} = 10.2\,\%$; Fig. 4d). Table II summarizes the kernel–embedding comparison, including the MATLAB default baseline, and the selected model is used for all subsequent figures and field reconstructions.

C. Uncertainty quantification

Split conformal prediction yields distribution-free uncertainty on event-disjoint test events. Observed coverage is 87.6 %/93.7 %/97.9 % for nominal 90 %/95 %/99 % intervals, and mean interval widths are $0.521\,\mathrm{m\,s^{-1}}$, $0.712\,\mathrm{m\,s^{-1}}$, and $1.140\,\mathrm{m\,s^{-1}}$. The corresponding half-widths are $q_{90} = 0.261\,\mathrm{m\,s^{-1}}$, $q_{95} = 0.356\,\mathrm{m\,s^{-1}}$, and $q_{99} = 0.570\,\mathrm{m\,s^{-1}}$, computed using the finite-sample order-statistic rule. Fig. 5 summarizes coverage statistics, interval widths, and nonconformity scores; Fig. 4(b) stitches three held-out *dry* validation

TABLE II

EVENT-GROUPED CROSS-VALIDATION ACROSS KERNELS AND EMBEDDING DIMENSIONS. VALUES ARE MEAN METRICS FOR EACH KERNEL–m PAIR (GENERATED AUTOMATICALLY FROM THE CURRENT RUN).

Kernel	m	RMSE (m/s)	MAE (m/s)	R^2	CV(RMSE) (%)
squaredexponential	8	0.166	0.098	0.585	0.095
matern32	8	0.164	0.097	0.596	0.089
matern52	8	0.164	0.097	0.593	0.092
default	8	0.167	0.099	0.579	0.088
squaredexponential	10	0.165	0.096	0.580	0.099
matern32	**10**	**0.161**	**0.094**	**0.598**	**0.102**
matern52	10	0.162	0.094	0.596	0.114
default	10	0.164	0.095	0.581	0.102
squaredexponential	12	0.167	0.098	0.568	0.107
matern32	12	0.165	0.097	0.580	0.108
matern52	12	0.166	0.098	0.576	0.101
default	12	0.167	0.098	0.567	0.108
squaredexponential	15	0.172	0.102	0.553	0.111
matern32	15	0.169	0.101	0.569	0.115
matern52	15	0.170	0.101	0.564	0.109
default	15	0.171	0.102	0.555	0.110

events to show how the conformal bands track transient motion.

D. Field demonstration and kinetic energy

River Calder deployments were post-processed using the validated model to reconstruct projected speed and total kinetic energy (Fig. 6). Four representative episodes are shown; across 5 field episodes, peak total energy spans $0.118\,\mathrm{J}$ to $0.168\,\mathrm{J}$. The predicted speed time series show initiation of motion, peak transport, and decay phases, with conformal bands widening during energetic intervals and tightening during quiescence. Angular-rate magnitude is plotted with a $\pm 2\sigma$ band using the gyroscope noise scale from static measurements (noise envelope, not a calibrated CI). Shaded envelopes depict a $95\,\%$ energy band derived from the conformal speed bounds. The final column decomposes total energy into translational and rotational components; translational shares at the dominant peak range from $10\,\%$ to $77\,\%$, while rotational complements span $23\,\%$–$90\,\%$ across events.

V. DISCUSSION

A. Conformal prediction and risk-aware kinetic-energy inference

A central contribution of this study is the use of split conformal prediction to attach distribution-free uncertainty to speed-magnitude estimates and, by propagation, to kinetic-energy budgets. The selected GPR configuration (Matérn 3/2, $m = 10$) is chosen using event-grouped cross-validation to balance accuracy and stability. Conformal calibration converts point predictions into finite-sample intervals (Fig. 5); empirical coverage is $87.6\,\%/93.7\,\%/97.9\,\%$ on disjoint test events. Propagating these intervals through $E_{\mathrm{tot}}(t) = \frac{1}{2}Mv(t)^2 + \frac{1}{2}I_{\mathrm{eq}}\omega(t)^2$ yields interpretable energy envelopes (Fig. 6) that support robust entrainment/cessation thresholding, site-to-site comparisons, and operational monitoring for distribution shift. The explicit static gating further reduces false motion during quiescent periods while preserving calibrated coverage. For practitioners, half-widths of $0.261\,\mathrm{m\,s^{-1}}$–$0.570\,\mathrm{m\,s^{-1}}$ imply that transport/no-transport decisions should use thresholds above this uncertainty band, and the resulting energy envelopes provide conservative bounds on peak budgets.

The comparative benchmark indicates statistically comparable performance under the current fold count: LSBoost has the lowest RMSE ($0.158\,\mathrm{m\,s^{-1}}$), while GPR remains within $0.001\,\mathrm{m\,s^{-1}}$ of the strongest non-GPR comparator (LSBoost), and the paired-fold RMSE test gives $p = 0.812$. We therefore treat rank ordering as indicative and select GPR as the best compromise for this pipeline: it remains competitive across RMSE/MAE/R^2/NSE, is sample-efficient for limited labeled fluvial events through kernel regularization, and integrates directly with the uncertainty workflow used here (stable short-window predictions plus split-conformal calibration).

Relative to prior smart-particle and IMU-based sediment studies, the main advance here is not simply another feasibility demonstration. Earlier work established instrumented-grain sensing as viable, improved calibration and packaging, and used inertial signatures to interpret entrainment, vibration regimes, and motion-state transitions [5]–[8], [13]–[17]. Those studies substantially improved process interpretation, but they generally did not target direct projected-speed reconstruction with event-disjoint evaluation, explicit cross-model benchmarking on identical folds, or calibrated uncertainty propagation into kinetic-energy estimates. The contribution of the present workflow is therefore methodological as well as applicational: it converts short IMU windows into transport-relevant kinematic and energetic quantities while retaining an uncertainty layer that is interpretable for field deployment.

Empirical coverage is slightly below nominal for the lower target levels, which is not unexpected in this setting because Hankel rows are temporally correlated and lab-to-field transfer introduces mild distribution shift. We therefore present empirical rates alongside nominal targets and interpret the resulting bands as operational risk bounds rather than strict exchangeability guarantees.

B. Why integration-free learning matters

Inferring velocity by integrating accelerations is notoriously sensitive to small biases, misalignment, and gravity leakage; even milligravity errors accumulate quickly, restricting credible operation to very short windows [36]. The integration-free approach used here instead learns a mapping from orientation-invariant features to projected speed, leveraging both dynamic acceleration and angular-rate norms. Hankel embeddings retain short-term temporal structure without long sequences, allowing the regressor to capture impacts, rolling transitions, and transient bursts that are poorly handled by direct integration. The result is a model that tracks rapid onsets (Fig. 3f), generalizes across held-out events, and remains stable during tumbling while preserving uncertainty calibration in the field.

C. Limitations, extensions, and future work

As with any data-driven approach, reliability is strongest within the support of the training distribution. Stress tests indicate graceful degradation under noise and missingness, but

large distribution shifts will eventually erode both accuracy and coverage. Practical safeguards include monitoring conformal residuals during deployment, refreshing calibration with a small number of site-specific labeled events, and periodically revisiting the static thresholds as sensors age.

Appendix Fig. 9 stress-tests the trained model by injecting feature noise, simulating missingness, and probing out-of-distribution (OOD) distance. Errors increase monotonically with perturbation strength and with OOD distance, while residuals remain approximately homoscedastic. These diagnostics show that the estimator degrades gracefully under realistic sensor imperfections, but they also delineate the regime where re-calibration or additional labeled events become necessary. In practice, elevated residuals or sustained interval inflation can be used as a trigger to refresh the model before coverage deteriorates further.

Several extensions are immediately tractable: (i) integrating high-g accelerometer summaries to capture impacts and saltation peaks; (ii) mode-aware training that separates rolling, sliding, and saltation to improve interpretability; (iii) transfer learning or lightweight fine-tuning when moving across sites, grain sizes, or housing geometries; and (iv) sequential conformal methods to provide joint coverage across windows for event-level decision support.

VI. Conclusions

We demonstrate that gravity-compensated, integration-free learning on short IMU windows can recover projected speed and kinetic-energy time series with calibrated uncertainty across laboratory and field conditions. The approach uses a complementary filter, Hankel delay embeddings of dynamic acceleration and angular-rate magnitudes, a Matérn 3/2 Gaussian Process regressor, and split conformal prediction. The estimator attains competitive accuracy on held-out events and retains distribution-free coverage when deployed in the River Calder, where energy reconstructions exhibit realistic transient maxima and component splits aligned with angular-rate activity. Because the approach avoids explicit acceleration integration, it remains stable during impacts and tumbling while providing interval-valued outputs that support risk-aware interpretation and operational decisions. The framework is readily extensible and offers a scalable path to risk-aware kinematics and energetics from IMUs in geomorphic settings where conventional sensing is difficult. A practitioner-facing synthesis of deployment protocols and interpretation remains an important next step.

CRediT Authorship Contribution Statement

Georgios Maniatis conceptualized the methods, designed and executed the flume and field experiments, designed and implemented the processing workflow, performed requirements capture, co-designed the sensors, and drafted the manuscript. Jeffrey A. Tuhtan led sensor design, contributed significantly to the implementation of the ML workflow, and revised the manuscript. Gert Toming implemented the sensor design, constructed and verified the sensors, and revised the manuscript. Edward Curley contributed significantly to the Calder experiments and revised the manuscript. Charlie Gadd contributed significantly to the implementation of the ML workflow, data and code curation and corrections, and revised the manuscript. Richard Williams co-designed and supervised the Calder experiments, and revised the manuscript. Trevor Hoey contributed significantly to the conceptualization of the sensors and experiments, the geomorphic and hydraulic reasoning, and revised the final draft.

Acknowledgment

The development of Kivi was funded by the Royal Society (grant `RGS\R1\221251`, awarded to Georgios Maniatis, 2022). Jeffrey A. Tuhtan and Gert Toming were funded in part by the Estonian Centre of Excellence in IT (EXCITE), the Estonian Research Council Grant PRG2198, and by the European Union and the Estonian Research Council through project TEM-TA141. For the purpose of open access, the author(s) has (have) applied a Creative Commons Attribution (CC-BY) license to any Author Accepted Manuscript version arising from this submission.

References

[1] J. R. Gray, J. B. Laronne, and J. D. G. Marr, "Bedload-surrogate monitoring technologies," U.S. Geological Survey, Reston, VA, Scientific Investigations Report 2010-5091, 2010. [Online]. Available: https://pubs.usgs.gov/sir/2010/5091/

[2] M. T. Perks, P. S. Dalrymple, A. Eltner, S. Baumgartner, N. Rüther, M. Rühling, M. Detert, J. Aberle, A. Sukhodolov, M. Tamminga, C. Cierpka *et al.*, "Towards harmonisation of image velocimetry for river surface flow observations," *Earth System Science Data*, vol. 12, no. 2, pp. 1545–1559, 2020.

[3] C. D. Rennie and P. V. Villard, "Site specificity of bed load measurement using an acoustic doppler current profiler," *Journal of Geophysical Research: Earth Surface*, vol. 109, no. F3, p. F03003, 2004.

[4] D. Rickenmann, "Bedload transport measurements with geophones and other passive acoustic methods," *Journal of Hydraulic Engineering*, vol. 143, no. 6, p. 03117004, 2017.

[5] E. Akeila, Z. Salcic, and A. Swain, "Smart pebble for monitoring riverbed sediment transport," *IEEE Sensors Journal*, vol. 10, no. 11, pp. 1705–1717, 2010.

[6] E. Akeila, Z. Salcic, N. Kularatna, B. Melville, and A. Dwivedi, "Testing and calibration of smart pebble for river bed sediment transport monitoring," in *2007 IEEE Sensors*, 2007, pp. 1201–1204.

[7] G. Maniatis, T. Hoey, R. Hodge, D. Rickenmann, and A. Badoux, "Inertial drag and lift forces for coarse grains on rough alluvial beds measured using in-grain accelerometers," *Earth Surface Dynamics*, vol. 8, no. 4, pp. 1067–1099, 2020.

[8] K. Al-Obaidi and M. Valyrakis, "A sensory instrumented particle for environmental monitoring applications: Development and calibration," *IEEE Sensors Journal*, vol. 21, no. 8, pp. 10 153–10 166, 2021.

[9] G. Maniatis, G. Toming, T. Hoey, and J. A. Tuhtan, "Kivi: the smartest of pebbles," in *40th IAHR World Congress*, Vienna, Austria, 2023, pp. 2911–2917.

[10] S. L. Brunton and J. N. Kutz, *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*, 2nd ed. Cambridge, UK: Cambridge University Press, 2022.

[11] A. N. Angelopoulos and S. Bates, "Conformal prediction: A gentle introduction," *Foundations and Trends in Machine Learning*, vol. 16, no. 4, pp. 494–591, 2023. [Online]. Available: https://doi.org/10.1561/2200000101

[12] M. Church, "Bed material transport and the morphology of alluvial river channels," *Annual Review of Earth and Planetary Sciences*, vol. 34, pp. 325–354, 2006.

[13] J. Yuan, P. Wang, M. Wang, and M. Wang, "Experimental investigations on the response of bedload using a smart pebble accelerometer," *Sustainability*, vol. 15, no. 15, p. 11972, 2023.

[14] P. Wang, J. Yuan, M. Wang, and M. Wang, "Sediment vibration characteristics based on accelerometer measurements," *Scientific Reports*, vol. 13, no. 1, p. 1205, 2023.

[15] J. Jiao, Y. Sun, Y. An, Q. Liu, and X. Wang, "Design and calibration of spherical particles with embedded inertial measurement unit: Preserving kinematic equivalence," *Acta Mechanica Sinica*, vol. 42, no. 4, p. 324749, 2026. [Online]. Available: https://doi.org/10.1007/s10409-025-24749-x

[16] A. Sgarabotto, I. Manzella, K. Roskilly, M. J. Clark, G. L. Bennett, C. Luo, and A. M. A. Franco, "Cobble motion characterisation with smart sensors through laboratory experiments for ground-based landslide monitoring," *Geomorphica*, vol. 1, no. 1, p. 42, 2025. [Online]. Available: https://doi.org/10.59236/geomorphica.v1i1.42

[17] J. Peng, D. Chen, M. A. Hassan, G. Maniatis, L. Wang, and R. Nie, "Experiments on kinematic characteristics and energy dissipation in rockfall movement on a slope," *Physics of Fluids*, vol. 36, no. 10, p. 106625, 2024.

[18] M. Hamawi, J. P. L. Johnson, S. Bilek, J. M. Turowski, and J. B. Laronne, "Integrating Smartrock and seismic monitoring to investigate bedload dynamics in glacierized catchments," *EGUsphere*, 2025, preprint. [Online]. Available: https://doi.org/10.5194/egusphere-2025-591

[19] T. Geay, S. Zanker, C. Misset, and A. Recking, "Passive acoustic measurement of bedload transport: Toward a global calibration curve?" *Journal of Geophysical Research: Earth Surface*, vol. 125, no. 8, p. e2019JF005242, 2020.

[20] Y. Xie, B. W. Melville, A. Y. Shamseldin, C. N. Whittaker, and Y. Yang, "Direct measurement of the inertial drag and lift forces on entrained coarse particles at various protrusion heights," *Earth Surface Processes and Landforms*, vol. 48, no. 3, pp. 371–385, 2023.

[21] J. F. Fuentes-Pérez, F. J. Sanz-Ronda, and J. A. Tuhtan, "An open surface drifter for river flow field characterization," *Sensors*, vol. 22, no. 24, p. 9918, 2022.

[22] R. Mooney, G. Corley, A. Godfrey, L. Quinlan, and G. ÓLaighin, "Inertial sensor technology for elite swimming performance analysis: A systematic review," *Sensors*, vol. 16, no. 1, p. 18, 2016.

[23] P. Picerno, "25 years of lower limb joint kinematics by using inertial and magnetic sensors: A review of methodological approaches," *Gait & Posture*, vol. 51, pp. 239–246, 2017.

[24] C. Monoli, J. A. Tuhtan, L. Piccinini, and M. Galli, "Wearable technologies for monitoring aquatic exercises: A systematic review," *Clinical Rehabilitation*, vol. 37, no. 6, pp. 791–807, 2022.

[25] A. Serani, P. Dragone, F. Stern, and M. Diez, "On the use of dynamic mode decomposition for time-series forecasting of ships operating in waves," *Ocean Engineering*, vol. 267, p. 113235, 2023.

[26] G. Palma, A. Serani, K. McTaggart, S. Aram, D. W. Wundrow, D. Drazen, and M. Diez, "Bayesian dynamic mode decomposition for real-time ship motion digital twinning," arXiv:2411.14839, 2024, arXiv preprint. [Online]. Available: https://arxiv.org/abs/2411.14839

[27] L. Rosafalco, P. Conti, A. Manzoni, S. Mariani, and A. Frangi, "EKF–SINDy: Empowering the extended kalman filter with sparse identification of nonlinear dynamics," *Computer Methods in Applied Mechanics and Engineering*, vol. 431, p. 117264, 2024. [Online]. Available: https://doi.org/10.1016/j.cma.2024.117264

[28] M. S. Hemati, M. O. Williams, and C. W. Rowley, "Dynamic mode decomposition for large and streaming datasets," *Physics of Fluids*, vol. 26, no. 11, p. 111701, 2014. [Online]. Available: https://doi.org/10.1063/1.4901016

[29] I. Gibbs and E. J. Candès, "Conformal inference for online prediction with arbitrary distribution shifts," *Journal of Machine Learning Research*, vol. 25, no. 162, pp. 1–36, 2024. [Online]. Available: https://jmlr.org/papers/v25/22-1218.html

[30] I. Gibbs and E. J. Candès, "Adaptive conformal inference under distribution shift," in *Advances in Neural Information Processing Systems*, vol. 34, 2021, pp. 1660–1672. [Online]. Available: https://proceedings.neurips.cc/paper/2021/hash/0d441de75945e5acbc865406fc9a2559-Abstract.html

[31] A. Cini, A. Jenkins, D. Mandic, C. Alippi, and F. M. Bianchi, "Relational conformal prediction for correlated time series," in *Proceedings of the 42nd International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, vol. 267. PMLR, 2025, pp. 10 949–10 965. [Online]. Available: https://proceedings.mlr.press/v267/cini25a.html

[32] J. Lei, M. G'Sell, A. Rinaldo, R. J. Tibshirani, and L. Wasserman, "Distribution-free predictive inference for regression," *Journal of the American Statistical Association*, vol. 113, no. 523, pp. 1094–1111, 2018. [Online]. Available: https://doi.org/10.1080/01621459.2017.1307116

[33] R. A. Hodge, T. B. Hoey, and L. S. Sklar, "Bed load transport in bedrock rivers: The role of sediment cover in grain entrainment, translation, and deposition," *Journal of Geophysical Research: Earth Surface*, vol. 116, no. F4, p. F04028, 2011.

[34] A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Analytical Chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964.

[35] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[36] S. O. H. Madgwick, A. J. L. Harrison, and R. Vaidyanathan, "Estimation of imu and marg orientation using a gradient descent algorithm," in *2011 IEEE International Conference on Rehabilitation Robotics*, 2011, pp. 1–7.

AUTHOR BIOGRAPHIES

Georgios Maniatis Georgios Maniatis received the Ph.D. degree in geographical and earth sciences from the University of Glasgow, Glasgow, U.K., in 2017, and the M.Sc. degree in freshwater systems science from the University of Glasgow in 2012. He received the degree in environmental engineering from the Technical University of Crete, Chania, Greece, in 2011. He is a Senior Lecturer in Physical Geography with the School of Applied Sciences, University of Brighton, U.K. His research interests include sediment transport, instrumented particles, and river morphodynamics.

Jeffrey Tuhtan Jeffrey A. Tuhtan received the B.Sc. degree in civil engineering from California Polytechnic State University, San Luis Obispo, CA, USA, in 2004, and the M.Sc. degree in water resources engineering and management and the Dr.Eng. degree from the University of Stuttgart, Germany, in 2007 and 2011, respectively. He is currently with the Centre for Biorobotics, Tallinn University of Technology, where he leads the Environmental Sensing and Intelligence Group. His current research interests include environmental intelligence and bio-inspired underwater sensing.

Gert Toming Gert Toming received the B.Sc. and M.Sc. degrees in information technology from the University of Tartu, Tartu, Estonia, in 2007 and 2010, respectively. He received the Ph.D. degree in information and communication technology from the Faculty of Information Technology, Tallinn University of Technology, Tallinn, Estonia, in 2017. He is currently a Researcher with the Centre for Environmental Intelligence and Sensing in the Department of Computer Systems, Tallinn University of Technology. His research focuses on sensor- and camera-based measurements in harsh underwater environments.

Edward Curley Edward received the degree in geographical and earth sciences from the University of Glasgow, Glasgow, U.K., in 2021, and the M.Sc. degree in conservation and resource management from the University of Liverpool in 2016. They received the degree in Zoology from the University of Liverpool in 2015. Edward is a Research Associate of the University of Glasgow, U.K. Edward's research interests include ecohydrology, remote sensing, and catchment-scale restoration.

Charlie Gadd Charlie Gadd received the B.Sc. degree in Geography from the University of Brighton, U.K., in 2022. He is currently working toward the Ph.D. degree at Royal Holloway, University of London, U.K., focusing on applications of machine learning in smart transport with a particular emphasis on conformal methods. His research interests include statistical modelling, algorithmic learning, and remote sensing as applied to fluvial geomorphology and transportation systems.

Richard Williams Richard Williams received the B.Sc. degree in Geography from the University of Cambridge, U.K., in 2004, the M.Sc. degree in Science of the Environment from Lancaster University, U.K., in 2005, and the Ph.D. degree from Aberystwyth University, U.K., in 2014. He is currently a science leader at Earth Sciences New Zealand (ESNZ), a public research organisation, where he is the General Manager for ESNZ's land and water science mission. He is also an Honorary Professor of River Science at the University of Glasgow, U.K. Prior to moving to ESNZ in 2024, he held a professorship at the University of Glasgow. His research spans temperate and tropical river systems, with interests including flood estimation in data-poor environments, multi-decadal river morphodynamics from satellite remote sensing, and monitoring and modelling river restoration and natural flood-risk management.

Trevor Hoey Trevor Hoey is Professor of River Science in the Department of Civil and Environmental Engineering, Brunel University of London, London, U.K. He is a fluvial geomorphologist with interests in coarse-grained sediment transport and the dynamics of bedrock–alluvial rivers. His recent work includes river dynamics and flood risk in the Philippines, the impacts of mining waste on river catchments, rough-bed river hydraulics, and sediment transport mechanics.

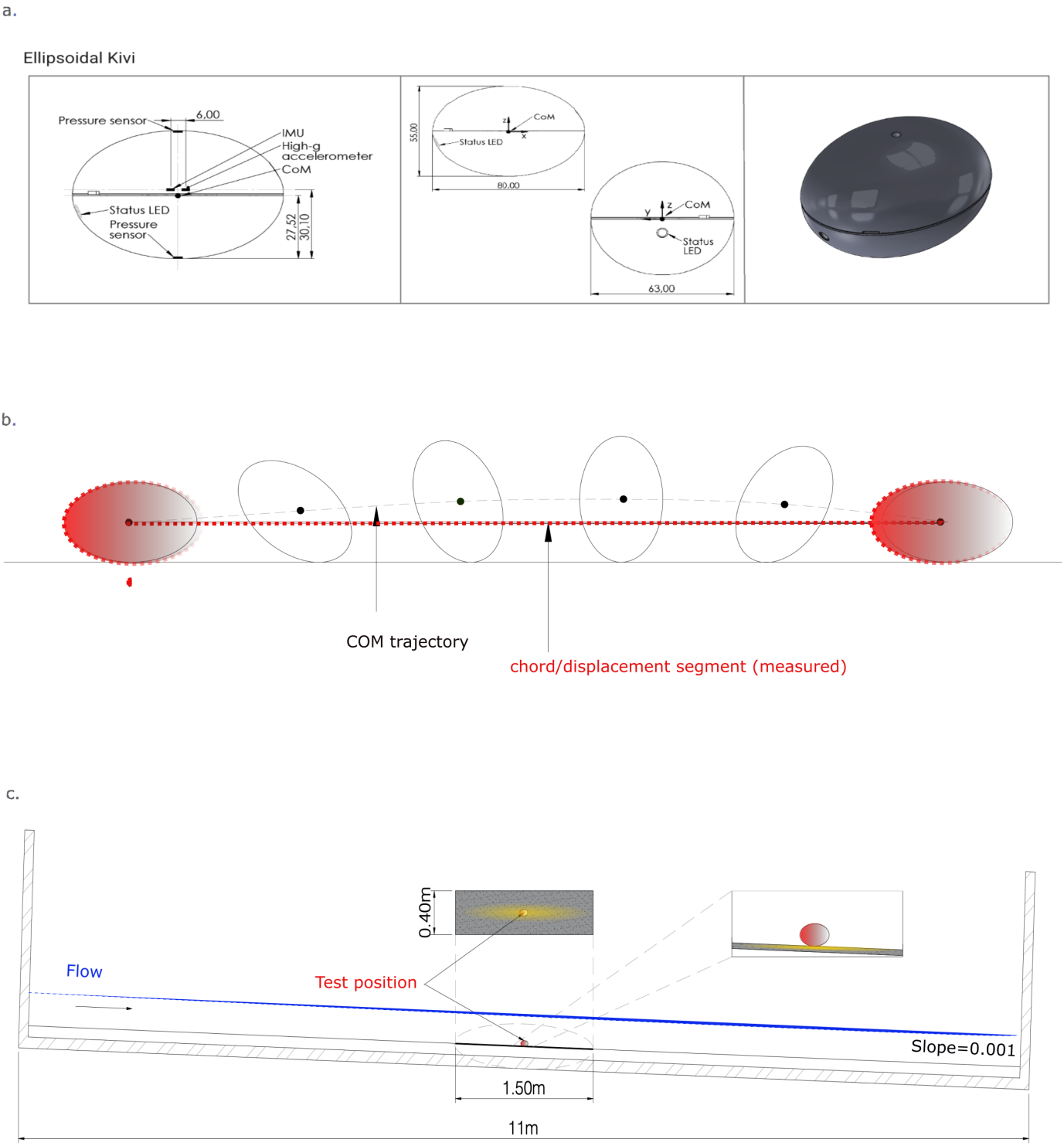


Fig. 1. Instrumented ellipsoidal Kivi and experimental configuration. (a) Instrumented ellipsoidal “Kivi” showing the location of the centre of mass, inertial sensors, pressure sensors and status LED; (b) definition of centre-of-mass motion during rolling; (c) flume set-up and measurement section.

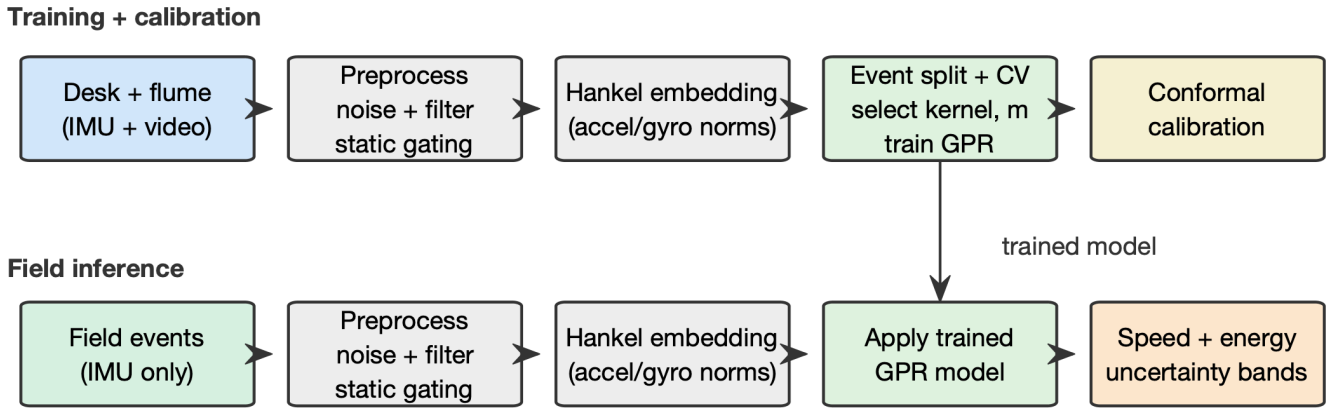


Fig. 2. **Workflow overview across training and field inference.** Desk and flume events provide labeled training data (IMU + video), which pass through preprocessing and Hankel embedding before event-disjoint model selection, GPR training, and conformal calibration. The trained model is then applied to field IMU data to produce projected speed, kinetic energy, and uncertainty bands.

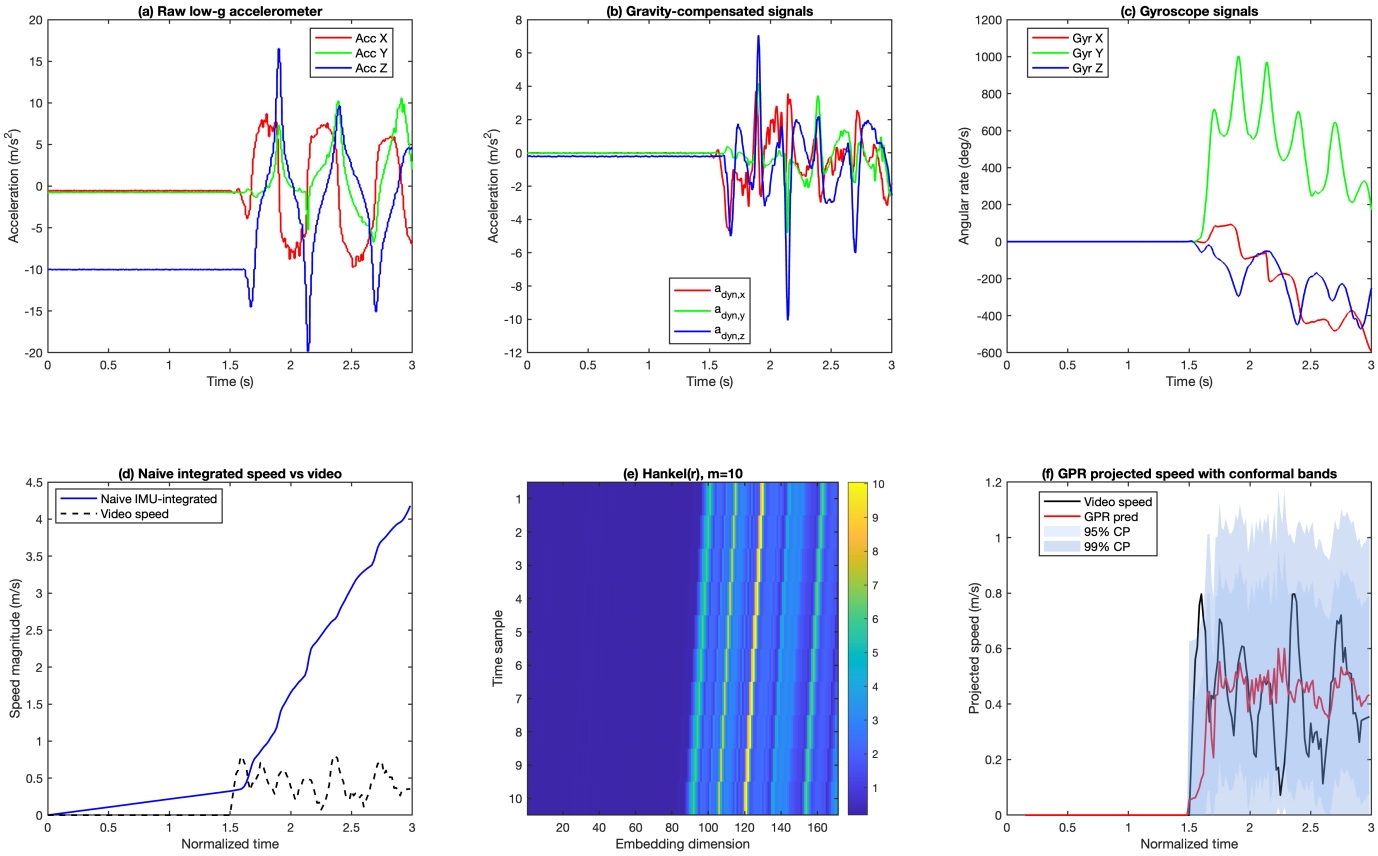


Fig. 3. **End-to-end processing and learning pipeline.** Example dry desk event with motion onset around $1.5\,\mathrm{s}$. (a) Raw low-g accelerometer signals on an event window. (b) Complementary-filter g ravity-compensated a cceleration. (c) G yroscope s ignals (dominant y -component bursts indicate rolling; peaks align with video in (d)). (d) Naive IMU-integrated speed versus video reference. (e) Hankel embedding structure (delay dimension from the trained model, $m = 10$). (f) GPR projected speed with $95\,\%/99\,\%$ conformal bands and tracked ground truth (static segments clamped via gyro gating; first m samples omitted for Hankel alignment).

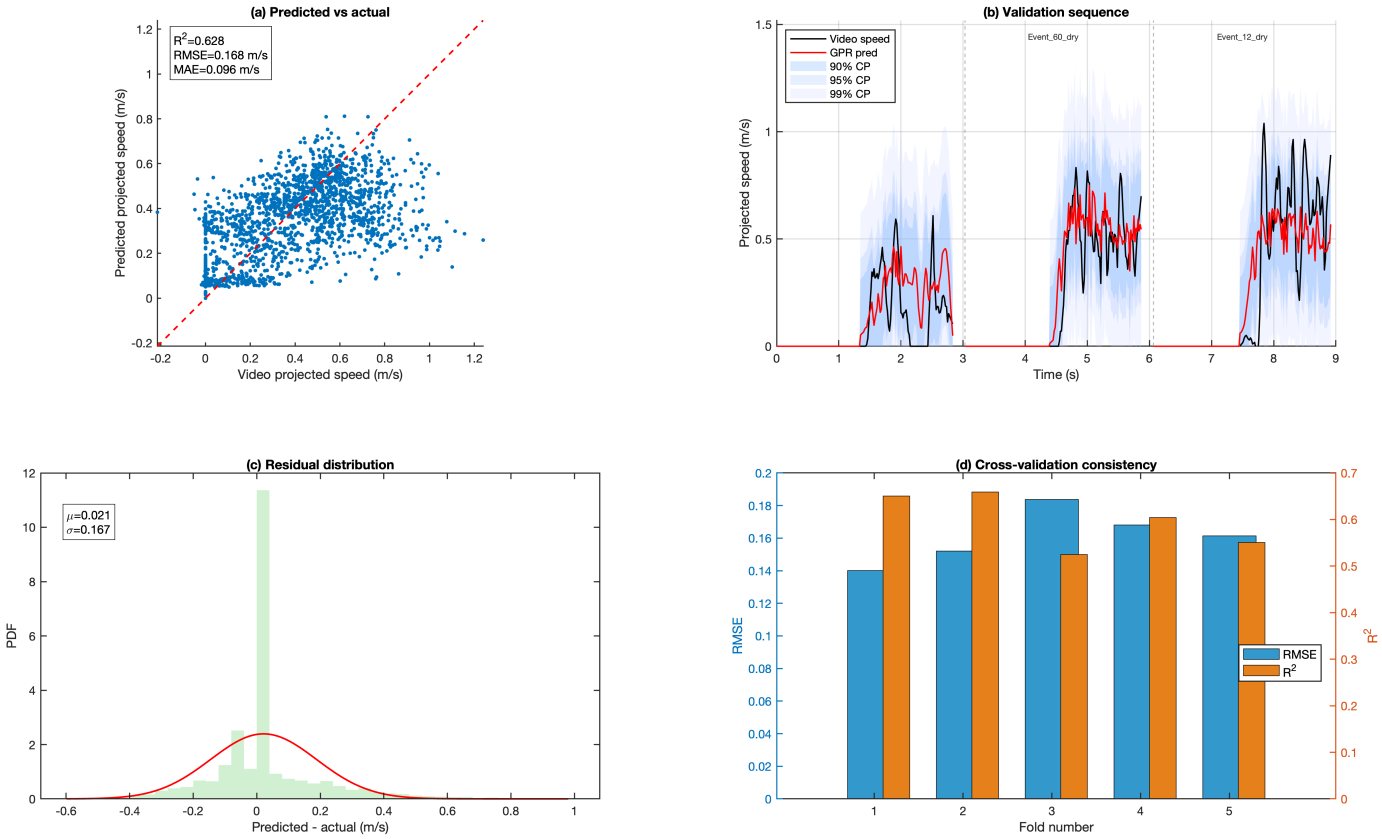


Fig. 4. **Model performance and validation.** (a) Predicted versus video-projected speed on the hold-out set ($R^2 = 0.628$, $\mathrm{RMSE} = 0.168\,\mathrm{m\,s^{-1}}$, $\mathrm{MAE} = 0.096\,\mathrm{m\,s^{-1}}$). (b) Stitched sequence of three held-out *dry* validation events with video reference and conformal bands. (c) Residual distribution (predicted minus actual) with near-zero mean. (d) Event-grouped cross-validation consistency across folds.

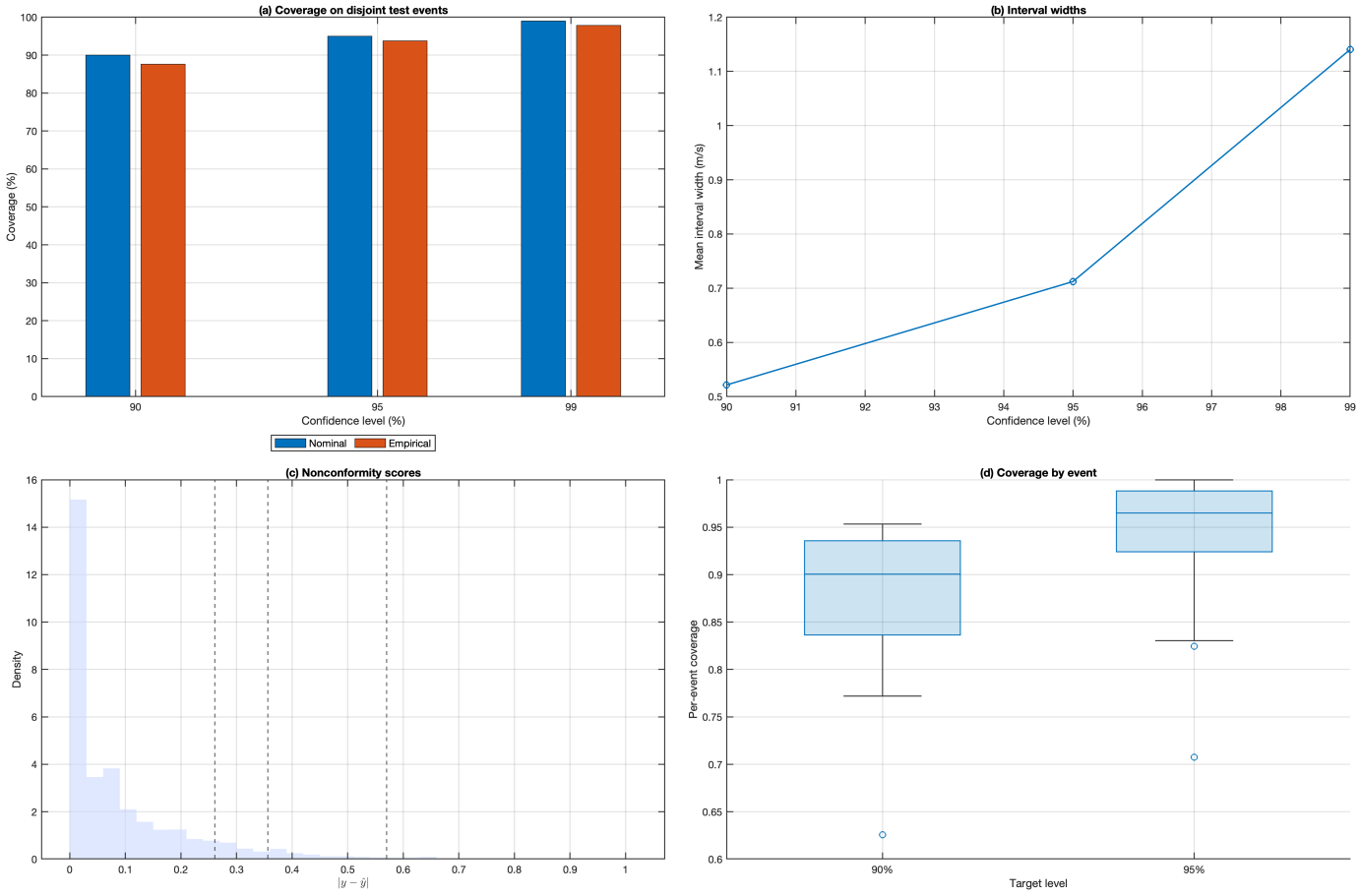


Fig. 5. **Uncertainty quantification summary.** (a) Nominal versus empirical coverage at the target levels. (b) Interval widths for each target level. (c) Nonconformity score distribution with quantile markers. (d) Per-event coverage distribution for the lower target levels.

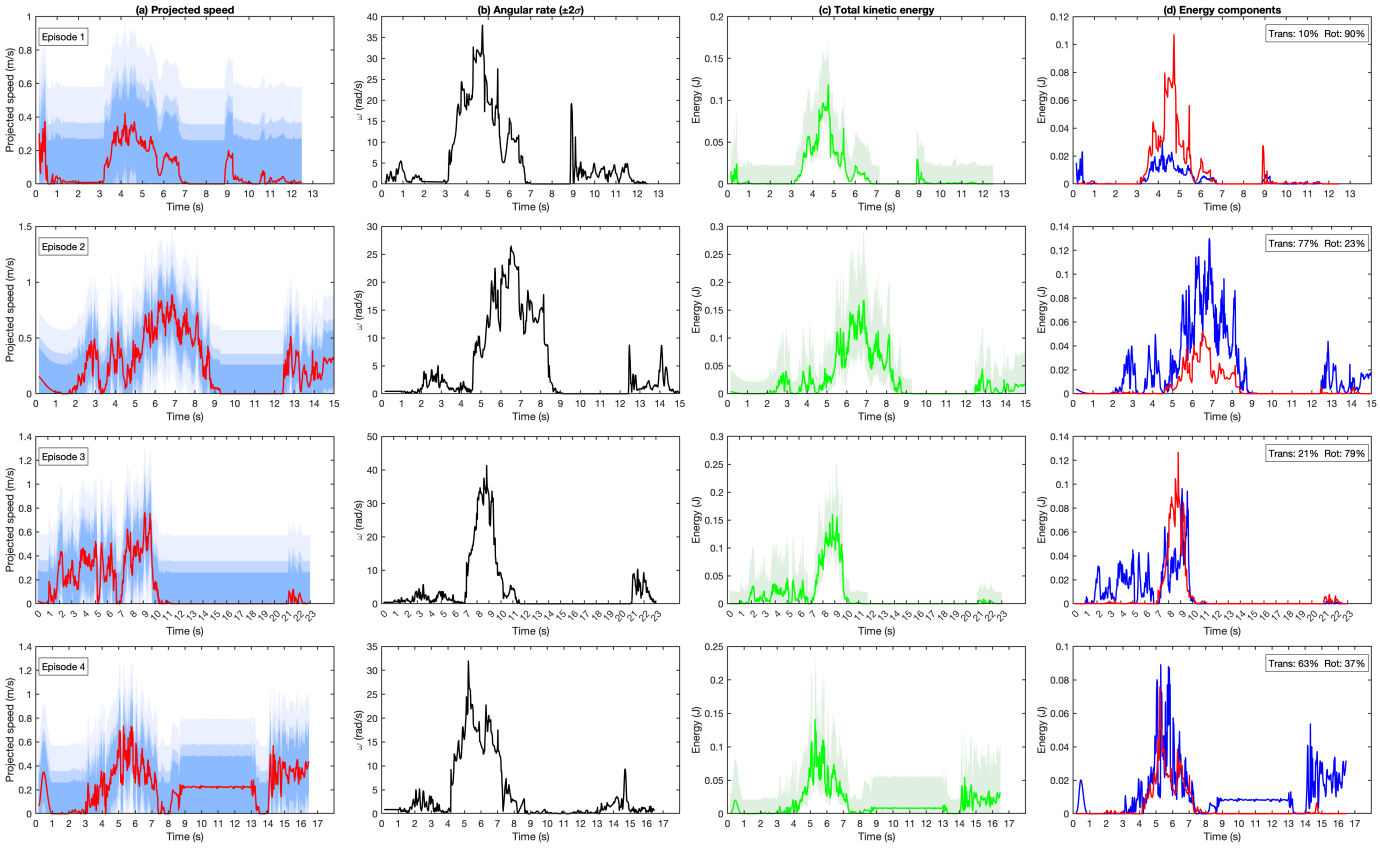


Fig. 6. **Field demonstration: River Calder.** Rows correspond to four representative deployment episodes. Column (a): predicted projected speed with conformal bands. Column (b): angular-rate magnitude with a centred $\pm 2\sigma$ noise envelope (gyro noise scale from static characterization). Column (c): total kinetic energy with a $95\,\%$ band derived from conformal speed bounds. Column (d): energy components (E_{trans}, E_{rot}); annotations report the translational/rotational percentage at the dominant peak.

TABLE III

PAIRED SIGNIFICANCE TESTS FOR COMPARATOR MODELS VERSUS GPR ON OUTER FOLDS.

Comparator	Metric	p-value	Folds	Interpretation
LSTM	RMSE	0.8125	5	Not significant at $\alpha = 0.05$
LSTM	MAE	0.1875	5	Not significant at $\alpha = 0.05$
LSTM	R2	0.8125	5	Not significant at $\alpha = 0.05$
LSTM	NSE	0.8125	5	Not significant at $\alpha = 0.05$
SVR_RBF	RMSE	0.4375	5	Not significant at $\alpha = 0.05$
SVR_RBF	MAE	0.3125	5	Not significant at $\alpha = 0.05$
SVR_RBF	R2	0.4375	5	Not significant at $\alpha = 0.05$
SVR_RBF	NSE	0.4375	5	Not significant at $\alpha = 0.05$
LSBoost	RMSE	0.8125	5	Not significant at $\alpha = 0.05$
LSBoost	MAE	0.8125	5	Not significant at $\alpha = 0.05$
LSBoost	R2	0.8125	5	Not significant at $\alpha = 0.05$
LSBoost	NSE	0.8125	5	Not significant at $\alpha = 0.05$

APPENDICES

Appendix A: Paired Statistical Comparison Details

Table III reports paired Wilcoxon signed-rank tests on outer-fold metrics for each comparator versus GPR. At $\alpha = 0.05$, none of the reported paired tests are statistically significant with five folds, so model ordering is interpreted as indicative performance under the current sample size rather than definitive dominance.

Supplementary figures that support the preprocessing, kernel sweep, and robustness checks are provided below (Appendix Fig. 7, Appendix Fig. 8, and Appendix Fig. 9).

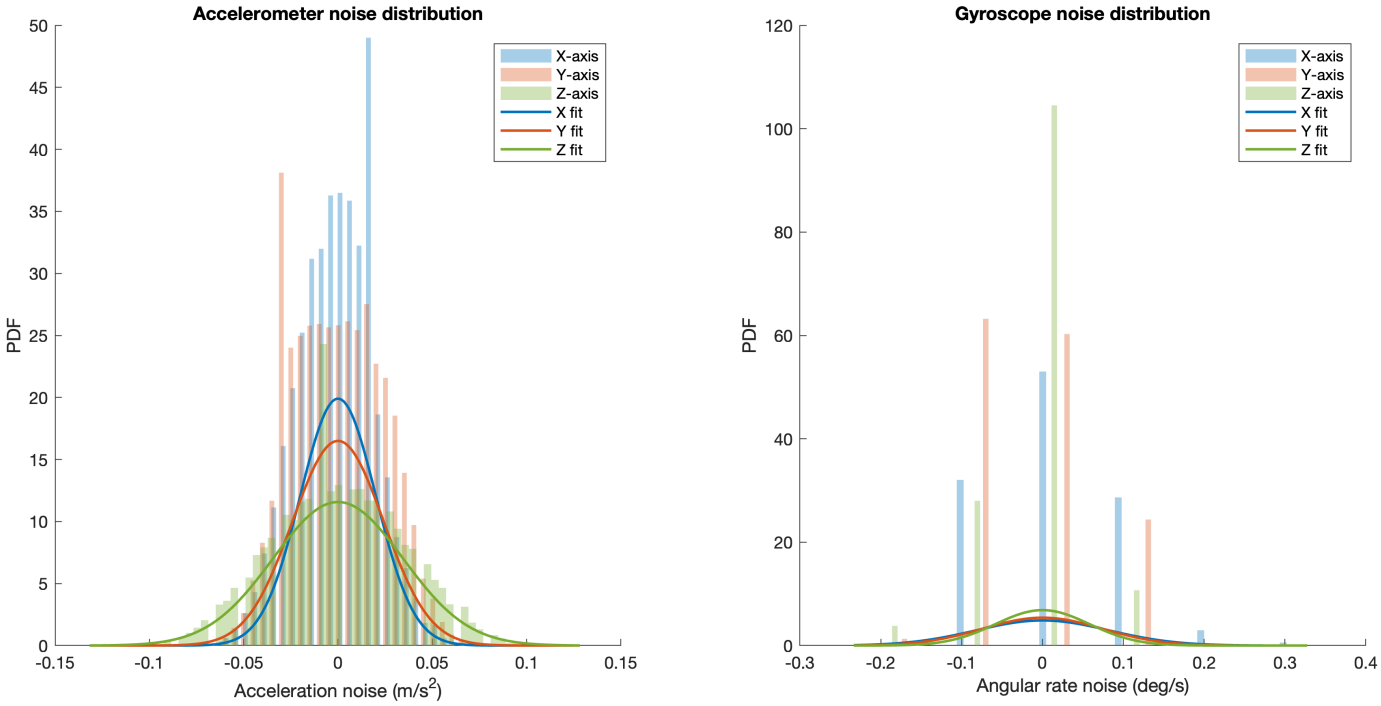


Fig. 7. **Static sensor noise characterization.** Accelerometer and gyroscope noise histograms with Gaussian fits, used to set the deadband and tilt-gating thresholds.

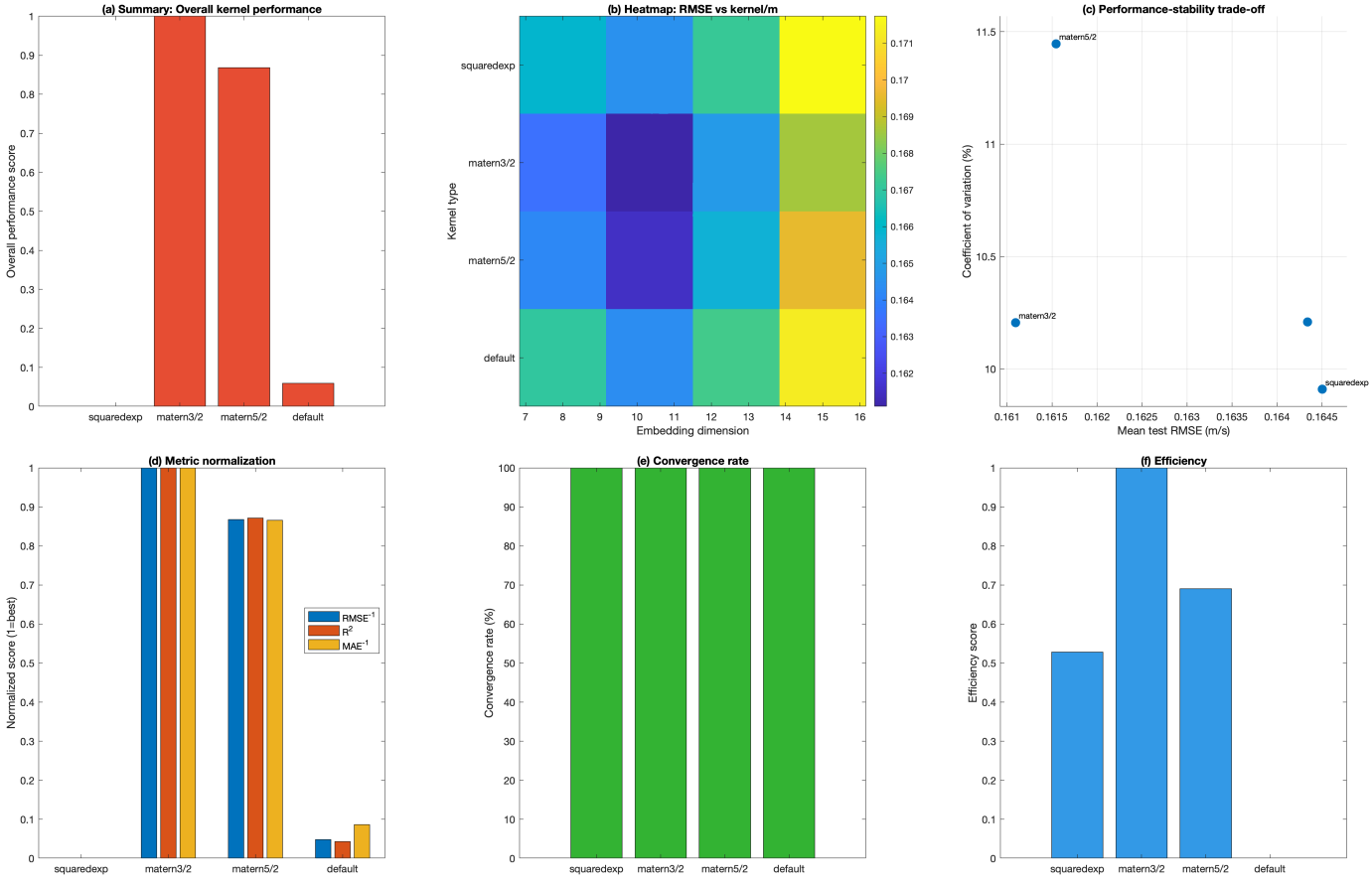


Fig. 8. **Comprehensive kernel and embedding analysis.** Overall scores, relative RMSE by delay dimension, performance–stability trade-off, metric normalization, convergence rate, and efficiency derived from event-grouped cross-validation. Panels (a), (d), and (f) use min–max normalization across kernels, so the worst kernel maps to zero and its bar can appear absent.

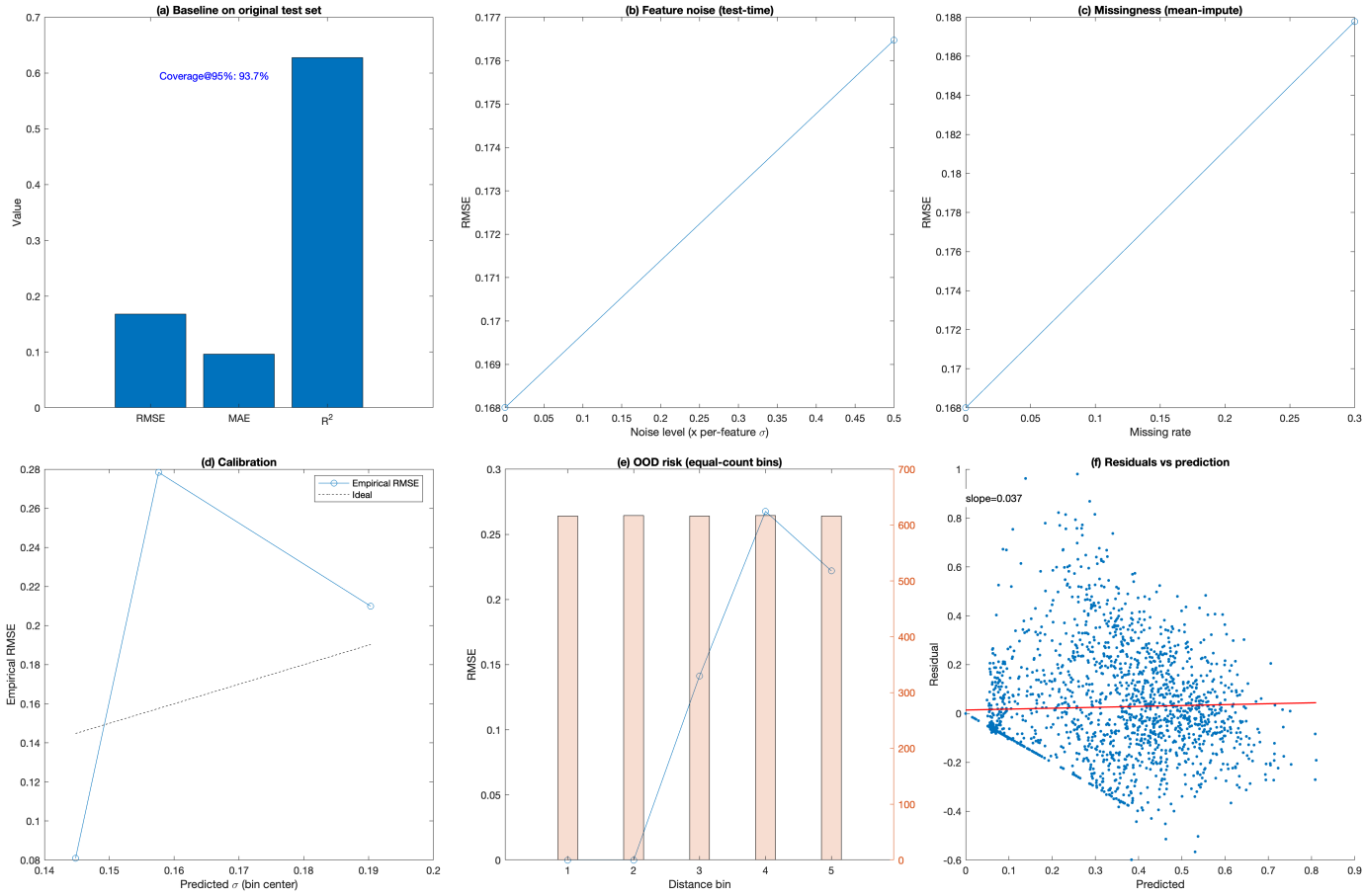


Fig. 9. **Fast robustness assessment of the already-trained model.** Baseline metrics, noise sensitivity, missingness sensitivity, calibration, OOD risk, and residual diagnostics computed from the current run.