

## Article

# UAV-Based River Velocity Estimation Using Optical Flow and FEM-Supported Multiframe RAFT Extension

Andrius Kriščiūnas <sup>1,\*</sup>, Vytautas Akstinas <sup>2</sup>, Dalia Čalnerytė <sup>1</sup>, Diana Meilutytė-Lukauskienė <sup>2</sup>,  
Karolina Gurjazkaitė <sup>2</sup>, Tautvydas Fyleris <sup>3</sup> and Rimantas Barauskas <sup>1</sup>

<sup>1</sup> Department of Applied Informatics, Kaunas University of Technology, Studentų St. 50, LT-51368 Kaunas, Lithuania

<sup>2</sup> Laboratory of Hydrology, Lithuanian Energy Institute, Breslaujos St. 3, LT-44403 Kaunas, Lithuania

<sup>3</sup> Department of Software Engineering, Kaunas University of Technology, Studentų St. 50, LT-51368 Kaunas, Lithuania

\* Correspondence: andrius.krisciunas@ktu.lt

## Highlights

### What are the main findings?

- A hybrid dataset was created by linking FEM-based hydrodynamic simulations with georeferenced UAV RGB video for supervised optical flow learning.
- A multiframe RAFT model with a pre-correlation Fuse-GRU module produces stable and physically consistent surface velocity fields, with average angular errors lower than 15°.

### What are the implications of the main findings?

- The method enables accurate, non-contact river surface flow velocity estimation from UAV video under diverse hydraulic and illumination conditions.
- The framework supports hydrodynamic model validation and scalable UAV-based hydrometric monitoring.

## Abstract

Quantifying river surface flow velocity is essential for hydrodynamic modelling, flood forecasting, and water resource management. Traditional in situ methods provide accurate point measurements but are costly and limited in spatial coverage. Unmanned aerial vehicles (UAVs) offer a flexible, non-contact alternative for high-resolution monitoring. Optical flow is a tracer-independent technique for deriving velocity fields from RGB video, making it well suited to UAV-based surveys. However, its operational use is hindered by the limited availability of annotated datasets and by instability under low-texture or noisy conditions. This study combines a Finite element method (FEM)-based physical flow model with UAV video to generate reference datasets and introduces a modified Recurrent All-Pairs Field Transforms (RAFT) architecture based on multiframe sequences. A Gated Recurrent Unit fusion module (Fuse-GRU) is incorporated prior to correlation computation, improving robustness to illumination changes and surface homogeneity while maintaining computational efficiency. The proposed model delivers stable, physically consistent velocity estimates across multiple rivers and flow conditions. Accuracy improves with higher spatial resolution and moderate temporal spacing. Compared to field measurements, the average angular difference ranged from 8 to 15°. The high error values were mainly caused by inaccuracies in the physical model and by complex river features. These findings confirm that multiframe optical flow can reproduce realistic river flow patterns with accuracy



Academic Editor: Pablo Rodríguez-González

Received: 30 January 2026

Revised: 18 March 2026

Accepted: 19 March 2026

Published: 21 March 2026

**Copyright:** © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

comparable to physically-based simulations, thereby supporting UAV-based hydrometric monitoring and model validation.

**Keywords:** multiframe deep learning; physics-informed modeling; river surface velocity; optical flow

---

## 1. Introduction

Understanding river surface flow velocity is essential for hydrodynamic modelling, flood risk management, sediment transport studies, ecosystem monitoring, and hydraulic engineering applications [1,2]. Traditional in situ techniques, such as current meters, float tracking, and acoustic Doppler current profilers (ADCP), remain reliable for point or profile-based velocity measurements [1,3], but are often costly, time-consuming, and limited in spatial and temporal coverage [4,5]. Their use is further constrained in hazardous, remote, or rapidly changing environments, where fast, large-scale, and non-invasive measurements are required [2,6]. In this context, unmanned aerial vehicles (UAVs) have emerged as a transformative technology for river monitoring, offering rapid deployment, flexible operation, and the ability to collect high-resolution imagery from previously inaccessible locations [7]. These advances have led to the growing adoption of image-based velocimetry methods that use UAV-derived video or imagery to retrieve surface velocity information over wide river reaches [1,8].

The family of image-based velocimetry techniques includes large-scale particle image velocimetry (LSPIV) [9], large-scale particle tracking velocimetry (LSPTV), space–time image velocimetry (STIV) [10,11], and optical flow methods [12,13]. LSPIV and LSPTV rely on visible tracers on the water surface, either naturally occurring, such as foam or floating debris, or artificially introduced particles [1,14]. STIV estimates surface flow velocity by analyzing spatiotemporal patterns of tracers in video sequences, providing a non-intrusive alternative for river flow monitoring [12]. Although these methods can provide accurate velocity fields in turbulent flows, their reliability decreases significantly in homogeneous, low-texture conditions or during floods, when tracer seeding is impractical or unsafe [4,14]. Various post-processing algorithms, such as Time Frequency Analysis (TiFA) applied to LSPIV results, aim to improve river surface velocity estimation in low tracer density conditions [15]. Optical flow, by contrast, estimates pixel displacements directly between consecutive frames, allowing the retrieval of surface velocity vectors without explicit reliance on tracers [2]. Early studies demonstrated that optical flow could produce spatially consistent velocity fields comparable to LSPIV, and subsequent developments extended its use to UAV imagery [3,6,16].

Recent research has focused on improving the robustness of optical flow approaches for natural rivers by integrating deep learning techniques. Architectures such as Recurrent All-Pairs Field Transforms (RAFT) [17] and Convolutional Neural Networks using pyramid, warping, and cost volume (PWC-Net) have significantly enhanced the ability to detect motion in challenging conditions, handling platform instability, reflections, and low contrast, which often limit classical methods [18,19]. Traditional optical flow methods typically use median filtering, spatial pyramids, and optimization approaches to minimize energy functions [20], or variational models [21], and usually process a single case. An important advantage of deep learning models is their good generalization, as networks like RAFT learn features and necessary patterns from data representing diverse environmental conditions [17,22]. Such models can operate on RGB video with limited textural content, where traditional particle-based approaches would struggle. Furthermore, real-time imple-

mentations have been proposed, such as UAV platforms integrating edge-computing units with convolutional neural networks and optimized optical flow algorithms, demonstrating the feasibility of near-operational hydrometric systems [23].

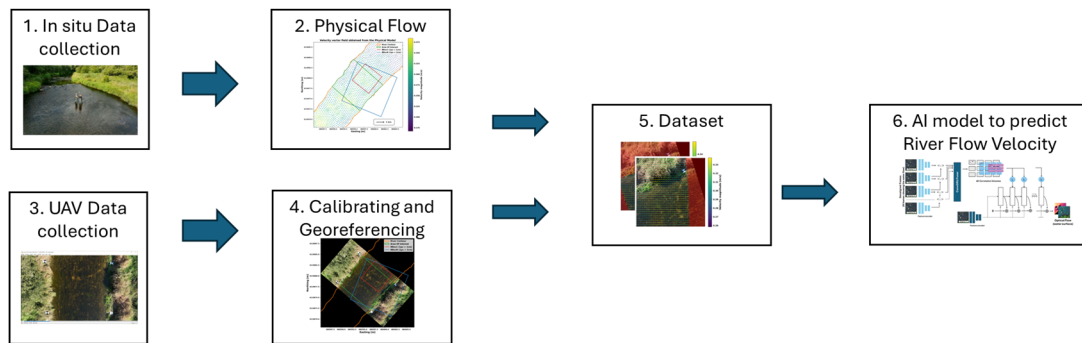
There is considerable interest in integrating UAV-based RGB data with other sensors to address the limitations of single-modality observations. Thermal cameras enhance tracer visibility by detecting subtle surface temperature gradients, Doppler radar provides independent flow validation, and bathymetric sonar or light detection and ranging (LiDAR) technology offer geometric context for discharge estimation [24–26]. These multi-sensor workflows have proven effective in specific case studies, such as combining thermal and RGB imagery to resolve velocity fields with deviations as low as 0.01 m/s [27], or using UAV bathymetry alongside velocimetry for discharge estimation in complex morphologies [28]. However, employing multiple sensors introduces additional complexity, costs, and payload requirements, which may limit practical deployment, particularly in emergency scenarios.

Despite the promise of sensor fusion, the most widely available and practical option remains the use of RGB cameras, which are standard on nearly all UAV platforms. Focusing on RGB video offers unmatched accessibility and deployment speed but introduces challenges that must be addressed to ensure accuracy. River reaches with low turbulence or limited surface tracers provide little textural contrast for tracking, while reflections, glare, and illumination changes reduce the reliability of optical flow estimation [2,16]. UAV motion adds further noise, requiring stabilization and geometric corrections such as deshaking, lens distortion correction, and orthorectification [6]. When these pre-processing steps are adequately implemented, studies have shown that UAV-based RGB velocimetry can achieve accuracy comparable to conventional field instruments. For instance, Eltner et al. [27] and Torres et al. [29] demonstrated deviations below 10% compared to ADCP and field measurements, whereas Koutalakis et al. [30] highlighted limitations in intermittent streams where dense vegetation and poor textural conditions compromised results.

Previous work by Kriščiūnas et al. [31] introduced a framework for UAV-based river flow velocity determination using optical flow recognition. It highlighted two major challenges: the lack of suitable datasets for training robust models and the inherent limitations of frame-to-frame optical flow estimation. Motivated by these gaps, this paper presents a modified Multi-Frame RAFT architecture (MF-RAFT) with integration of a gated recurrent fusion unit module (Fuse-GRU) to enable river flow prediction from aerial video RGB streams. A comprehensive analysis of the results is conducted for different spatial resolutions, temporal strides, and valid-pixel coverage groups across datasets obtained from three river stretches.

## 2. Materials and Methods

The general workflow of this study is summarized in Figure 1. It comprises two complementary dataset preparation paths. Steps 1 and 2 involve in situ data collection and the construction of a physical flow model for the specific river segment using the finite element method (FEM), to obtain reference velocity fields expressed as vector fields. Steps 3 and 4 involve UAV-based RGB video acquisition of the same segment, followed by calibration and georeferencing to correct perspective distortions and ensure spatial consistency. Both dataset branches are integrated in Step 5, where reference and UAV-derived data are combined into a single dataset. This dataset is then used in Step 6 to develop and evaluate an artificial intelligence model for predicting river flow velocity from UAV imagery.



**Figure 1.** Research Workflow.

Although presented sequentially in Figure 1, the methodological details are discussed thematically in the following subsections. Specifically, the optical flow formulation underlying velocity estimation is provided in Section 2.1, the artificial intelligence model is described in Section 2.2, the experimental study area and data collection are detailed in Section 2.3, and the dataset preparation procedures are explained in Section 2.4.

### 2.1. Optical Flow Formulation for River Velocity Subsection

Optical flow refers to the apparent motion of brightness patterns between two consecutive frames in an image sequence. The classical formulation relies on the brightness constancy assumption, which states that the intensity  $I(x, y, t)$  of a pixel at location  $(x, y)$  and time  $t$  remains unchanged as it moves over time. For small displacements, this constraint can be expressed as in Equation (1), following the classical formulation of Horn and Schunck [32].

$$\frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v + \frac{\partial I}{\partial t} = 0 \quad (1)$$

Here,  $(u, v)$  denotes the horizontal and vertical components of the optical flow vector. Equation (1) provides one constraint for two unknowns, leading to the well-known aperture problem, which is typically addressed by introducing spatial smoothness constraints or more advanced formulations. The terms  $\frac{\partial I}{\partial x}u$  and  $\frac{\partial I}{\partial y}v$  represent the change in pixel intensity in the horizontal and vertical components, respectively. The temporal derivative term  $\frac{\partial I}{\partial t}$  represents the change in pixel intensity between consecutive frames and is essential for linking the spatial displacement of features with their motion over time.

In river monitoring using pipeline UAV-based video, image sequences are calibrated and georeferenced so that each pixel corresponds to a fixed ground location and the spatial resolution remains constant; that is, one pixel (px) represents a predefined constant value  $C$  in meters (m) ( $1 \text{ px} = C \text{ m}$ ). Under these conditions, the displacement of features in the image plane can be directly converted into physical velocity vectors. If the flow velocity field within the river segment remains approximately constant during the observation period, the optical flow vectors remain unchanged between consecutive frames, provided that the time interval  $\Delta t$  is constant:

$$d_i \approx d_{i+1} \approx \dots \approx d_{i+N-1} \quad (2)$$

Here,  $d_i = (u_i, v_i)$  denotes the displacement vector between frame  $t$  and  $t + \Delta t$ , and  $N$  is the number of frames. Equation (2) highlights the temporal coherence property, which allows us to extend inference and model training beyond a simple two-frame input to sequences of  $N$  frames, thereby improving robustness against noise, illumination changes, and local texture deficiencies.

Once displacement  $d = (u, v)$  is estimated, the corresponding physical flow velocity  $V$  (in meters per second) can be derived as follows:

$$V = \frac{\sqrt{u^2 + v^2}C}{\Delta t} \quad (3)$$

Here,  $C$  is the pixel-to-meter conversion factor obtained during georeferencing and  $\Delta t$  is the frame interval. Equation (3) establishes the link between optical flow displacements and physical velocity, enabling UAV-based optical methods to provide quantitative estimates of river flow.

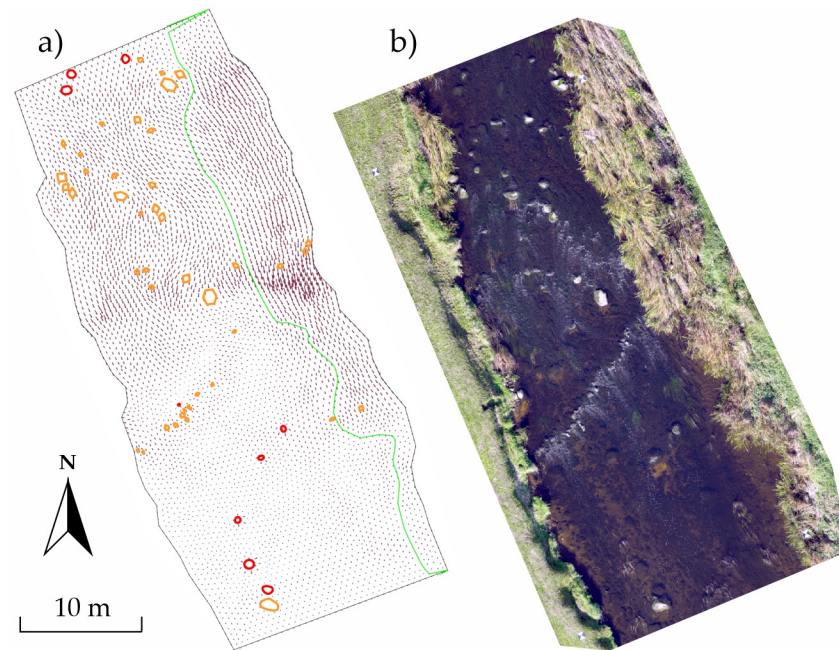
## 2.2. SWE-FE Model

The 2D model for each river segment was developed in the finite element (FE) software COMSOL Multiphysics 6.2, using the shallow water equations (SWE) application mode, following the scheme presented in [33]. The SWE describe shallow flow in a 2D region, represented by river bottom height elevation varying with coordinates and water surface height. The SWE modelling results were calibrated and validated using field measurements of flow velocity at 0.6 of the total depth (from the surface). Bottom resistance was incorporated as bottom shear, calculated using a hydraulic resistance formula with an empirically determined coefficient. Flow-hindering stresses due to vegetation at specific points were set proportional to velocity and depth, with the coefficient selected according to the pre-determined vegetation type (vegetation-free, bank vegetation, or dense vegetation zones). The coefficients for each river stretch were adjusted by comparing computational results with measured data and selecting those that provided the best match for average flow velocity, the lowest relative velocity errors, and a balance between positive and negative error values. The SWE-FE model was constructed using the discontinuous Galerkin approach [34] to discretize the SWE, and the Lax-Friedrichs flux [35] to ensure numerical stability. Numerical integration in time was performed using the explicit Runge-Kuta method until the flow became stationary or near-stationary. The output of simulation using FEM is presented as vectors of flow velocities within the calculation area.

The orthomosaic UAV image of one of the analyzed river stretches (Mūša) is shown in Figure 2b. The geometry of the SWE-FE model was created using data acquired from both direct field measurements and high-resolution UAV aerial images. The direct measurements were used in the SWE-FE model to define the bottom heights and water surface elevations. The 2D Delaunay triangulation was created from the measured points, with linear shape functions used to interpolate values at the SWE-FE model mesh nodes. The in situ measurements of flow velocity vectors (speed and direction at 0.6 of the depth) were also used as reference values to evaluate the simulation results and adjust the coefficients during the model calibration stage. The vegetation type was manually defined after analysis of the UAV aerial images. The SWE-FE model mesh was generated by taking into account the results of semi-automatic recognition of boulders above and below water from the UAV images [36]. The polygons defining the banks and the velocity field obtained using the FEM simulation are presented in Figure 2a.

## 2.3. AI Model for Processing UAV Video Sequences

Recent advances in optical flow estimation have been largely driven by deep learning models, with RAFT establishing itself as a state-of-the-art architecture due to its accuracy and robustness across diverse benchmarks [17]. RAFT iteratively refines dense flow fields through recurrent updates, making it suitable for applications requiring high precision. However, its architecture was originally designed for frame-to-frame estimation, which may limit its ability to fully exploit temporal consistency in longer image sequences.

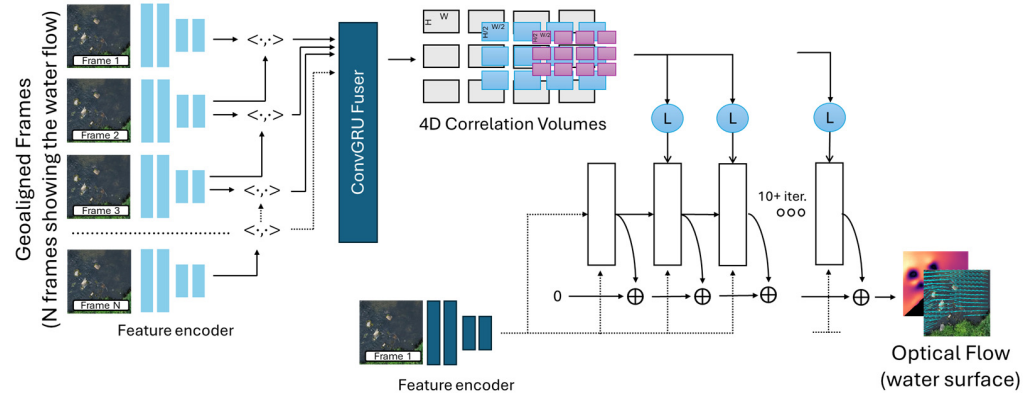


**Figure 2.** Example of the FEM simulation results represented by the velocity field (orange and red lines indicate boulders above and below water respectively; the green line marks the boundary between areas of different vegetation densities) (a), and the corresponding UAV image (b).

To address this limitation, several studies have investigated multi-frame optical flow methods. For example, VideoFlow proposes a tri-frame optical flow module combined with motion propagation across temporal segments, effectively extending estimation beyond two consecutive frames [37]. StreamFlow introduces an in-batch multi-frame pipeline to compute flow across several frames simultaneously, reducing computational redundancy and improving temporal coherence [38]. Other approaches use spatiotemporal learning. The spatiotemporal recurrent transformers for multi-frame optical flow estimation (SSTM) employ recurrent transformers to capture motion dependencies across longer video sequences [39]. In contrast, the Self-Teaching Multi-frame Unsupervised RAFT with Full-Image Warping (SMURF) method introduces temporal self-supervision into RAFT-like architectures to improve stability and generalization [40]. These developments demonstrate the feasibility of extending frame-to-frame optical flow into multi-frame paradigms.

In this research, we extend the RAFT architecture by proposing a multi-frame input variant, MF-RAFT, in which the model processes a sequence of  $N$  consecutive frames rather than a single frame pair. Each frame is first processed by the standard feature encoder, after which the extracted features from frame pairs are fused using a Fuse-GRU before correlation volume computation (Figure 3). This design integrates temporal information from multiple frames before dense correlations are calculated, thereby enhancing robustness in scenarios where single frame-pair estimation is unstable, such as low-texture river surfaces or variable illumination.

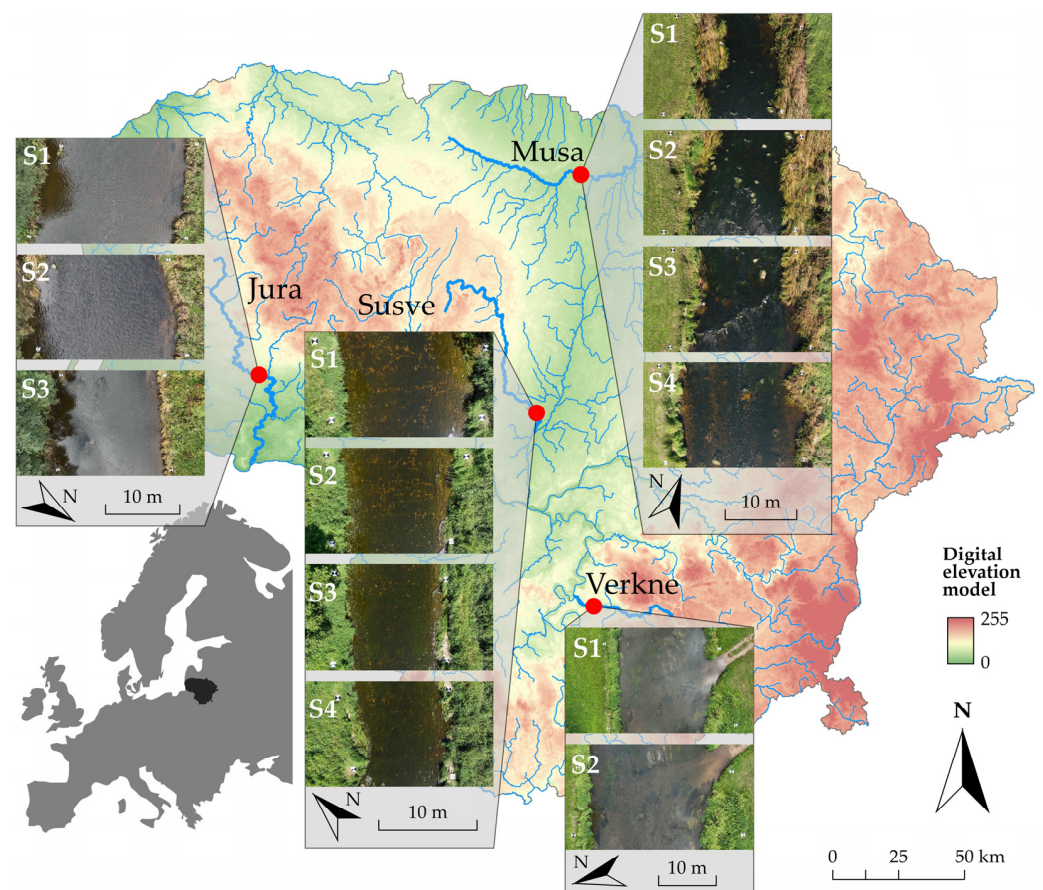
It should be noted that several alternative design choices are possible. The fusion block could, in principle, be applied after the correlation volume, allowing the model to operate on higher-level matching features. However, this approach would significantly increase computational cost due to the dimensionality of the correlation tensor. By placing Fuse-GRU before correlation computation, the model achieves a more favorable balance between temporal context integration and efficiency, enabling practical training and inference on UAV-derived river flow datasets. Importantly, the fused representations ultimately contribute to the estimation of displacement vectors  $(u, v)$  as defined in Section 2.1, which are subsequently converted into physical velocities  $V$ .



**Figure 3.** Modified RAFT architecture with multi-frame input and integrated Fuse-GRU.

2.4. Experimental Area and Data Collection

The study area comprises the same river segments examined in [31], located in Lithuania. The research focused on four shallow river sites with moderate flow velocities and minimal vegetative cover, offering well-distributed surface tracers visible in RGB imagery. These site characteristics supported both ground measurements and modeling using FEM as well as provided suitable conditions for UAV data acquisition. A more detailed description of the geographic context, hydraulic conditions, and field campaign logistics is available in [31]. The general layout of the study area and objects is shown in Figure 4.



**Figure 4.** The general study area and the river surfaces of the four selected rivers in Lithuania (in all cases, river flow oriented downward).

Flow velocity measurements and UAV-based video data were collected during dedicated field campaigns. Point measurements of ground truth flow velocity were taken

at 0.6 of the total depth (from the surface) using a Valeport Model 801 electromagnetic flowmeter (accuracy  $\pm 0.005$  m/s). Each point measurement was georeferenced using coordinates determined with a GeoMax Zenith 40 GNSS GPS receiver (accuracy  $\pm 0.015$  m). The duration of measurement campaign depends on the number of points under consideration and takes up to one day. UAV-based video frames were calibrated using ground control markers positioned along the riverbanks. The calibration procedure enabled correction of lens distortions, while georeferencing ensured that each pixel was associated with a fixed geographic location. This step established a constant spatial resolution (1 px = 1 cm) across all sequences, which is a prerequisite for the subsequent conversion of optical flow displacements into physical flow velocities as described in Section 2.1. Table 1 summarizes the main characteristics of the data collected at the experimental sites, including river name, measurement date, river discharge (RD), and the position and number of frame sequences (PFS). In the PFS column, the notations S1, S2, S3, and S4 indicate the locations along the river where UAV videos were recorded. The multiplier “ $\times 2$ ” denotes that two sequences were acquired at the same location with a  $180^\circ$  rotation, ensuring bidirectional coverage of the river reach.

**Table 1.** Summary of UAV-based river measurement campaigns, S1, S2, S3, and S4 indicate the locations along the river where UAV videos were recorded.

No.	River	Date	River Discharge (RD, m <sup>3</sup> /s)	Position Frame Sequences (PFS)
1	Můša	23 September 2024	0.506	S1 $\times$ 2 S2 $\times$ 2 S3 $\times$ 2 S4 $\times$ 2
2		11 July 2023	0.548	S1 $\times$ 2 S2 $\times$ 2
3	Šušvě	22 September 2023	0.491	S1 $\times$ 2 S2 $\times$ 2 S3 $\times$ 2 S4 $\times$ 2
4		26 June 2024	0.697	S1 $\times$ 2 S2 $\times$ 2 S3 $\times$ 2 S4 $\times$ 2
5	Verkné	17 July 2023	1.420	S1 $\times$ 2 S2 $\times$ 2
6	Jūra	13 July 2023	1.010	S1 S2 S3
7		14 September 2023	3.240	S1 $\times$ 2 S2 $\times$ 2 S3 $\times$ 2

### 2.5. Dataset Preparation

Based on the calibrated and georeferenced UAV sequences, a dataset was constructed for training and evaluation of the proposed AI model. Although the sequences were already spatially aligned, additional steps were taken to ensure that the dataset remained unbiased and suitable for machine learning. First,  $K$  random target points  $\{p_1, p_2, \dots, p_K\}$  were selected within the ground control point (GCP)-constrained zone, excluding riverbanks and shoreline areas to avoid non-flow features. Around each selected point, square image patches of varying side lengths were extracted, with the set of considered patch sizes denoted by:

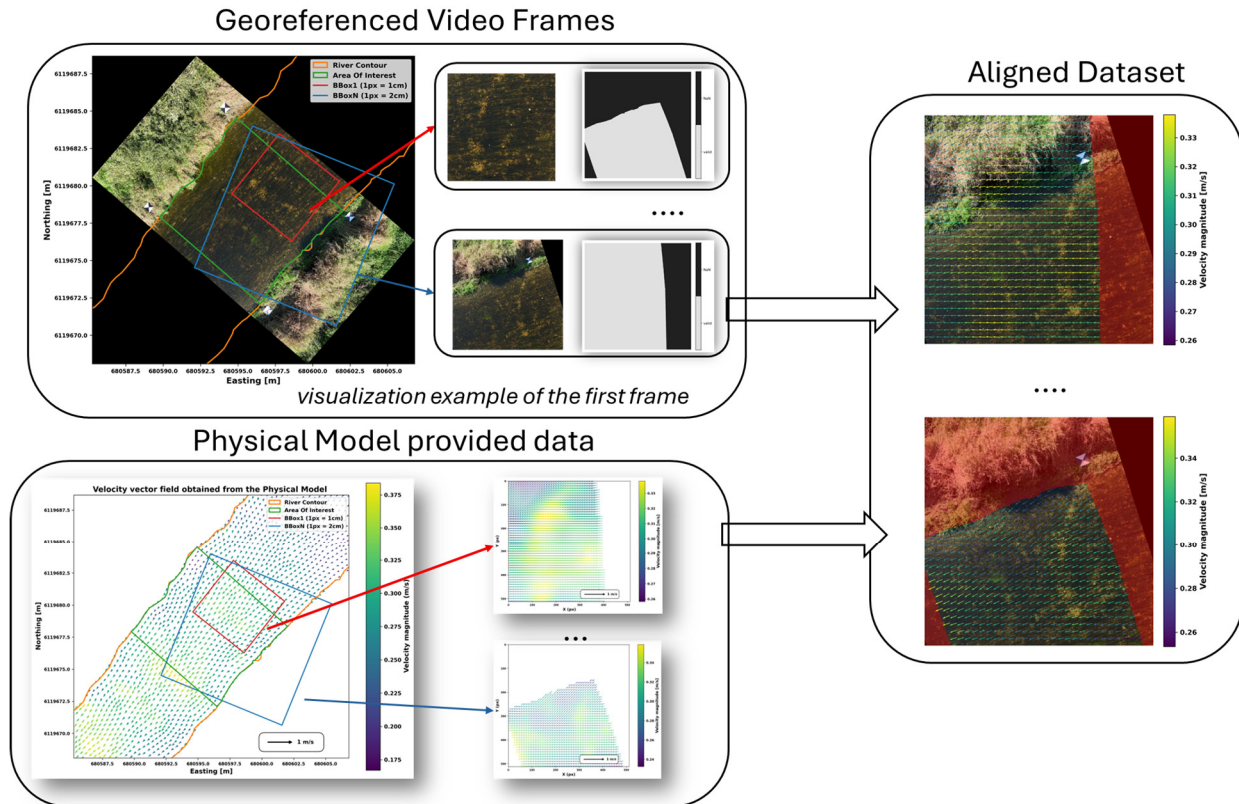
$$S = \{s_1, s_2, \dots, s_{|S|}\} \quad (4)$$

For each pair  $(p_k, s_j)$ , a corresponding patch was generated and then randomly rotated by an angle  $\theta \in [0^\circ, 90^\circ]$ , with additional random multiples of  $90^\circ$  applied to further diversify orientations. It is important to note that such rotations cannot be achieved by simply cropping the RGB imagery and vector field rasters and applying the same pixel-level transformation. A naive operation of this kind would break the georeferencing and, more critically, invalidate the physical meaning of the velocity vectors. While the georeferenced imagery is stored in a north-up orientation, rotation in pixel space would not preserve the correct coordinate reference system. Furthermore, the velocity components  $(u, v)$  are defined relative to the global axes; rotating only the image would leave the vector

field inconsistent with the rotated patch. Therefore, the vector field itself must be rotated component-wise by the same angle  $\theta$ , using a 2D rotation matrix:

$$\begin{bmatrix} u' \\ v' \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \tag{5}$$

Figure 5 schematically illustrates this process, showing that RGB imagery and the associated velocity vectors remain consistent after an arbitrary-angle rotation.



**Figure 5.** Schematic illustration of dataset patch generation. A random target point within the GCP-constrained zone is selected, a square patch of predefined size is extracted, and both the RGB imagery and the associated velocity vectors are rotated by an arbitrary angle  $\theta$ . This procedure ensures that the image content and flow fields remain physically consistent after rotation.

To introduce angular variability while preserving numerical stability, the initial random rotation was restricted to the interval  $[0^\circ, 90^\circ]$ . This was then combined with an additional one to three random multiples of  $90^\circ$ , effectively distributing patch orientations across the full  $[0^\circ, 360^\circ]$  range. This hybrid strategy enhanced augmentation diversity and ensured that image rotations remained computationally stable and geospatially consistent. Figure 6 illustrates an example in which a patch, first rotated within  $[0^\circ, 90^\circ]$ , is subsequently rotated by an additional  $180^\circ$ , demonstrating that both the imagery and velocity fields remain properly aligned.

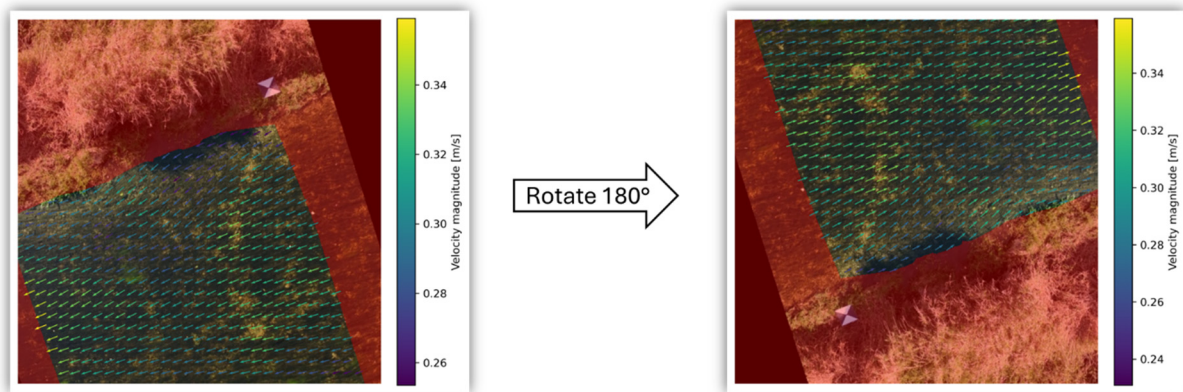
In addition to spatial diversity, temporal variability was incorporated by introducing multiple frame strides. We define a set

$$D = \left\{ \Delta_{\text{frames}}^{(1)}, \Delta_{\text{frames}}^{(2)}, \dots, \Delta_{\text{frames}}^{(|D|)} \right\} \tag{6}$$

where each element specifies the number of frames skipped between consecutive samples. For a given stride  $\Delta_{\text{frames}}^{(d)} \in D$ , the corresponding physical time interval  $\Delta T^{(d)}$  is expressed as:

$$\Delta T^{(d)} = \frac{\Delta_{\text{frames}}^{(d)}}{\text{FPS}} \quad (7)$$

where FPS refers to frames per second.



**Figure 6.** Example of a rotated patch after an additional 180-degree rotation. The RGB imagery and velocity vectors remain aligned, demonstrating that the applied transformation preserves consistency between spatial appearance and flow direction.

This construction ensures that the dataset reflects river dynamics across multiple temporal scales: small values of  $\Delta_{\text{frames}}$  capture short-term fluctuations, while larger values account for slower, large-scale motions. Moreover, by including several stride values, the resulting dataset is not implicitly tied to a single acquisition frame rate, thereby improving robustness when applying the trained model to UAV surveys or video sequences acquired at different FPS. Such temporal augmentation thus complements spatial augmentation, broadening the variability of training instances while preserving the physical interpretability of the data.

Finally, the total number of dataset instances generated through this procedure can be expressed as:

$$\Delta N_{\text{instances}} = K \cdot |S| \cdot |D| \quad (8)$$

Here,  $K$  denotes the number of randomly sampled target points,  $|S|$  the number of distinct patch sizes, and  $|D|$  the number of temporal stride values. This strategy yielded a dataset in which river segments were represented without directional bias (i.e., not constrained to a consistent leftward or rightward flow). At the same time, each instance remained explicitly linked to its underlying river segment, enabling a principled partitioning of training and testing subsets according to distinct segments. Such a design prevents data leakage between training and test sets and ensures that model evaluation reflects true generalization performance under independent conditions.

## 2.6. Dataset Configuration and Partitioning

The UAV-based video data (Section 2.4) were used to construct four complementary datasets for model training and evaluation. Independence between training and validation subsets was ensured by either separating entire measurement campaigns (temporal independence) or partitioning spatial zones within the same campaign (spatial independence).

Each video sequence was georeferenced and then converted into dataset instances following the procedure described in Section 2.5 (Equation (8)). In this study, three temporal stride values were applied,  $D = \{4, 5, 6\}$  corresponding to  $\Delta T$  intervals defined by the UAV acquisition frame rate. A total of  $K = 180$  randomly selected target points were generated within the GCP-constrained zone, excluding non-flow regions such as riverbanks or shorelines. For each target point, six spatial resolutions were considered,

$S = \{0.010, 0.012, 0.014, 0.016, 0.018, 0.020\}$  m/px, representing ground sampling distances at 0.002 m. increments. According to Equation (8), this configuration resulted in a total of  $N = K \times |S| \times |D| = 180 \times 6 \times 3 = 3240$  independent samples per dataset, ensuring both spatial and temporal variability across multiple scales.

For river-specific datasets, one measurement campaign (Table 1) was completely withheld for validation, while all remaining measurements from the same and other rivers were used for training. Each position frame sequence (PFS) could include multiple sequences (denoted  $\times 2$ ) corresponding to videos acquired with a  $180^\circ$  rotation at the same location; each sequence was treated independently during dataset construction. The resulting dataset configuration and partitioning are summarized in Table 2, which is directly derived from the measurement campaigns presented in Table 1. The number of measured points in Verkné river was not high compared to the other river segments, therefore it was used for training only to increase the diversity of training data.

**Table 2.** Partitioning of training and validation subsets.

Dataset	Training Rivers (RD)	Validation River (RD)3	Train Sequences	Val Sequences	Train Instances	Val Instances
Jūra	Jūra (1.01), Mūša (0.548; 0.506), Šušvė (0.491; 0.697), Verknė (1.42)	Jūra (3.24)	35	6	113,400	19,440
Mūša	Mūša (0.548), Jūra (1.01; 3.24), Šušvė (0.491; 0.697), Verknė (1.42)	Mūša (0.506)	33	8	106,920	25,920
Šušvė	Šušvė (0.491), Mūša (0.548; 0.506), Jūra (1.01; 3.24), Verknė (1.42)	Šušvė (0.697)	33	8	106,920	25,920

The partitioning strategy (Table 2) ensured that validation was conducted on measurements not seen during training or on independent spatial zones within the same campaign. This design prevents overlap between training and validation subsets and allows model performance to be assessed under conditions approximating real deployment scenarios across individual rivers.

## 2.7. Evaluation Metrics of Results

The performance of the proposed method was assessed using complementary measures: training loss, endpoint error (EPE), average angular error (AAE), and flow outlier rate (FL). Together, these metrics capture both the convergence behavior during optimization and the quantitative accuracy of the resulting optical flow fields.

The training loss follows the standard RAFT formulation proposed by Teed and Deng [17] adapted to include valid-pixel masking. It is a multi-iteration sequence loss that computes the pixel-wise Euclidean distance between the predicted and reference flow fields, weighted by an exponentially decaying factor ( $\gamma = 0.8$ ) across recurrent refinement steps. Only pixels within valid flow masks are included in the computation. This loss design encourages gradual convergence while maintaining flow consistency in spatially coherent regions.

The endpoint error (EPE) measures the Euclidean distance between the predicted and reference flow vectors in pixel space. For a predicted vector  $(u_p, v_p)$  and a reference vector  $(u_r, v_r)$ , it is defined as:

$$\text{EPE}_{\text{px}} = \sqrt{(u_p - u_r)^2 + (v_p - v_r)^2} \quad (9)$$

This metric directly expresses the discrepancy in terms of pixel displacements. To allow interpretation in physical velocity units,  $EPE_{\text{px}}$  values were converted to meters per second as follows:

$$EPE_{\text{phys}} = \frac{EPE_{\text{px}} \cdot R}{\Delta T} \quad (10)$$

Here,  $R$  denotes the spatial resolution [m/px] determined by the selected zoom ratio, and  $\Delta T$  is the physical time interval between frames defined by the chosen temporal stride. Since both  $R$  and  $\Delta T$  vary across dataset configurations, the physical interpretation of  $EPE_{\text{phys}}$  is not constant but depends on the experimental setup. For instance, an error of 1 px at a resolution of  $R = 0.012$  m/px and a temporal stride of  $\Delta T = 0.17$  s corresponds approximately to a velocity error of 0.07 m/s.

The average angular error (AAE) complements magnitude-based evaluation by focusing on directional deviations. It is computed as:

$$AAE = \arccos \left( \frac{u_p u_r + v_p v_r + 1}{\sqrt{(u_p^2 + v_p^2 + 1)(u_r^2 + v_r^2 + 1)}} \right) \quad (11)$$

Here, the additional term +1 in both the numerator and denominator ensures numerical stability for small vector magnitudes. The result, expressed in degrees, reflects the orientation consistency of the velocity field between prediction and reference.

By combining the loss-based convergence criterion with EPE and AAE, the evaluation framework captures both the learning dynamics and the resulting flow accuracy in terms of magnitude, direction, and spatial robustness across various flow regimes.

### 3. Results

#### 3.1. Model Training

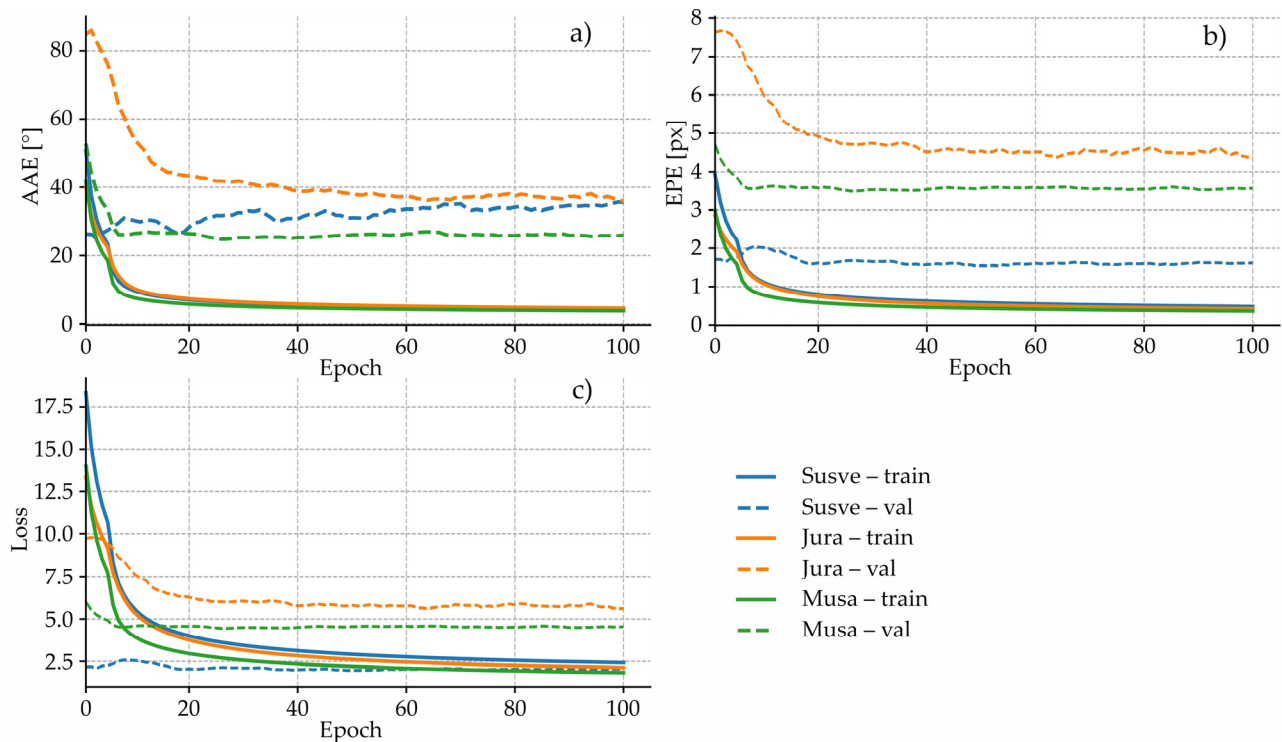
Three models were trained independently, each corresponding to one of the dataset configurations described in Section 2.6 (Jūra, Mūša, and Šušvė). All training was conducted on an HPC cluster equipped with NVIDIA H100 GPUs, with 64 GB of system memory and 8 CPU cores allocated for data loading and preprocessing. The implementation used the PyTorch 2.6 framework.

Before training, all samples were resized and arranged to fixed dimensions ( $512 \times 512$ ), ensuring consistency for images, ground-truth flow, and masks across the dataset. Each training instance used sequences of ten consecutive frames per sample, with identical data loading parameters to maintain comparability between datasets. Windows were selected deterministically from the start of each sequence, and short sequences were right-padded to reach the full temporal length.

Optimization fused RAFT-style recurrent refinement scheme with 12 update iterations per forward pass. The training objective employed a masked RAFT sequence loss [17] with exponentially decaying weights across iterations ( $\gamma = 0.8$ ), applied only to valid pixels. Validation loss was computed as a masked L1 difference between the upsampled prediction and the reference flow. In addition to the loss, three evaluation metrics were monitored during training: endpoint error (EPE), average angular error (AAE), and the fraction of outlier pixels (FL-error), defined by absolute (3 px) and relative (5%) thresholds.

All models were trained for 100 epochs using the AdamW optimizer with a fixed learning rate of  $1 \times 10^{-4}$ . Gradient clipping with a global norm of 1.0 was applied to stabilize optimization. The batch size was adjusted according to the available GPU memory capacity, ensuring full utilization of the H100 hardware without memory overflow. One training epoch takes approximately 90 min.

The training and validation loss curves, along with the evolution of EPE, AAE, and FL-error metrics across epochs, are shown in Figure 7. These plots illustrate the convergence behavior of all three models and enable a comparative evaluation of training stability and generalization performance.



**Figure 7.** Training progress of the four RAFTWAT models: (a) Endpoint error (EPE); (b) Average angular error (AAE); (c) Training and validation loss as functions of training epochs for Jūra, Mūša, and Šušvė. Solid lines denote training results, and dashed lines denote validation results.

As shown in Figure 7a–c, all models converged rapidly within the first 20–30 epochs, followed by stable metric evolution for the remaining epochs. Both EPE and AAE decreased monotonically, indicating consistent improvements in flow accuracy and directional stability. Among the configurations, the river-specific models exhibited similar convergence behavior, confirming that the training framework generalized well across different river environments.

It is important to note that these results are expressed in pixel space, representing relative displacements rather than absolute physical velocities. As both the spatial resolution ( $R$ ) and temporal stride ( $\Delta T$ ) vary across dataset configurations, direct comparison in physical units (m/s) is not straightforward. Conversion to physical flow magnitudes depends on these parameters and is discussed in the following section, where velocity predictions are quantitatively analyzed under real-world scaling conditions.

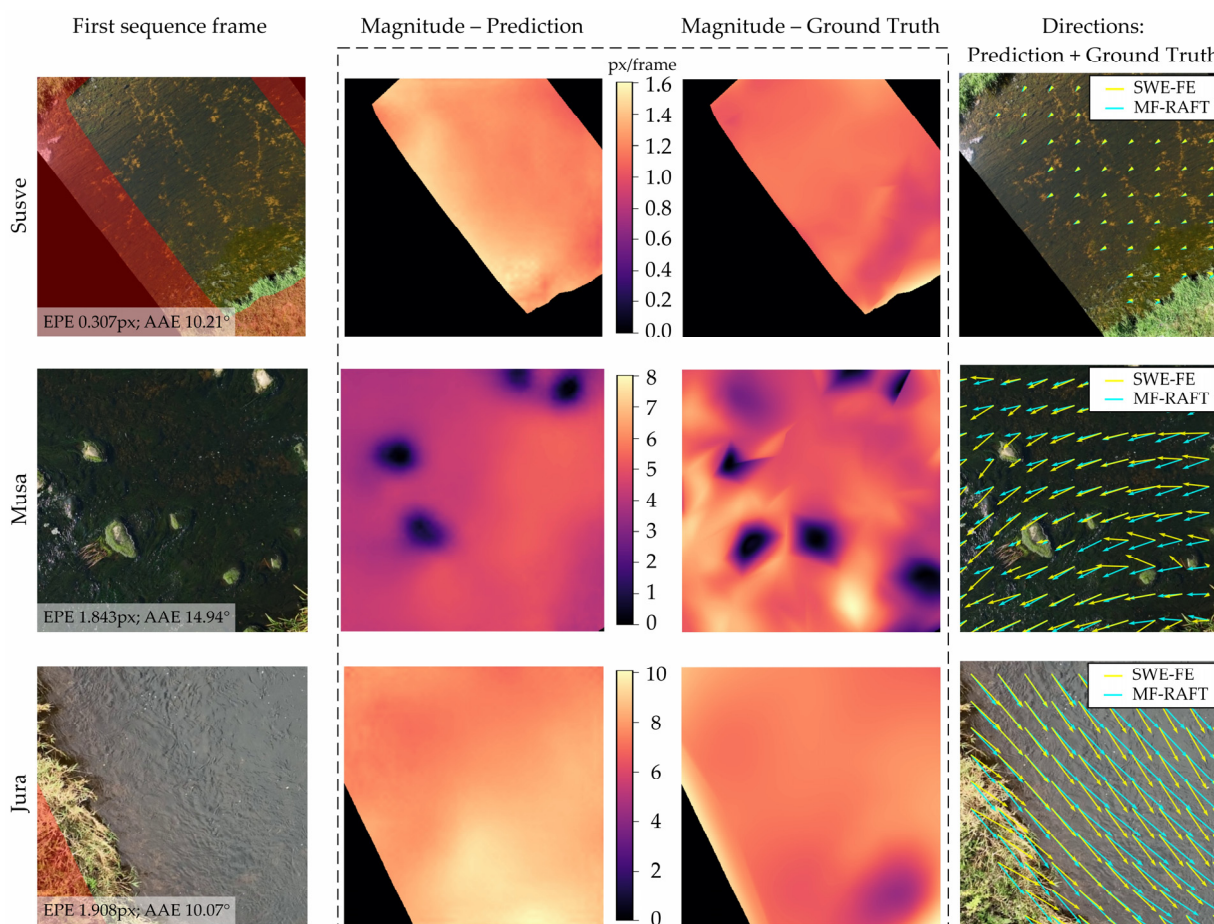
Furthermore, the reference flow fields used for training and validation were derived from SWE-FE model, which inherently approximate the true hydrodynamic conditions. While obtained flow fields provide physically consistent supervision, they may still contain systematic deviations due to boundary simplifications, mesh resolution, or numerical diffusion. Consequently, the reported error metrics reflect the model's consistency with the SWE-FE reference rather than the absolute physical accuracy of the flow.

Despite these limitations, the observed trends provide clear evidence that the learning framework successfully captures the dominant kinematic and directional flow characteristics across various river environments. The resulting river-specific models demonstrate

strong generalization capability and serve as a robust foundation for the detailed physical evaluation presented in Section 3.2.

### 3.2. Preview of Visual Results

To provide an intuitive understanding of the model's performance, Figure 8 presents representative qualitative examples from three UAV test sites. Each case corresponds to a distinct river reach with different flow regimes and surface characteristics. For each site, the first valid frame, predicted velocity magnitude, SWE-FE reference field, and directional vector comparison are shown.



**Figure 8.** Qualitative comparison of predicted and reference velocity fields for selected UAV test sequences in Šušvė, Mūša, and Jūra. Error metrics (EPE and AAE) are reported for each sequence.

In the Šušvė case, the image resolution is 0.02 m/px, resulting in generally shorter velocity vectors, as each pixel represents a larger physical area and overall flow velocities are lower. Nevertheless, the model successfully reconstructs the full river cross-section and consistently captures the overall flow pattern in agreement with the SWE-FE reference. At the Mūša site, several submerged stones are visible. The SWE-FE model retains near-zero velocities around these obstacles, depending on how precisely the solid boundary was delineated. The proposed model reproduces most of these low-velocity regions but tends to slightly smooth smaller obstacles—a likely consequence of both limited texture and small-scale inaccuracies in the SWE-FE reference contours. In the Jūra sequence, the riverbanks partially overlap in the region of interest, and some vegetation occludes the edges. The visible boundary vectors in this area correspond to zones also present in the training set, which may contribute to better local consistency. Overall, the examples show that at low flow magnitudes, the model achieves mean endpoint errors (EPE) of around 0.3 px, while

for higher velocities and coarser resolutions, EPE values reach 1.8–1.9 px. Directional discrepancies remain limited, typically within 10–15°, confirming that the Fuse-GRU architecture maintains good directional stability even under challenging optical conditions.

### 3.3. Comprehensive Performance Analysis

To comprehensively assess the robustness and behavior of the proposed model under different conditions, a quantitative analysis was conducted across multiple configurations of spatial and temporal parameters. This section examines how variations in spatial resolution, temporal stride, and valid-pixel coverage affect the accuracy and stability of the predicted flow fields. The primary evaluation metric is the endpoint error (EPE), which quantifies the Euclidean distance between predicted and reference flow vectors. For interpretability in physical terms, EPE values were expressed as  $EPE_{phys}$  (Section 2.7), allowing direct comparison of the actual flow velocity discrepancies between model predictions and reference data.

#### 3.3.1. Analysis Across Spatial Resolutions

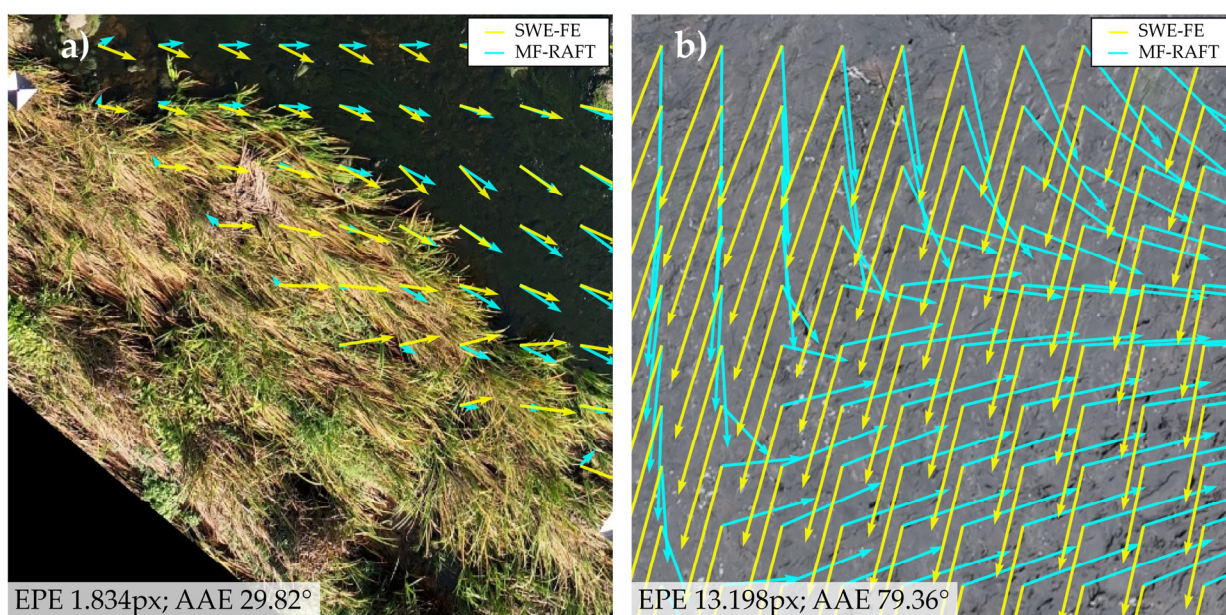
To evaluate how image scale influences the accuracy of predicted flow fields, a quantitative analysis was performed at multiple spatial resolutions for all three datasets. Table 3 shows the validation performance of the proposed model at resolutions from 0.010 m/px to 0.020 m/px, including the mean and standard deviation of the physical endpoint error ( $EPE_{phys}$ ), mean absolute percentage error (MAPE), angular error (AAE), and Root Mean Squared Error (RMSE). This comparison enables investigation of how changes in pixel size, and thus the level of visible texture and physical displacement per pixel, influence the accuracy and stability of flow magnitude and direction estimation.

**Table 3.** Validation performance across datasets and spatial resolutions (mean ± standard deviation) if all data points are used or if 15% of the lowest velocity values are filtered out (Šušvė: >0.170 m/s; Mūša: >0.271 m/s; Jūra: >0.403 m/s).

Dataset	Resolution (m/px)	Full Dataset				15% Lowest Velocity Values Filtered Out			
		$EPE_{phys}$ (m/s)	MAPE (%)	AAE (°)	RMSE (m/s)	$EPE_{phys}$ (m/s)	MAPE (%)	AAE (°)	RMSE (m/s)
Jūra	0.010	0.438 ± 0.116	32.2 ± 22.7	45.06 ± 13.11	0.172	0.434 ± 0.109	22.4 ± 11.4	44.15 ± 12.11	0.152
	0.012	0.405 ± 0.122	36.0 ± 27.0	41.01 ± 13.29	0.183	0.398 ± 0.112	22.4 ± 12.1	40.69 ± 11.93	0.152
	0.014	0.361 ± 0.124	34.6 ± 29.3	35.56 ± 13.59	0.179	0.352 ± 0.114	22.0 ± 12.2	34.60 ± 12.02	0.152
	0.016	0.348 ± 0.129	32.9 ± 29.0	33.90 ± 14.26	0.180	0.341 ± 0.119	22.3 ± 12.6	32.90 ± 12.76	0.157
	0.018	0.313 ± 0.128	29.1 ± 27.2	30.26 ± 14.56	0.171	0.311 ± 0.116	22.2 ± 12.3	29.43 ± 12.50	0.157
	0.020	0.309 ± 0.129	30.7 ± 28.9	28.48 ± 14.21	0.180	0.307 ± 0.115	23.4 ± 12.9	27.53 ± 11.90	0.165
Mūša	0.010	0.319 ± 0.103	53.3 ± 109.3	34.35 ± 14.17	0.250	0.333 ± 0.096	44.9 ± 12.3	34.10 ± 12.48	0.260
	0.012	0.305 ± 0.109	53.6 ± 121.8	32.67 ± 14.17	0.249	0.321 ± 0.101	44.8 ± 12.8	32.47 ± 12.35	0.260
	0.014	0.299 ± 0.115	53.6 ± 125.7	29.85 ± 13.74	0.257	0.317 ± 0.107	45.3 ± 13.4	29.85 ± 12.18	0.268
	0.016	0.291 ± 0.119	54.6 ± 139.3	28.32 ± 13.57	0.256	0.309 ± 0.111	45.1 ± 13.8	28.12 ± 12.02	0.267
	0.018	0.292 ± 0.124	55.7 ± 152.1	28.10 ± 13.36	0.260	0.312 ± 0.114	45.4 ± 13.9	28.00 ± 11.72	0.273
	0.020	0.286 ± 0.129	56.2 ± 160.5	27.44 ± 13.76	0.260	0.309 ± 0.119	45.7 ± 13.9	27.35 ± 12.10	0.274
Šušvė	0.010	0.140 ± 0.043	31.5 ± 14.0	33.66 ± 10.90	0.079	0.144 ± 0.042	32.1 ± 13.3	34.28 ± 10.71	0.081
	0.012	0.148 ± 0.048	31.3 ± 15.4	35.31 ± 11.67	0.080	0.152 ± 0.048	31.7 ± 14.6	35.78 ± 11.75	0.081
	0.014	0.142 ± 0.054	30.8 ± 16.8	32.36 ± 12.72	0.078	0.145 ± 0.054	30.6 ± 15.8	32.74 ± 12.65	0.079
	0.016	0.144 ± 0.058	30.0 ± 17.0	32.79 ± 13.33	0.077	0.147 ± 0.057	29.8 ± 16.0	33.02 ± 12.98	0.078
	0.018	0.147 ± 0.063	30.7 ± 17.9	32.81 ± 14.28	0.080	0.151 ± 0.061	30.4 ± 16.8	33.27 ± 13.55	0.080
	0.020	0.150 ± 0.069	31.3 ± 18.5	33.09 ± 15.17	0.081	0.156 ± 0.065	31.1 ± 17.4	33.84 ± 13.79	0.082

As shown in Table 3, a gradual improvement in accuracy is observed as the spatial resolution becomes coarser, particularly in the Jūra dataset, where angular errors decrease from 45° at 0.010 m/px to 28° at 0.020 m/px. A similar trend is observed for  $EPE_{phys}$  whereas MAPE remains relatively stable. For Mūša, the same overall tendency is visible, though percentage errors remain higher due to local turbulence and specular reflections that

distort optical cues. In contrast, Šušvė shows minimal variation across scales, indicating that smoother, low-velocity flow conditions make the model less sensitive to pixel size. When the resolution decreases, a larger portion of the river scene fits within a single frame, providing richer spatial context for motion interpretation and facilitating smoother, more generalized predictions. Conversely, at higher resolutions, the model observes smaller, localized areas where lighting variations, ripples, or vegetation motion may dominate the signal, occasionally leading to unstable estimates. To further illustrate these effects, representative examples of failure cases are shown in Figure 9. After filtering out 15% of the lowest velocity values for each case,  $EPE_{phys}$ , AAE, and RMSE values do not change significantly. MAPE values drop for the Jūra and Mūša rivers, demonstrating that the model tends to generate proportionally higher errors for low velocities. However, in the Šušvė case, the change in MAPE values is not significant, showing that the model works robustly with respect to velocity in low-complexity river segments without boulders or ripples.



**Figure 9.** Examples of challenging cases with reduced prediction accuracy: (a) Mūša—local disturbances from vegetation and surface ripples; (b) Jūra—high turbulence, low texture, and reflection effects causing angular deviations between prediction and ground truth.

In the Mūša case (Figure 9a), dense vegetation and fine-scale surface ripples locally disturb the predicted flow vectors, although the overall direction remains consistent. In the Jūra example (Figure 9b), a limited field of view, high turbulence, and weak texture information result in large angular deviations between the predicted and reference vectors. These examples demonstrate that, while the model performs robustly in most scenarios, local inconsistencies can arise when visual patterns are ambiguous or physically unstable—particularly in areas where vegetation motion, wind-driven waves, or reflections dominate the observed surface dynamics.

### 3.3.2. Analysis Across Temporal Strides

To investigate the influence of temporal spacing between frames, the model was evaluated using different temporal strides ( $\Delta_{frames} = 4, 5, \text{ and } 6$ ) across all datasets. Table 4 summarizes the corresponding validation metrics, including the physical endpoint error ( $EPE_{phys}$ ), mean absolute percentage error (MAPE), and angular error (AAE). This analysis provides insight into how temporal separation affects the model's ability to maintain flow consistency and accurately track displacements over time.

**Table 4.** Validation performance across datasets and temporal strides (mean  $\pm$  standard deviation) if all data points are used or if 15% of the lowest velocity values are filtered out (Šušvė:  $>0.170$  m/s; Mūša:  $>0.271$  m/s; Jūra:  $>0.403$  m/s).

Dataset	$\Delta_{frames}$	Full Dataset				15% Lowest Velocity Values Filtered Out			
		$EPE_{phys}$ (m/s)	MAPE (%)	AAE (°)	RMSE (m/s)	$EPE_{phys}$ (m/s)	MAPE (%)	AAE (°)	RMSE (m/s)
Jūra	4	0.335 $\pm$ 0.126	33.8 $\pm$ 28.9	31.01 $\pm$ 13.45	0.180	0.326 $\pm$ 0.115	22.5 $\pm$ 12.7	29.98 $\pm$ 11.84	0.156
	5	0.361 $\pm$ 0.124	32.6 $\pm$ 27.5	35.43 $\pm$ 13.84	0.177	0.355 $\pm$ 0.114	22.3 $\pm$ 12.3	34.62 $\pm$ 12.24	0.155
	6	0.391 $\pm$ 0.123	31.3 $\pm$ 25.6	40.69 $\pm$ 14.22	0.176	0.389 $\pm$ 0.114	22.6 $\pm$ 11.8	39.93 $\pm$ 12.54	0.158
Mūša	4	0.298 $\pm$ 0.118	53.8 $\pm$ 146.6	30.66 $\pm$ 13.80	0.246	0.314 $\pm$ 0.110	43.0 $\pm$ 14.0	30.62 $\pm$ 12.26	0.255
	5	0.297 $\pm$ 0.116	54.5 $\pm$ 134.4	29.95 $\pm$ 13.76	0.255	0.315 $\pm$ 0.107	45.2 $\pm$ 13.3	29.80 $\pm$ 12.09	0.267
	6	0.301 $\pm$ 0.115	55.3 $\pm$ 123.3	29.75 $\pm$ 13.82	0.265	0.321 $\pm$ 0.106	47.3 $\pm$ 12.7	29.53 $\pm$ 12.08	0.278
Šušvė	4	0.157 $\pm$ 0.063	32.5 $\pm$ 19.3	33.22 $\pm$ 12.92	0.084	0.160 $\pm$ 0.061	31.8 $\pm$ 18.1	33.65 $\pm$ 12.52	0.084
	5	0.142 $\pm$ 0.055	29.4 $\pm$ 16.1	33.11 $\pm$ 13.05	0.076	0.146 $\pm$ 0.054	29.6 $\pm$ 15.2	33.61 $\pm$ 12.60	0.077
	6	0.138 $\pm$ 0.050	30.9 $\pm$ 14.3	33.69 $\pm$ 13.07	0.078	0.142 $\pm$ 0.048	31.4 $\pm$ 13.7	34.21 $\pm$ 12.60	0.08

As shown in Table 4, the effect of temporal stride varies with river conditions. For the Jūra dataset, increasing  $\Delta_{frames}$  from 4 to 6 leads to progressively higher endpoint and angular errors (from 0.335 m/s to 0.391 m/s and from 31° to 41°, respectively) indicating that larger time gaps reduce temporal coherence and make motion correspondence less stable. This behavior is typical of faster and more turbulent flows, where displacements between frames can become too large for reliable optical flow matching. In contrast, Mūša and Šušvė exhibit relatively stable or slightly improved accuracy at larger strides. For Šušvė, where surface motion is smoother and dominated by slow laminar flow, increasing  $\Delta_{frames}$  from 4 to 6 reduces  $EPE_{phys}$  from 0.157 m/s to 0.138 m/s. This suggests that in low-velocity regimes, a longer temporal gap can enhance the detectability of meaningful motion by emphasizing displacements above the sensor's noise threshold. If the lowest 15% velocities are filtered out, the trends remain the same, with significantly smaller MAPE values for the Jūra and Mūša rivers. Overall, these results demonstrate that the optimal temporal stride depends strongly on the flow regime: shorter intervals are preferable for rapid or turbulent motion, while slower flows can tolerate longer separations without significant loss of accuracy.

### 3.3.3. Effect of Valid-Pixel Coverage

Since training and evaluation used flow masks that excluded regions with unreliable or missing motion information, an additional analysis was conducted to quantify the influence of visible (valid) area coverage on model performance. The valid-pixel ratio represents the proportion of pixels included in the loss and metric computation, averaged across validation subsets. Lower coverage indicates that a larger portion of the image was masked out, providing less spatial context for motion estimation.

As shown in Table 5, there is a clear relationship between coverage ratio and model accuracy. In all datasets, higher valid-pixel coverage consistently leads to lower endpoint and angular errors, confirming that richer spatial context improves both magnitude and directional estimates. For example, in the Jūra dataset, increasing valid coverage from below 60% to above 80% reduces  $EPE_{phys}$  from 0.400 m/s to 0.340 m/s and AAE from 42° to 32°. A similar tendency is evident for Šušvė, where fully visible frames ( $>80\%$ ) achieve the lowest errors ( $EPE_{phys} \approx 0.097$  m/s,  $AAE \approx 20^\circ$ ). In contrast, the Mūša dataset shows minor changes in  $EPE_{phys}$  but a clear decrease in angular error with higher coverage, suggesting that when a larger portion of the river surface is visible, the model can better constrain flow directions even if velocity magnitude errors remain similar. This effect likely arises because vegetation and shadowed regions near the banks were excluded in low-coverage samples, reducing the available texture for reliable optical tracking. In the

lowest 15% of velocities are filtered out, the trends remain the same, with significantly smaller MAPE values for the Jūra and Mūša rivers.

**Table 5.** Validation performance across datasets and valid-pixel coverage groups (mean ± standard deviation). Reported metrics: physical endpoint error ( $EPE_{phys}$ ), mean absolute percentage error (MAPE), angular error (AAE), and root mean squared error (RMSE) when all data points are used or when the 15% lowest velocity values are filtered out (Šušvė: >0.170 m/s; Mūša: >0.271 m/s; Jūra: >0.403 m/s).

Dataset	Valid Pixels	Total Instances	Full Dataset				15% Lowest Velocity Values Filtered Out			
			$EPE_{phys}$ (m/s)	MAPE (%)	AAE (°)	RMSE (m/s)	$EPE_{phys}$ (m/s)	MAPE (%)	AAE (°)	RMSE (m/s)
Jūra	<60%	5544	0.400 ± 0.123	40.5 ± 25.7	42.12 ± 13.83	0.201	0.399 ± 0.109	27.7 ± 13.8	43.41 ± 12.37	0.173
	60–80%	4965	0.362 ± 0.131	34.9 ± 28.2	35.51 ± 14.75	0.187	0.290 ± 0.117	19.8 ± 11.9	25.53 ± 12.48	0.148
	>80%	8931	0.340 ± 0.122	26.4 ± 27.9	31.85 ± 13.33	0.158	0.353 ± 0.118	17.7 ± 10.6	30.88 ± 11.77	0.140
Mūša	<60%	9843	0.298 ± 0.113	54.8 ± 148.6	34.77 ± 14.85	0.232	0.299 ± 0.098	43.9 ± 13.1	33.15 ± 12.48	0.235
	60–80%	6987	0.299 ± 0.115	54.4 ± 143.4	29.66 ± 13.61	0.254	0.338 ± 0.123	46.7 ± 13.5	28.80 ± 12.09	0.298
	>80%	9090	0.299 ± 0.122	54.3 ± 113.2	25.45 ± 12.79	0.281	0.334 ± 0.114	46.5 ± 13.7	23.80 ± 11.39	0.307
Šušvė	<60%	14,604	0.171 ± 0.063	33.4 ± 18.5	40.02 ± 14.28	0.086	0.163 ± 0.056	32.6 ± 16.4	38.13 ± 13.03	0.082
	60–80%	7533	0.120 ± 0.050	27.8 ± 15.0	27.04 ± 12.22	0.072	0.127 ± 0.054	27.7 ± 14.8	26.41 ± 12.38	0.078
	>80%	3783	0.097 ± 0.039	27.5 ± 12.4	20.08 ± 9.69	0.069	0.105 ± 0.042	26.2 ± 12.5	20.15 ± 9.90	0.073

These results emphasize that adequate surface visibility is crucial for maintaining both quantitative and directional accuracy. Even with temporal fusion, insufficient valid-pixel coverage limits the model’s ability to infer coherent flow structures, highlighting the importance of high-quality imagery and consistent illumination during UAV data acquisition.

### 3.4. Independent Validation Using Field Measurements

To ensure that the MF-RAFT velocity estimates are physically consistent and generalize beyond the training data, an independent validation was carried out using in situ flow measurements collected at two river segments, Mūša (0.506) and Šušvė (0.697). Notably, the MF-RAFT inference used models trained without these specific segments, ensuring an unbiased evaluation (see Section 2.6). For comparison, velocity fields were processed with the temporal stride fixed at  $\Delta_{frames} = 5$ , and only cases with at least 60% visible flow surface were included across all spatial resolutions, as this threshold provides sufficient spatial context for reliable motion estimation and minimizes artifacts from occluded or low-texture regions. The resulting dataset enabled a three-way comparison among the measured velocities, the SWE-FE physical model, and the MF-RAFT predictions, providing an independent assessment of model consistency and accuracy.

The results in Table 6 shows a strong correspondence between the MF-RAFT-derived velocities and the field measurements, confirming that the proposed approach captures the main flow structures with physically meaningful accuracy. In the Šušvė segment (discharge 0.697 m<sup>3</sup>/s), the physical endpoint error between the measured and physics-based velocities was as low as 0.08 m/s, indicating excellent agreement between the numerical model and direct observations. Compared to the MF-RAFT predictions, both the measured and derived using SWE-FE velocities remained closely aligned, with  $EPE_{phys}$  values of 0.17 m/s and 0.15 m/s, respectively. The angular error (AAE) was also small—4.6° between measured and simulated velocities, and below 10° in comparisons involving the MF-RAFT estimates—showing that the physical model reproduces realistic directional patterns that the MF-RAFT successfully captures. These results suggest that the physical flow behavior in this low-turbulence reach is well approximated by both the SWE-FE model and the proposed inference method.

**Table 6.** Comparison of flow velocity estimation accuracy between measured, SWE-FE, and MF-RAFT-derived results (mean  $\pm$  standard deviation). Reported metrics: physical endpoint error ( $EPE_{phys}$ ), and angular error (AAE).

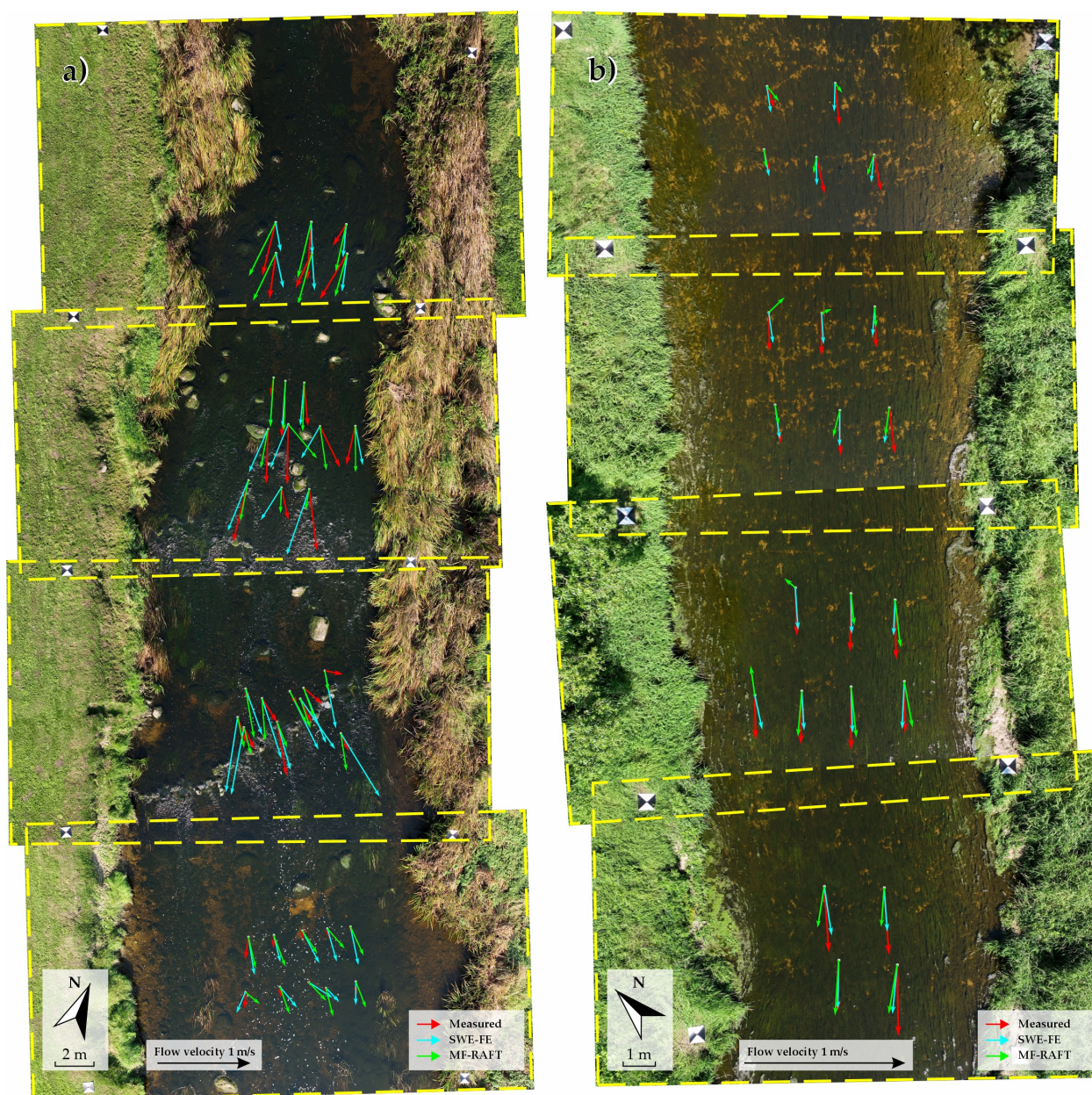
River (Physically Unique Points)	Comparison Types	$EPE_{phys}$ (m/s)	AAE ( $^{\circ}$ )
Mūša (300)	Measured vs. MF-RAFT	$0.279 \pm 0.242$	14.75
	SWE-FE vs. Measured	$0.311 \pm 0.191$	15.45
	SWE-FE vs. MF-RAFT	$0.364 \pm 0.244$	18.58
Šušvė (94)	Measured vs. MF-RAFT	$0.170 \pm 0.141$	9.17
	SWE-FE vs. Measured	$0.086 \pm 0.041$	4.62
	SWE-FE vs. MF-RAFT	$0.155 \pm 0.152$	8.4

In contrast, the Mūša segment (discharge  $0.506 \text{ m}^3/\text{s}$ ) exhibits a more complex flow regime, characterized by submerged stones, shallow zones, and turbulence induced by vegetation. Here, the discrepancies are more pronounced, with  $EPE_{phys}$  values of  $0.28 \text{ m/s}$  for the measured versus MF-RAFT comparison,  $0.31 \text{ m/s}$  for physics versus measured, and  $0.36 \text{ m/s}$  for physics versus MF-RAFT. The angular differences also increase, reaching  $14\text{--}18^{\circ}$ , reflecting the challenges of accurately resolving local vortices and near-bed flow variations.

Notably, the large errors (particularly in the SWE-FE vs. measured case) are mainly due to overestimations by the SWE-FE model near areas of high shear and stone-induced flow separation, where the measured velocities are substantially lower. These localized discrepancies are consistent with the visual examples shown in Figure 10a (middle segments), where velocities derived using SWE-FE are visibly higher than both measured and MF-RAFT-predicted vectors. All comparisons in Figure 10 were performed at a uniform image resolution of  $0.018 \text{ m/px}$ , ensuring a consistent spatial scale across both river sites.

For visualization and quantitative comparison, only points with more than 60% visible surface coverage were included, in accordance with the previously established dependence of accuracy on the valid-pixel ratio. Where more than ten valid measurement points were available, only the ten with the highest visible-area coverage were retained to maintain visual clarity and consistency between segments. In the Šušvė segments, however, this upper limit was not reached, with typically four to seven valid measurement points per area due to the relatively sparse sampling zones and larger bounding regions. It should also be noted that the parsed regions were extracted not only along the main riverbanks but were also constrained by the physical separation of ground markers used during field georeferencing. In the Mūša S1 segment (Figure 10a), physical measurements did not begin exactly at the upstream boundary of the captured area; therefore, these uppermost points are not visible in the figure.

Figure 10 presents the results for segments of varying complexity. The Mūša segments (Figure 10a, middle segments) exhibit high complexity due to the presence of boulders above and below the water, which generate ripples and increased reflections. These effects significantly impact the MF-RAFT predictions and lead to discrepancies in the results compared to both the SWE-FE model and in situ measurements. In the Mūša segments with boulders only below the water (Figure 10a, top and bottom segments), the MF-RAFT predictions show better agreement with the SWE-FE model results. The Šušvė river (Figure 10b) displays low complexity, as there are no boulders in the analyzed segment. Consequently, agreement between the SWE-FE model, in situ measurements, and MF-RAFT results is evident at most of the points selected for visualization. Overall, the visual comparisons indicate consistent directional agreement among all three data sources, despite local discrepancies in zones of complex flow. This coherence across both river sites supports the reliability of the proposed MF-RAFT approach for reconstructing realistic velocity patterns under varying hydrodynamic conditions.



**Figure 10.** Examples of field measurement locations used for independent validation in (a) Mūša and (b) Šušvė river segments (yellow dashed rectangles). All comparisons are shown for  $\Delta_{\text{frames}} = 5$ , spatial resolution = 0.018 m/px, and visible coverage > 60%. Limits of ten points per segment were applied for readability.

#### 4. Discussion

The proposed framework demonstrates that UAV-based RGB video analysis, combined with physically informed data generation and a multiframe deep MF-RAFT model, can produce stable and physically meaningful estimates of river velocity under various hydraulic conditions. Introducing a recurrent fusion module (Fuse-GRU) before correlation computation allows the model to exploit temporal coherence across multiple frames, reducing sensitivity to illumination changes, specular reflections, and low-texture surfaces that typically limit frame-to-frame optical flow. The multiframe RAFT extension thus offers a practical balance between temporal robustness and computational efficiency suitable for UAV hydrometric applications. Because the proposed approach relies on deep learning, its performance is strongly influenced by the quality of the training data. When the training dataset (UAV observations, in situ measurements for the physical model, and the physical

model itself) is prepared with sufficient accuracy, the MF-RAFT model is expected to perform reliably. Nevertheless, environmental factors such as lighting conditions or wind can introduce substantial variability and negatively affect the results.

The hybrid data design, linking SWE-FE flow simulations with UAV-derived imagery, was essential for developing a reliable training dataset. Vector fields generated using FEM provided spatially coherent physical references that are otherwise difficult to obtain from field measurements alone. Although FEM solutions are numerical approximations, residual discrepancies were mainly confined to zones of strong shear, obstacle-induced separation, and vegetated areas. In several of these regions, the optical-flow model locally outperformed the SWE-FE baseline when compared with field data, suggesting that the learning-based system was able to extract subtle visual cues reflecting real dynamics beyond the simplified depth-averaged hydrodynamic representation.

Resolution and cadence analyses offered practical insights for UAV mission planning. Coarser ground sampling distances (approximately 0.010–0.020 m/px) improved performance by increasing spatial context and reducing small-scale radiometric artefacts. Similar benefits could be achieved by expanding the network's input size (e.g., from  $512 \times 512$  px to  $512 \times 1024$  px), although computational efficiency constraints led to the use of smaller patches in this study. Temporal stride effects were dependent on flow regime: shorter intervals benefited fast and turbulent flows, while longer intervals enhanced motion detectability in slow, laminar conditions. Accuracy was strongly influenced by valid-pixel coverage, with at least 80% visible water surface consistently yielding the lowest endpoint and angular errors.

Future work should focus on expanding the scale and diversity of the dataset, including automatically generated geometry of SWE-FE model based on UAV-derived orthophotogrammetry and semi-synthetic hydrodynamic representations. Using dynamic sequence lengths could further enhance temporal context, enabling the model to capture longer flow histories and suppress transient disturbances such as wind-driven ripples. Integrating these improvements, along with physics-informed loss functions, uncertainty quantification, and edge-efficient implementations, would strengthen the framework's applicability for operational, real-time UAV-based river monitoring.

## 5. Conclusions

This study presents a physics-informed, UAV-based approach for estimating river velocity using an enhanced MF-RAFT architecture. By combining FEM-based hydrodynamic modeling with UAV RGB video, a dense and physically coherent training dataset is created, supporting supervised learning under realistic flow conditions. The novelty of the proposed architecture lies in its ability to predict flow using multiple consecutive frames as input rather than relying solely on pairwise flow estimation [17,40]. While traditional particle-based techniques [4,10] focus on tracking discrete tracers or brightness patterns across two frames, the proposed method considers the entire image sequence as a unified source of information. In addition, the framework aims to estimate the actual river velocity, not merely the river surface velocity [12,15,16]. This is achieved by integrating SWE-FE that produces a velocity vector field, which is then linked to observed RGB image features, enabling a more meaningful interpretation of image-derived motion cues. The proposed MF-RAFT model, extended with a pre-correlation Fuse-GRU module, effectively integrates temporal information from consecutive frames, improving stability, robustness, and physical consistency compared with traditional frame-pair methods.

A comprehensive evaluation across multiple rivers demonstrated that the framework reproduces spatially coherent and physically realistic velocity fields, achieving accuracy comparable to in situ measurements and physically based simulations. The method re-

mains resilient under variable illumination, surface texture, and flow regimes, confirming its suitability for UAV-based hydrometric applications. However, severe environmental conditions (e.g., ripples caused by wind) can affect the results. The large computational resources required to train the MF-RAFT model on the high-resolution data is another limitation of using the proposed framework.

Overall, the developed system marks progress towards operational, data-driven hydrometry that integrates computer vision with physical modeling. By combining UAV flexibility, physics-aware learning, and computational efficiency, it offers a scalable, non-invasive tool for continuous river monitoring, flood risk assessment, and hydraulic model validation. Future developments, including automated dataset generation, dynamic temporal modeling, and real-time edge deployment, will further enhance its potential for large-scale environmental observation and water resource management.

**Author Contributions:** Conceptualization, A.K., V.A., D.Č., D.M.-L., K.G., T.F. and R.B., methodology, A.K., V.A., D.Č., D.M.-L., K.G., T.F. and R.B.; software, A.K., D.Č., T.F. and R.B.; validation, A.K., V.A., D.Č., D.M.-L. and K.G.; formal analysis, V.A., D.Č., D.M.-L. and R.B.; investigation, A.K., V.A., D.Č., D.M.-L., K.G., T.F. and R.B.; resources, A.K., V.A., T.F. and R.B.; data curation, A.K., V.A. and D.Č.; writing—original draft preparation, A.K. and D.Č.; writing—review and editing, A.K., V.A., D.Č., D.M.-L., K.G. and R.B.; visualization, A.K., V.A., D.Č. and D.M.-L. supervision, A.K. and R.B.; project administration, R.B.; funding acquisition, R.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Research Council of Lithuania (LMTLT) under the programme of Researcher Groups' projects, the scientific study 'Development of Combined Physical Behavior and Artificial Intelligence Models to Determine Hydromorphology of Rivers by Indirect Measurements (ArtHyReS)' (Agreement Number S-MIP-23-88).

**Data Availability Statement:** The data that support the findings of this study are available from the paper authors upon reasonable request.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Muste, M.; Fujita, I.; Hauet, A. Large-scale Particle Image Velocimetry for Measurements in Riverine Environments. *Water Resour. Res.* **2008**, *44*. [[CrossRef](#)]
2. Tauro, F.; Petroselli, A.; Grimaldi, S. Optical Sensing for Stream Flow Observations: A Review. *J. Agric. Eng.* **2018**, *49*, 199–206. [[CrossRef](#)]
3. Eltner, A.; Sardemann, H.; Grundmann, J. Technical Note: Flow Velocity and Discharge Measurement in Rivers Using Terrestrial and Unmanned-Aerial-Vehicle Imagery. *Hydrol. Earth Syst. Sci.* **2020**, *24*, 1429–1445. [[CrossRef](#)]
4. Detert, M.; Johnson, E.D.; Weitbrecht, V. Proof-of-concept for Low-cost and Non-contact Synoptic Airborne River Flow Measurements. *Int. J. Remote Sens.* **2017**, *38*, 2780–2807. [[CrossRef](#)]
5. Tauro, F.; Tosi, F.; Mattoccia, S.; Toth, E.; Piscopia, R.; Grimaldi, S. Optical Tracking Velocimetry (OTV): Leveraging Optical Flow and Trajectory-Based Filtering for Surface Streamflow Observations. *Remote Sens.* **2018**, *10*, 2010. [[CrossRef](#)]
6. Manfreda, S.; Miglino, D.; Saggi, K.C.; Jomaa, S.; Eltner, A.; Perks, M.; Peña-Haro, S.; Bogaard, T.; van Emmerik, T.H.M.; Mariani, S.; et al. Advancing River Monitoring Using Image-Based Techniques: Challenges and Opportunities. *Hydrol. Sci. J.* **2024**, *69*, 657–677. [[CrossRef](#)]
7. Pizarro, A.; Valera-Gran, D.; Navarrete-Muñoz, E.-M.; Dal Sasso, S.F. The Use of Unmanned Aerial Systems for River Monitoring: A Bibliometric Analysis Covering the Last 25 Years. *Hydrology* **2024**, *11*, 80. [[CrossRef](#)]
8. Fujita, I.; Kunita, Y. Application of Aerial LSPIV to the 2002 Flood of the Yodo River Using a Helicopter Mounted High Density Video Camera. *J. Hydro-Environ. Res.* **2011**, *5*, 323–331. [[CrossRef](#)]
9. Bodart, G.; Le Coz, J.; Jodeau, M.; Hauet, A. Quantifying and Reducing the Operator Effect in LSPIV Discharge Measurements. *Water Resour. Res.* **2024**, *60*, e2023WR034740. [[CrossRef](#)]
10. Legleiter, C.J.; Kinzel, P.J.; Engel, F.L.; Harrison, L.R.; Hewitt, G. A Two-Dimensional, Reach-Scale Implementation of Space-Time Image Velocimetry (STIV) and Comparison to Particle Image Velocimetry (PIV). *Earth Surf. Process. Landf.* **2024**, *49*, 3093–3114. [[CrossRef](#)]

11. Lu, J.; Yang, X.; Wang, J. Velocity Vector Estimation of Two-Dimensional Flow Field Based on STIV. *Sensors* **2023**, *23*, 955. [CrossRef]
12. Fujita, I.; Aya, S. Refinement of LSPIV Technique for Monitoring River Surface Flows. In *Proceedings of the Building Partnerships*; American Society of Civil Engineers: Reston, VA, USA, 2000; pp. 1–9.
13. Fujita, I.; Kunita, Y. Space-Time Image Analysis and Numerical Simulation of Flash Caused by Torrential Rain in Urbanized Area. In *Proceedings of the 34th IAHR World Congress—Balance and Uncertainty*; Engineers Australia: Barton, Australia, 2011.
14. Pearce, S.; Ljubičić, R.; Peña-Haro, S.; Perks, M.; Tauro, F.; Pizarro, A.; Dal Sasso, S.; Strelnikova, D.; Grimaldi, S.; Maddock, I.; et al. An Evaluation of Image Velocimetry Techniques under Low Flow Conditions and High Seeding Densities Using Unmanned Aerial Systems. *Remote Sens.* **2020**, *12*, 232. [CrossRef]
15. Yu, Q.; Rennie, C.D.; Ferguson, S.; Provan, M. TiFA: A New LSPIV Post-Processing Algorithm for River Surface Velocity Measurement under Low Tracer Density Conditions. *J. Hydrol.* **2025**, *661*, 133543. [CrossRef]
16. Yu, K.; Lee, J. Method for Measuring the Surface Velocity Field of a River Using Images Acquired by a Moving Drone. *Water* **2022**, *15*, 53. [CrossRef]
17. Teed, Z.; Deng, J. RAFT: Recurrent All-Pairs Field Transforms for Optical Flow (Extended Abstract). In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*; International Joint Conferences on Artificial Intelligence Organization: San Francisco, CA, USA, 2021; pp. 4839–4843.
18. An, G.; Du, T.; He, J.; Zhang, Y. Non-Intrusive Water Surface Velocity Measurement Based on Deep Learning. *Water* **2024**, *16*, 2784. [CrossRef]
19. Gao, L.; Zhang, Z.; Chen, L.; Li, H. River Surface Space–Time Image Velocimetry Based on Dual-Channel Residual Network. *Appl. Sci.* **2025**, *15*, 5284. [CrossRef]
20. Sun, D.; Roth, S.; Black, M.J. A Quantitative Analysis of Current Practices in Optical Flow Estimation and the Principles Behind Them. *Int. J. Comput. Vis.* **2014**, *106*, 115–137. [CrossRef]
21. Chantas, G.; Gkamas, T.; Nikou, C. Variational-Bayes Optical Flow. *J. Math. Imaging Vis.* **2014**, *50*, 199–213. [CrossRef]
22. Tlhomole, J.B.; Hughes, G.O.; Zhang, M.; Piggott, M.D. From PIV to LSPIV: Harnessing Deep Learning for Environmental Flow Velocimetry. *J. Hydrol.* **2025**, *649*, 132446. [CrossRef]
23. Salandra, M.L.; Colacicco, R.; Panza, S.; Fumai, G.; Dellino, P.; Capolongo, D. RivAir: A Custom-Designed UAV-Based Sensor for Real-Time Water Area Segmentation and Surface Velocity Estimation. *Int. J. Appl. Earth Obs. Geoinf.* **2025**, *142*, 104720. [CrossRef]
24. Koutalakis, P.; Zaimes, G.N. River Flow Measurements Utilizing UAV-Based Surface Velocimetry and Bathymetry Coupled with Sonar. *Hydrology* **2022**, *9*, 148. [CrossRef]
25. Zhou, Z.; Riis-Klinkvort, L.; Jørgensen, E.A.; Lindenhoff, C.; Frías, M.C.; Vesterhauge, A.R.; Olesen, D.H.; Lavish, M.; Dobrovolskiy, A.; Kadek, A.; et al. Measuring River Surface Velocity Using UAS-Borne Doppler Radar. *Water Resour. Res.* **2024**, *60*, e2024WR037375. [CrossRef]
26. Kinzel, P.J.; Legleiter, C.J.; Gazoorian, C.L. Reach-Scale Mapping of Surface Flow Velocities from Thermal Images Acquired by an Uncrewed Aircraft System along the Sacramento River, California, USA. *Water* **2024**, *16*, 1870. [CrossRef]
27. Eltner, A.; Mader, D.; Szopos, N.; Nagy, B.; Grundmann, J.; Bertalan, L. Using Thermal and RGB UAV Imagery to Measure Surface Flow Velocities of Rivers. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2021**, *XLIII-B2-2*, 717–722. [CrossRef]
28. Koutalakis, P.; Stamataki, M.D.; Tzoraki, O. UAV-Based Optical Methods to Estimate the Flow of Temporary Streams and Evaluate Their Environmental Status. *Preprints* **2023**, 2023060768.
29. Torres, W.; Torres, A.; Valencia, E.; Pinchao, P.; Escobar-Segovia, K.; Cando, E. Experimental Validation of the Remote Sensing Method for River Velocity Measurement Using an Open-Source PIV Scheme—Case Study: Antisana River in the Ecuadorian Andes. *Water* **2024**, *16*, 3177. [CrossRef]
30. Koutalakis, P.; Stamataki, M.-D.; Tzoraki, O. Enhancing the Monitoring Protocols of Intermittent Flow Rivers with UAV-Based Optical Methods to Estimate the River Flow and Evaluate Their Environmental Status. *Drones Auton. Veh.* **2024**, *1*, 10006. [CrossRef]
31. Kriščiūnas, A.; Čalnerytė, D.; Akstinas, V.; Meilutytė-Lukauskienė, D.; Gurjazkaitė, K.; Barauskas, R. Framework for UAV-Based River Flow Velocity Determination Employing Optical Recognition. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *134*, 104154. [CrossRef]
32. Horn, B.K.P.; Schunck, B.G. Determining Optical Flow. *Artif. Intell.* **1981**, *17*, 185–203. [CrossRef]
33. Barauskas, R.; Meilutytė-Lukauskienė, D.; Čalnerytė, D.; Kriščiūnas, A.; Gurjazkaitė, K.; Akstinas, V. Modelling Flow Dynamics in Shallow Lowland Rivers with Natural Obstacles. *Hydrol. Process.* **2026**. [CrossRef]
34. Khan, A.A.; Lai, W. *Modeling Shallow Water Flows Using the Discontinuous Galerkin Method*; CRC Press/Taylor & Francis Group: Boca Raton, FL, USA, 2014; ISBN 9781482226010.
35. CFD Module User’s Guide. Version: COMSOL 6.2. 1998–2023. Available online: <https://doc.comsol.com/6.2/docserver/#!/com.comsol.help.comsol/helpdesk/helpdesk.html> (accessed on 17 March 2026).
36. Akstinas, V.; Kriščiūnas, A.; Šidlauskas, A.; Čalnerytė, D.; Meilutytė-Lukauskienė, D.; Jakimavičius, D.; Fyleris, T.; Nazarenko, S.; Barauskas, R. Determination of River Hydromorphological Features in Low-Land Rivers from Aerial Imagery and Direct Measurements Using Machine Learning Algorithms. *Water* **2022**, *14*, 4114. [CrossRef]

37. Shi, X.; Huang, Z.; Bian, W.; Li, D.; Zhang, M.; Cheung, K.C.; See, S.; Qin, H.; Dai, J.; Li, H. VideoFlow: Exploiting Temporal Cues for Multi-Frame Optical Flow Estimation. In Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France, 2–6 October 2023.
38. Sun, S.; Liu, J.; Li, T.H.; Li, H.; Liu, G.; Gao, W. StreamFlow: Streamlined Multi-Frame Optical Flow Estimation for Video Sequences. *arXiv* **2023**, arXiv:2311.17099.
39. Ferede, F.A.; Balasubramanian, M. SSTM: Spatiotemporal Recurrent Transformers for Multi-Frame Optical Flow Estimation. *Neurocomputing* **2023**, *558*, 126705. [[CrossRef](#)]
40. Stone, A.; Maurer, D.; Ayvaci, A.; Angelova, A.; Jonschkowski, R. SMURF: Self-Teaching Multi-Frame Unsupervised RAFT with Full-Image Warping. *arXiv* **2021**, arXiv:2105.07014.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.