Machine Learning and Statistical Analysis of Measured Velocities by ADV of Free Surface Vortices at Intakes

Hamid Ahani[1], Saeed Reza Khodashenas[*,1] ,Hamed Sarkardeh[2]

1. Department of Water Science and Engineering, Agriculture Faculty, Ferdowsi University of Mashhad, Mashhad, Iran.
2. Department of Civil Engineering, Hakim Sabzevari University, Sabzevar, Iran.

***Corresponding Author:**
Saeed Reza Khodashenas,

Email: khodashenas@um.ac.ir
Received: 2025.11.18
Accepted: 2026.01.01
J. Hydraul. Struct., 12(3):54-69
DOI: 10.22055/jhs.2026.50962.1367

ABSTRACT

Free surface vortices in hydraulic intakes can lead to air entrainment, vibrations, and energy loss. This study aimed to optimize an Acoustic Doppler Velocimeter (ADV) for measuring tangential velocities in vortices. The used statistical methods include calculating the median, mode, and arithmetic mean without outlier removal; and removing outliers using truncated mean, Z-Score, box plot, histogram, and a two-standard deviation range. The employed machine learning models were Isolation Forest, One-Class SVM (Support Vector Machines), and DBSCAN (Density-Based Spatial Clustering of Applications with Noise). Experiments were conducted at Fr = 0.85, 1, and 1.27. In order to determine the data acquisition duration, a long-term experiment was conducted, and based on the output graph, 30 seconds was determined as the appropriate time. The results showed that the mean and median at all Fr provided smooth and physical profiles, while the mode method produced unrealistic fluctuations due to the multimodal nature of the data. The outlier removal methods produced almost overlapping profiles with an RMS deviation of less than 1.5%, indicating the stability of these methods and the appropriate quality of the ADV data. Among the machine learning methods, the Isolation Forest algorithm provided the most accurate and stable results, and its behavior was most consistent with valid statistical methods. Comparison of the results with the Rankine model also showed that the measured velocity profile accurately reproduces the linear growth of the velocity in the core and the drop in the outer regions of the vortex. Overall, this research shows that combining ADV data with robust data processing methods specifically, trimmed mean, Z-score, and Isolation Forest significantly increases the accuracy and reliability of eddy velocity profile determination and can be used in the analysis and design of eddy control systems in reservoirs.

KEYWORDS

Free-surface vortex, ADV device, Statistical analysis, Tangential velocity, Inlet intake.

1. Introduction

Vortices are a common occurrence in both natural and industrial systems on a wide range of scales. Free-surface vortices, in particular, are often observed at the intakes of hydropower plants. Free surface vortices form at the water surface and extend down into the intake [1]. The primary factors contributing to vortex formation include changes in boundary direction, non-uniform velocity distribution, and asymmetric flow near the intake [2].

A vortex is a vector quantity, ω, that locally quantifies the rotation of a fluid particle around its center of mass. Mathematically, it is defined as the curl of the velocity vector: $\omega = \nabla \times V$. Vortex flow is typically described using a local cylindrical coordinate system (r, θ, z), where z extends downward from the free surface along the vortex central axis. The corresponding velocities are radial, tangential, and axial, denoted as V_r, V_θ and V_z [1]. Vortices formed in hydraulic intakes are categorized into two types of surface and subsurface vortices. According to the Rankine vortex model [3], surface vortices are further divided into two kinds of forced and free vortex. In a forced vortex, V_θ increases linearly with the distance from the vortex center. Conversely, in a free vortex, V_θ is inversely proportional to this distance [2]. There are two well-known classification systems for vortices that form on the water surface. Hecker [4] physically divided vortices into six types. In a different classification, Sarkardeh et al. [5] categorized them into three classes A, B and C based on their significance and risk level. Table 1 shows the relationship between these two classification systems and the physical and visual characteristics of the vortices.

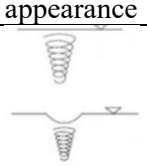
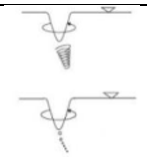
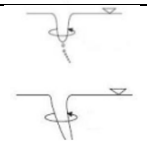
Formation of Surface vortices are an undesirable phenomenon in hydraulic intakes because they lead to air entrainment, vibration, and energy loss [2]. The vortices that form at the intake are a composite model of both forced and free vortices [3]. The strength of a vortex is calculated by its V_θ in the free vortex region, which is obtained from Equation (1) [6].

$$\Gamma = 2\pi r V_\theta \quad (1)$$

In Equation (1), r is the radial distance from the vortex center. Vortex strength (Γ) is a crucial factor for predicting vortex behavior, and it depends on the reservoir geometry, the intake Froude number (Fr), and the relative submergence (S/D) [7]. Understanding the vortex mechanism in hydraulic intakes requires a thorough investigation of the flow velocity.

To analyze a vortex, it is crucial to obtain orthogonal velocities, as vorticity is defined as the curl of the velocity, and orthogonal velocities allow for the precise calculation of local rotation. Velocity measurement methods include Particle Image Velocimetry (PIV) for 2D or 3D velocity fields, Laser Doppler Velocimetry (LDV) for local velocity, and ADV for 3D velocity. The ADV is a device that records velocity fluctuations in three perpendicular directions over a specific range and

Table 1. Vortex types and classes [5]

Vortex type	Definition [4]	Classification [5]	Physical appearance
1	Coherent surface swirl	Class C	
2	Surface dimple; coherent swirl at surface		
3	Dye core; coherent swirl through water column	Class B	
4	Vortex pulling floating trash but no air		
5	Vortex pulling air bubbles into the intake	Class A	
6	Full air core to the intake		

processes them with its software. The ADV velocity measurement is based on the Doppler principle, where it sends out sound waves at a specific frequency and receives their reflections. The frequency difference is proportional to the flow velocity. This device measures velocity at a point 5 cm below the transmitting sensor. Key parameters include the correlation factor (R^2) and the Signal-to-Noise Ratio (SNR). The correlation factor can be considered a quality parameter for the measured data and is recommended to be in the 70% to 100% range. Since the ADV velocity measurement is based on the Doppler principle, a short sound pulse with a known frequency is first sent vertically. The reflection of sound from the water is then received by three small probes, amplified by a converter, and analyzed by the device's processor card. The change in frequency between the transmitted and received pulses is a result of the water's flow velocity. To ensure proper device function, the reflected sound must be strong enough for the device to calculate the resulting frequency change. The strength of the return sound is represented by the signal-to-noise ratio. An SNR of at least 5 dB is recommended for data acquisition [8]. The ADV is used in hydraulic laboratories to measure 3D velocities, velocity profiles, and Reynolds stresses.

Early investigations on free surface vortices primarily focused on their symmetric behavior. Sarkardeh et al. [5] showed that vortex strength decreases as the intake wall becomes more vertical, while Camnasio et al. [9] linked reservoir geometry to flow patterns using ultrasonic profiling. Following this, Sarkardeh et al. [10] measured the flow field in a reservoir with a free-surface vortex using an ADV, graphically presenting their findings. With advances in experimental techniques, research shifted toward understanding the 3D and frequently asymmetric nature of vortices. PIV and high-speed particle image velocimetry (HSPIV) measurements by Suerich-Gulick et al. [1], Mulligan et al. [11], Naderi and Gaskin [12], and Naderi et al. [13] demonstrated the strong influence of approach flow on vortex asymmetry. More recent studies refined this understanding: Duinmeijer et al. [14] quantified axial flow transport in the vortex core, and Hamidi et al. [15] mapped 3D particle trajectories for different vortex classes. Alongside improved understanding of vortex mechanisms, numerous works have focused on mitigation and hydraulic control. These include circulation zone analysis near intakes [16], performance of anti-vortex plates [17], the effect of intake angle and sediment transport [18,19,20], and development of new suppression methods [21,22]. Taghizadeh et al. [23] provided a comprehensive review of the most recent vortex-formation mechanisms, highlighting the multi-factor and highly unsteady nature of intake vortices. Newer experimental studies have also expanded the understanding of vortex induced instability: Amini et al. [24] numerically examined vortex behavior under variable temperature fields; Zhang et al. [25] characterized vortex evolution and pressure fluctuations in pump sumps; Jiao et al. [26] analyzed suction vortex formation governed by clearance and inflow conditions; and Mondal et al. [27] experimentally captured air entrainment stages during free surface vortex development. Collectively, these recent findings emphasize that accurate, high-resolution velocity field measurement is crucial for understanding vortex dynamics. Given this need, the ADV has become a common tool for vortex studies. However, ADV generates large volumes of instantaneous, noisy, and highly fluctuating velocity data, which must be filtered using signal quality metrics such as SNR and R^2 before identifying a representative velocity value suitable for hydraulic interpretation. Traditional statistical filtering methods are often insufficient for fully removing outliers under strongly unsteady vortex conditions.

In parallel, recent developments show that machine learning is increasingly effective in analyzing complex hydraulic data. Roushangar et al. [28] successfully modeled energy dissipation over stepped

spillways; Ghorbani et al. [29,30] predicted discharge coefficients and suspended sediment concentration with ML models; Salmasi and Abraham [31] used hybrid ML methods to estimate inclined-gate discharge; and Salmasi et al. [32] demonstrated the superior performance of models such as RF, GEP, MARS, and M5P in predicting scour depth. These studies collectively highlight the capability of ML to extract reliable information from noisy, nonlinear hydraulic signals similar to ADV datasets recorded under vortex flow. The novelty of the present study lies in conducting, for the first time, a systematic comparison between statistical and ML based outlier removal approaches to improve ADV velocity data quality under vortex flow conditions. The objective is to determine the most reliable method for extracting representative 3D velocities, thereby enhancing the accuracy of vortex flow characterization.

2. Materials and Methods

Tests were performed in a rectangular reservoir with transparent walls, measuring 0.7 m in length, 0.6 m in width, and 0.7 m in height. A vertical intake with a 0.04 m diameter opening was used. To enable the observation of stronger vortices and minimize scaling effects, the distance from the center of the intake to the side walls was set to at least 4D (where D is the diameter of the intake pipe), as recommended by Anwar et al. [33]. To stabilize the incoming flow, two perforated plates were installed in the flow path. This flow conditioner not only eliminated water fluctuations but also ensured a uniform flow into the reservoir, allowing for the use of the maximum reservoir space. By selecting a closed-loop water circuit and using a variable-speed electric pump with adequate power, it was possible to generate strong vortices and control various flow rates at the intake and set the water surface at a fixed intake submergence (S). A flow meter was used to measure the volumetric flow rate of the water, with an accuracy of 0.01 L/s. Additionally, a scale with an accuracy of ±1 mm was installed on the reservoir wall to read the intake's submergence depth. Figure 1 shows a schematic view of the experimental setup.

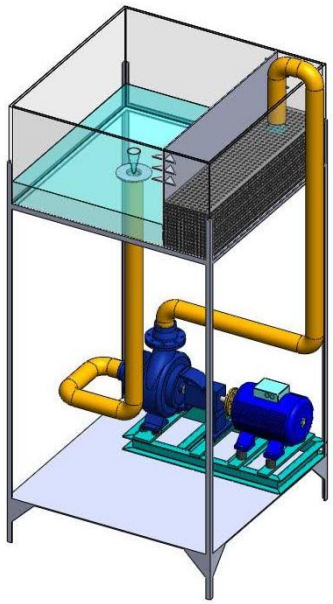


Figure 1. A Schematic view of the experimental setup

The experimental conditions were as follows: three tests were conducted with a constant relative submergence depth (S/D) of 3.25 and a constant data acquisition depth (Z) of 1.25 D, under three different Froude numbers (Fr = 0.85, 1, and 1.27). For measuring instantaneous velocities, a Nortek Vectrino ADV with four down-looking probes was used, capable of recording 3D velocity components. The device frequency was set to 25 Hz. The number of valid data points was not constant and could be reduced due to limitations such as a signal-to-noise ratio (SNR) below 5 dB or correlation coefficient (R^2, COR) below 70%. The raw data were filtered and validated using WinADV software to remove possible spikes and noise. To analyze the instantaneous velocity data obtained by the ADV and achieve reliable velocity values for flow analysis, a set of statistical and machine learning methods was applied. These analyses were divided into three categories: analysis using all data, analysis with outlier removal, and analysis using machine learning methods for outlier detection.

2.1. Statistical Analysis Using All Data

In this step, the statistical measures were calculated without removing any data. Median index which is the middle value of the data after it has been sorted in ascending order. This measure is robust against outliers and represents the center of the data distribution without being influenced by extreme values. Mode index which is the value that appears most frequently in the dataset. This measure is useful for identifying the dominant or recurring velocities in the flow and can indicate stable patterns in the instantaneous data. Arithmetic Mean (avg) index which is calculated as the sum of all data points divided by the total number of points. This is the simplest measure of central tendency but is sensitive to outliers, which can result in an over- or underestimated value in the presence of noise.

$$avg = \frac{\sum x_i}{n} \quad (2)$$

2.2. Statistical Analysis with Outlier Removal

To improve accuracy and eliminate the effect of noise or invalid data (e.g., spikes from acoustic interference or suspended particles), various methods were applied to identify and remove outliers. After their removal, the arithmetic mean was calculated from the remaining data. The following methods used were employed. Trimmed Mean (avg 8%): In this method, the data was sorted in ascending order, and the top and bottom 4% of the data were considered outliers. The arithmetic mean was then calculated from the remaining 92% of the data. This is a simple and effective method for reducing the effect of outliers without complex calculations. Z-Score: The Z-Score method identifies outliers based on their standard distance from the mean. The Z-Score for each data point x is calculated as:

$$z = \frac{x - \mu}{\sigma} \quad (3)$$

where μ is the mean of all data and σ is the standard deviation. Data points with an absolute Z-Score greater than 3 ($|Z| > 3$) were identified and removed as outliers. The arithmetic mean was then calculated from the remaining data. This method assumes a normal distribution, making it suitable for velocity data, which often follows a normal distribution. Box Plot: In a box plot, outliers are identified using the Interquartile Range (IQR). First, the first quartile (Q1, the 25th percentile) and the third quartile (Q3, the 75th percentile) are calculated, and then the IQR is determined by

$$IQR = Q3 - Q1 \quad (4)$$

Data points less than Q1 - 1.5 × IQR or

greater than Q3 + 1.5 × IQR are considered outliers. After removing these points, the arithmetic mean is calculated from the remaining data. This method is visual and non-parametric, making it useful for skewed distributions. Histogram: In a histogram, data is divided into bins, and the frequency of each bin is plotted. Outliers are typically identified at the tails of the distribution or in bins with very low frequency, as these points are far from the main distribution (e.g., a normal peak). After identifying and removing these outliers (either through visual inspection or a frequency threshold), the arithmetic mean is calculated from the remaining data. This method is effective for detecting distribution patterns. Standard Deviation Range (2 σ): The standard deviation (σ) is a measure of data dispersion around the mean and is calculated as:

$$\sigma = \frac{\sum(x_i - \mu)^2}{n} \quad (5)$$

In this method, for each data point, a range from (μ - 2σ) to (μ + 2σ) was considered. This range covers approximately 95% of a normal distribution. The arithmetic mean was then calculated only from the data within this range, eliminating the effect of outliers outside of it.

2.3. Machine Learning Methods for Outlier Detection

For more accurate outlier detection, especially in complex data with high noise, three machine learning models were applied. Each model identified the outliers, and the arithmetic mean was then calculated from the non-outlier data.

Isolation Forest: This algorithm is a decision tree-based method designed for outlier detection. It isolates data points through random partitioning and calculates the path length h(x) for each point x until it reaches a leaf node. Points that are isolated earlier (i.e., those with a lower average path length E (h(x)) are identified as outliers, as they are less connected to the main data groups. The anomaly score for each point is computed using the following formula:

$$s(x,n) = 2^{-\frac{E\,(h\,(x)\,)}{c\,(n)}} \quad (6)$$

where:

$$c\,(n) = 2H\,(n-1) - \frac{2\,(n-1)}{n} \quad (7)$$

represents the average path length in a random binary tree with n samples, and $H(i) \approx \ln(i) + 0.577$ (0.577 is the Euler-Mascheroni constant) is the harmonic number function. The symbols used in the equations are defined as follows: $s(x,n)$: Anomaly score for a point x given n samples, ranging from 0 to 1, where values close to 1 indicate outliers and values close to 0.5 suggest normal data. $h(x)$: Path length of point x, representing the number of nodes traversed in a decision tree until reaching a leaf. $E(h(x))$: Expected path length of point x, averaged over multiple trees. $c(n)$: Average path length for a random binary tree with n samples, used as a normalization factor. n: Total number of samples in the dataset. x: A data point in the dataset. i: Index variable used in the harmonic number calculation [34].

The parameter contamination (estimated outlier ratio, e.g., 0.1) is used to adjust the method's sensitivity. After removing the outliers, the mean is calculated from the remaining data. This method is suitable for large datasets and does not require a specific data distribution assumption.

One-Class SVM: This model is an unsupervised learning algorithm that learns the boundaries of normal data using the Support Vector Machine approach and identifies points outside these boundaries as outliers. The decision function for this model is defined as follows:

$$f(x) = sign\left(\sum_{i=1}^{n} \alpha_i\, k\,(x, x_i) - \rho\right) \quad (8)$$

Where $k(x, x_i)$ is the kernel function (typically the Gaussian kernel $\exp(-\Upsilon\|x - x_i\|^2)$, α_i are the Lagrange coefficients for support vectors, and ρ is the bias term

determining the decision boundary. The algorithm operates by solving the following optimization problem:

$$\min_{\omega,\xi,\rho} \frac{1}{2}\| \omega \|^2 + \frac{1}{\upsilon\, n}\sum_{i=1}^{n}\xi_i - \rho \quad (9)$$

subject to the constraint:

$$\left(\omega, \varphi(x_i)\right) \geq \rho - \xi_i, \quad \xi_i \geq 0, \quad i = 1, \ldots, n \quad (10)$$

where ω is the weight vector in the feature space, $\varphi(x_i)$ is the mapping of data points to the feature space, υ is a parameter controlling the proportion of outliers (typically between 0 and 1, e.g., 0.1), ξ_i are slack variables for points outside the boundary, and n is the total number of samples [35].

The model assumes that most data points are normal and detects outliers based on their distance from the distribution center. After removing the outliers, the arithmetic mean is calculated from the remaining data. This method is ideal for discrete, unclassified data such as velocities measured by ADV.

DBSCAN (Density-Based Spatial Clustering of Applications with Noise): This algorithm is a density-based clustering method that groups points based on proximity and a minimum number of neighbors (min_samples, e.g., 5). A point x is considered a core point if the number of points within its ε-neighborhood (e.g., ε = 1. 5) is at least min_samples, as defined by:

$$| N_\varepsilon(x) | \geq \min_samples \quad (11)$$

where:

$$N_\varepsilon(x) = \{x' \in D | \, dist(x, \acute{x}) \leq \varepsilon\} \quad (12)$$

is the set of points within the ε-neighborhood of x, and

$$dist(x, x') = \sqrt{\sum_{i=1}^{d}(x_i - x'_i)^2} \quad (13)$$

is the Euclidean distance between points x and x', with d representing the data dimensionality and x_i and x'_i the coordinates of x and x', respectively. Points that do not belong to any cluster or have low density (i.e., $N_\varepsilon(x)$ < min_samples) are identified as noise or outliers [36]. After removing these points, the mean is calculated from the data within the clusters. This method is suitable for data with irregular distributions and does not require a pre-determined number of clusters.

3. Results and Discussion

Figure 2 shows a picture of the formed vortices of Type A and B in the present experimental study.



Fr = 1.27 **Fr = 1** **Fr = 0.85**

Figure 2. Formed air-core vortex in the experimental model

To determine the optimal sampling duration, a long-term experiment of 600 seconds was conducted at a fixed point. This experiment resulted in the collection of approximately 15,000 instantaneous velocity data points. Figure 3 presents the distribution of these velocity values against their frequency of occurrence, illustrating the repetition rate of the recorded instantaneous velocities.

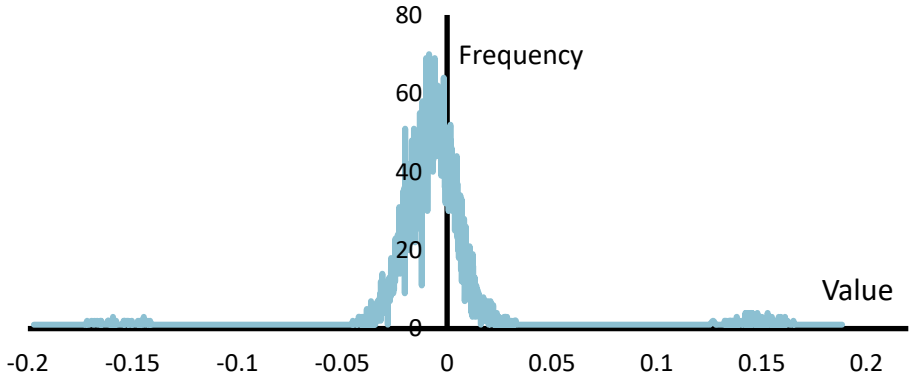


Figure 3. Distribution of instantaneous velocity values versus their frequency of occurrence

To identify the optimal sampling duration, the collected data were analyzed cumulatively over time intervals of 10, 20, 30, … up to 600 seconds. Figure 4 illustrates the variations of tangential velocity (V_θ) using different statistical methods and machine learning algorithms. The results indicate that significant fluctuations occur during the initial stage of the experiment, particularly within the first 30 seconds. However, after approximately 30 seconds, most methods exhibit a stable pattern, and the velocity variations become minimal. This trend is confirmed by the convergence of statistical approaches (mean, median, mode, and others) as well as machine learning methods (Isolation Forest, One-Class SVM, and DBSCAN). In other words, the agreement among these different approaches beyond this duration confirms the stabilization of the velocity trend. Accordingly, it can be concluded that a 30-second sampling interval is sufficient for reliable data acquisition using the Acoustic

Doppler Velocimeter (ADV). Within this duration, approximately 750 instantaneous velocity data points are collected, although the final number may be reduced due to the applied filters. This choice ensures that the collected data are statistically reliable while avoiding unnecessarily long sampling times.

Based on the analytical Rankine vortex model, vortices that form at the intake are a composite of two regions of a forced vortex (rotational) and a free vortex (irrotational). In the forced vortex region, V_θ increases linearly with the distance from the center of rotation. In contrast, in the free vortex region, the magnitude of V_θ is inversely proportional to the distance from the center. The boundary between these two regions is approximately at a distance equal to the radius of the intake from the axis of rotation [3]. Figure 5 shows the tangential velocity distribution profile based on the Rankine model.

Figures 6 to 8 show the results of statistical analyses performed on the radial cross section of the vortex obtained from the present experimental study at Fr = 0.85, 1.0 and 1.27 and with three states statistical analysis using all data, statistical analysis with outlier removal, and machine learning methods for outlier detection.

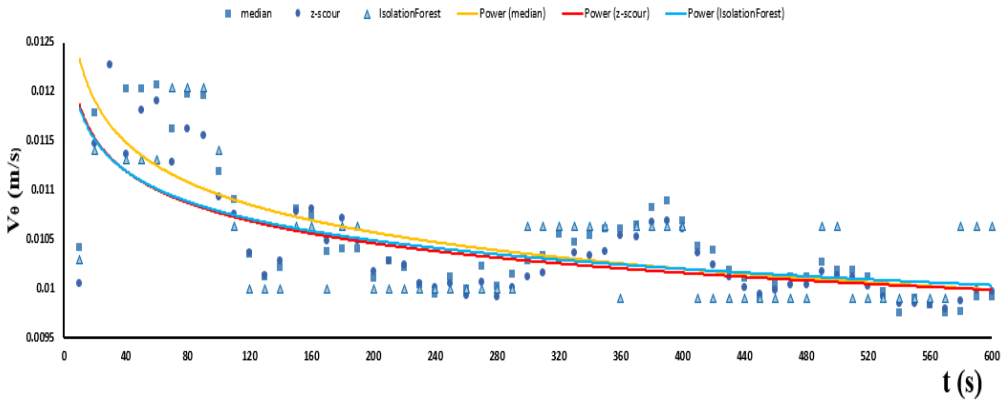


Figure 4. Statistical Analysis of Velocity Components (V_θ) Across Various Sampling Durations Using Statistical and Machine Learning Methods

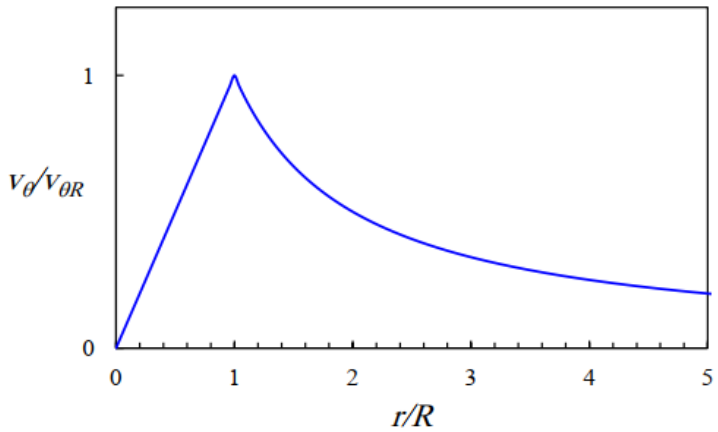


Figure 5. Tangential velocity distribution profile based on the Rankine model

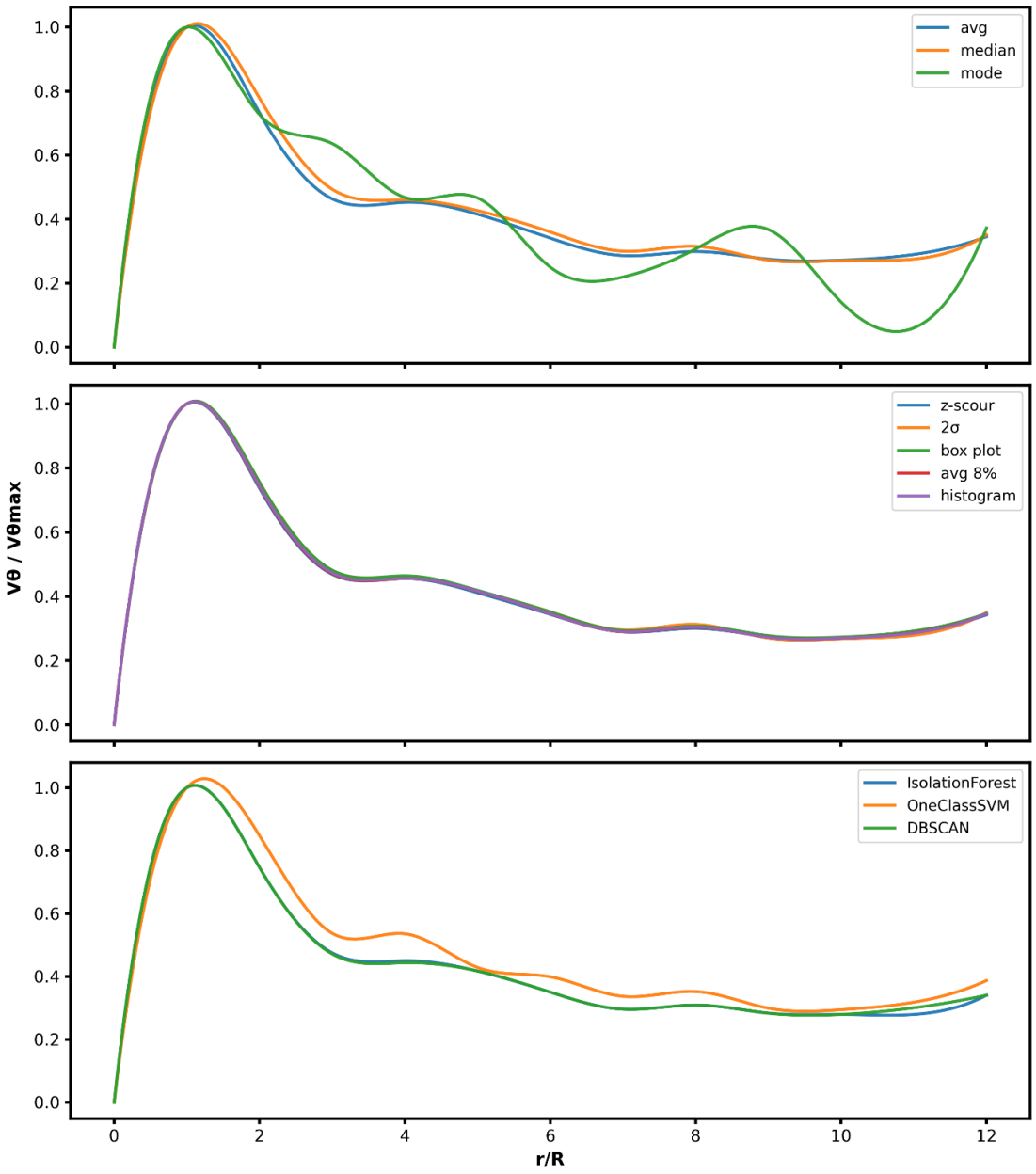


Figure 6. The statistical analyses conducted on a radial cross-section of the vortex for Fr = 0.85

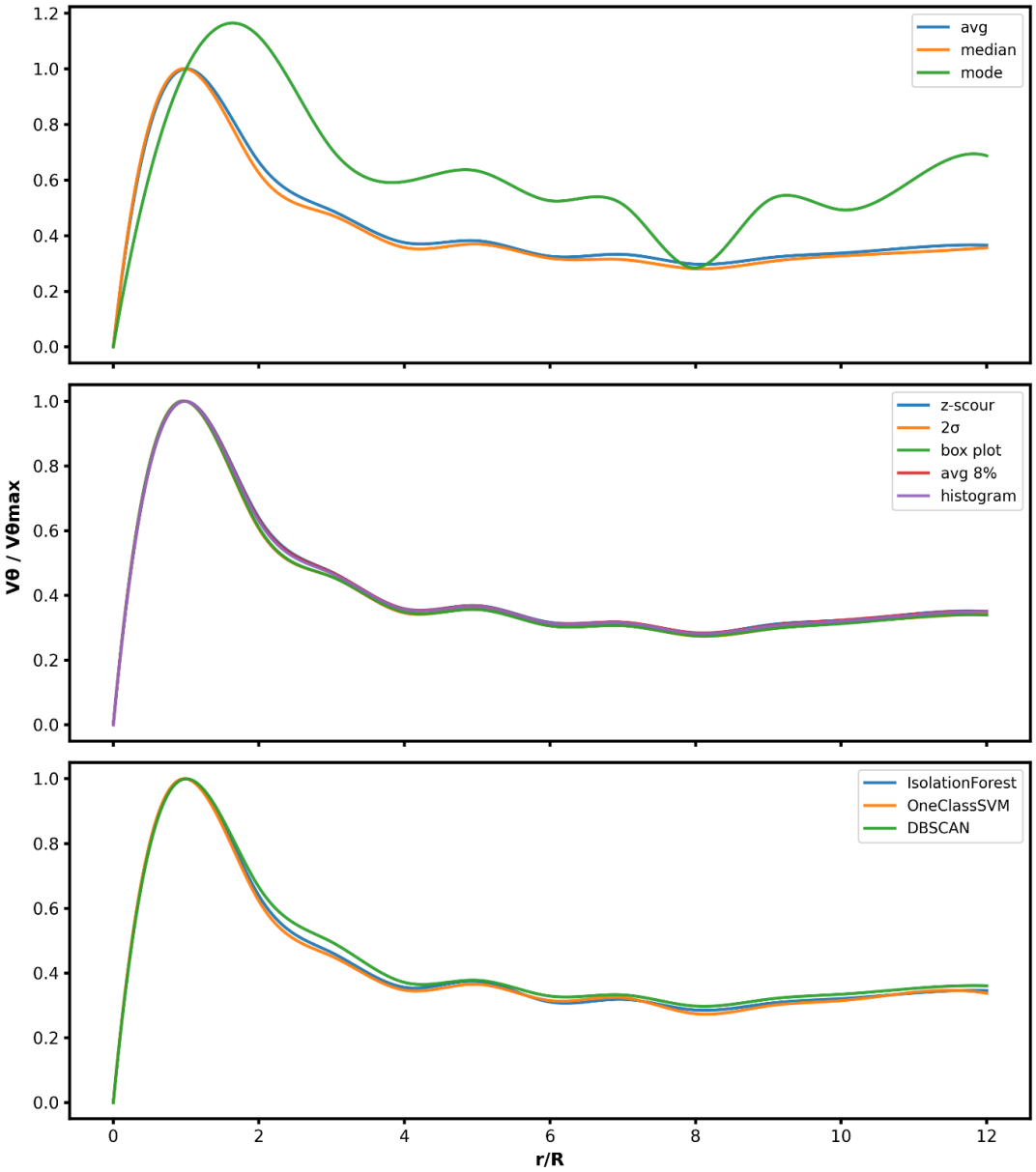


Figure 7. The statistical analyses conducted on a radial cross-section of the vortex for Fr = 1

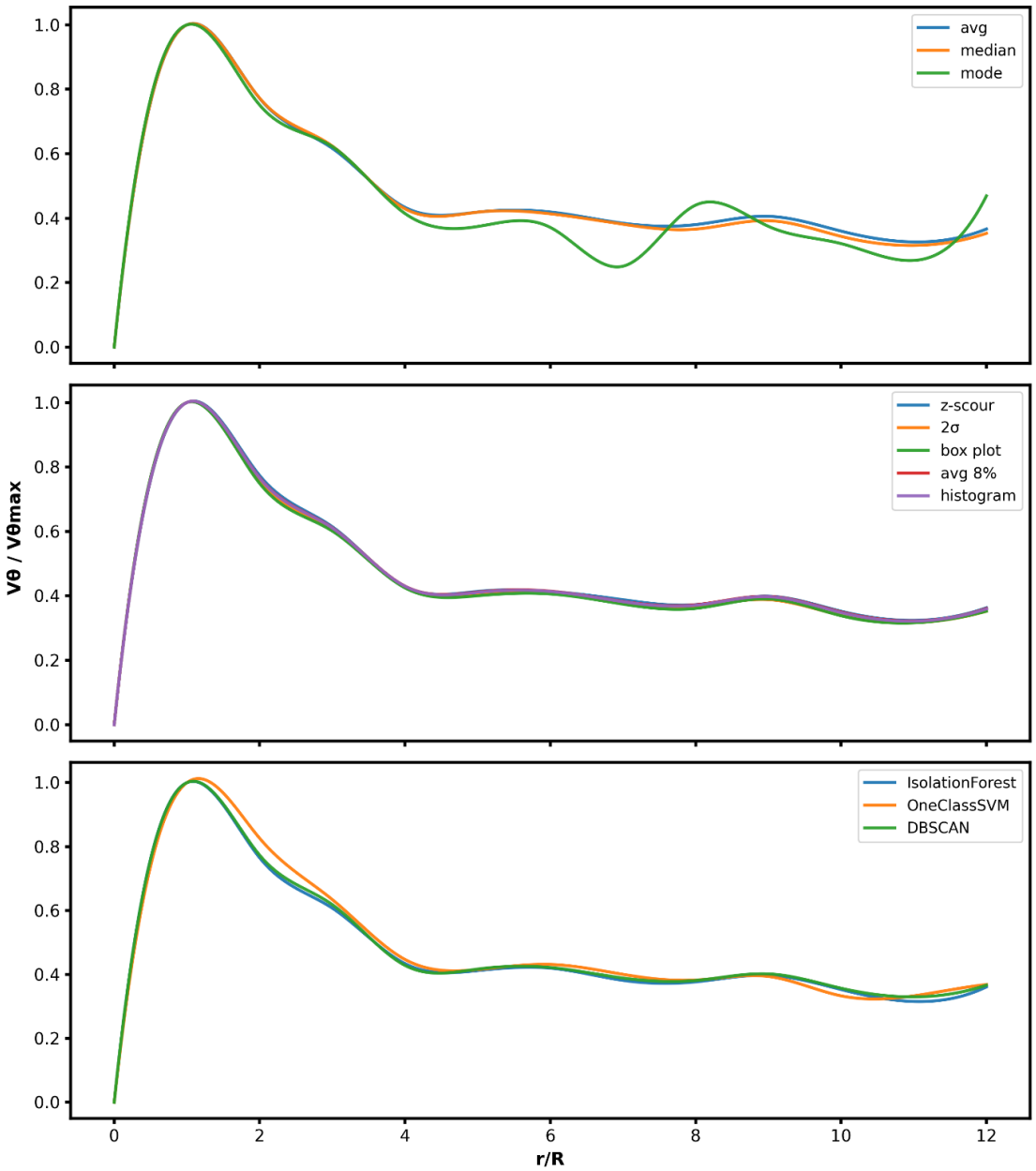


Figure 8. The statistical analyses conducted on a radial cross-section of the vortex for Fr = 1.27

The normalized tangential velocity profiles obtained for Fr = 0.85, 1, and 1.27 demonstrate that the vortex structure near the intake exhibits a stable and coherent behavior consistent with the Rankine vortex model. In all cases, the peak velocity occurs at approximately $r/R \approx 1$ with a normalized magnitude of unity. Beyond this radius, the tangential velocity decreases smoothly, reaching values of about 0.40–0.50 at $r/R = 4$ and 0.28–0.35 at $r/R = 10$. This corresponds to a reduction of nearly 55% in the intermediate region and around 70% in the far field.

In the statistical analysis without outlier removal, the mean and median produced smooth and physically consistent profiles for all Fr. Their differences remained within $0.02 - 0.03$ (approximately 2–3% of the peak value), indicating minimal skewness in the data distribution and acceptable quality of the raw ADV measurements. In contrast, the mode exhibited the poorest and most unstable behavior. For instance, oscillations of up to 15% were observed in Fr = 1 and fluctuations of 5 – 7 % appeared in Fr = 1.27. This nonphysical behavior is attributed to a fundamental limitation of the mode: when the data distribution is multimodal, the method returns only the first detected mode, resulting in artificial peaks and distorted profile shapes.

Applying outlier removal techniques including z-score, 2σ, box plot, average 8%, and histogram resulted in profiles that were nearly identical for all Fr. The differences among these methods remained within 2 – 3 %, and the RMS deviation was typically below 1.5%. This high degree of overlap indicates that the ADV data contain very limited noise and that removing outliers does not alter the fundamental shape of the velocity profile. These methods collectively produced the smoothest, most stable, and most Rankine like profiles, demonstrating that statistical outlier removal provides the most reliable framework for flow characterization across all cases.

Machine learning based outlier detection methods (Isolation Forest, DBSCAN, and One-Class SVM) also produced physically meaningful and smooth profiles, though with varying levels of agreement. Isolation Forest consistently delivered the most accurate and stable results across all Fr, with profiles closely matching those from the most reliable statistical methods. Its superiority arises from its structural mechanism: Isolation Forest isolates outliers through recursive partitioning, requiring fewer splits for points that deviate structurally from the main data cluster. Since the primary body of velocity data forms a coherent geometric trend, this method removes only genuinely anomalous points, minimizing false outlier classification. DBSCAN also performed well; however, due to its sensitivity to density based parameters, small deviations ($\leq$ 2 %) were observed, particularly at higher Fr. One-Class SVM slightly overestimated velocities in the mid radius region by about 0.02–0.03, but overall retained a physically reasonable trend.

The results also show that as the Fr increases from 0.85 to 1.27, fluctuations in the weaker methods increase, highlighting the need for more robust methods in higher-turbulence regimes.

The results of this study are remarkably similar to those of Cea et al. [37] in terms of the importance of processing ADV velocity data; however, the present study provides a more comprehensive and detailed approach. Both studies emphasize that raw ADV data, especially in turbulent and near-free-surface flows, contain significant amounts of spikes and noise, and that proper data processing has a decisive impact on the extraction of velocity profiles and turbulence parameters. In the study by Cea et al., threshold-based and phase-space filters were investigated and it was shown that the turbulence energy and frequency spectrum are strongly affected by the filtering quality.

4. Summary and Conclusions

This research aimed to optimize the Acoustic Doppler Velocimeter (ADV) for

measuring tangential velocities in free-surface vortices at hydraulic intakes. Experiments were conducted in a rectangular tank at Froude numbers of 0.85, 1, and 1.27 and the data were analyzed using statistical and machine learning methods. The results showed A comparison among the three Fr that although flow turbulence slightly increases from Fr = 0.85 to 1.27, the essential vortex structure remains stable. Both statistical and machine learning methods especially Isolation Forest were capable of reconstructing accurate and physically plausible profiles. Across all cases, the RMS deviation among reliable methods remained within 2 – 3 %, whereas the mode exhibited deviations as large as 15%, confirming its unreliability for vortex characterization. These findings demonstrate that the underlying vortex structure is robust with respect to the choice of post-processing method and that only methods with inherent statistical limitations (such as the mode) produce nonphysical results. Overall, the results indicate that statistical methods with outlier removal, along with Isolation Forest, provide the most accurate and trustworthy reconstruction of the vortex velocity profile. These methods yield smooth, noise resistant, and Rankine consistent curves and are therefore recommended for detailed hydrodynamic analysis of free surface vortices. In contrast, the mode consistently exhibited unstable behavior and should not be used for interpreting ADV measurements in similar studies. Future research could focus on fine-tuning the parameters of machine learning models for ADV data to achieve greater accuracy. These findings contribute to improving the design of hydraulic intakes and reducing the risks associated with vortex formation.

Funding Statement
This research received no external funding.

Conflict of Interest
The authors declare that they have no conflict of interest.

Author Contributions
All authors contributed to the study conception and design. Material preparation, data collection, and analysis, were performed by H.A, SR.K and H.S. The first draft of the manuscript was written by H.A and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

References
1. Suerich-Gulick F, Gaskin SJ, Villeneuve M, Parkinson É. Characteristics of free surface vortices at low-head hydropower intakes. Journal of Hydraulic Engineering. 2014 Mar 1;140(3):291-9. https://doi.org/10.1061/(ASCE)HY.1943-7900.0000826
2. Knauss J, editor. Swirling flow problems at intakes. Routledge; 2017 Nov 1.
3. Rankine WJ. A manual of applied mechanics. C. Griffin and Company, Limited; 1901.
4. Hecker GE, editor. Fundamentals of vortex intake flow. IAHR hydraulic structures design manual. Rotterdam: Balkema, [chapters 2 and 8]. 1987.
5. Sarkardeh H, Zarrati AR, Roshan R. Effect of intake head wall and trash rack on vortices. Journal of Hydraulic Research. 2010 Feb 1;48(1):108-12. https://doi.org/10.1080/00221680903565952
6. White F. Fluid Mechanics (5th ed.). New York, NY: McGraw-Hill. 2003.
7. Travis QB, Mays LW. Prediction of intake vortex risk by nearest neighbors modeling. Journal of Hydraulic Engineering. 2011 Jun 1;137(6):701-5. https://doi.org/10.1061/(ASCE)HY.1943-7900.0000344
8. Heravi Z, Azhdary Moghaddam M, Esmaili K, Givehchi M, Bahr Peyma A. Experimental Evaluation of Performance of the New Design of Bed Protection Models (F-jacks) in Altering the Flow Pattern around Bridge Piers. Journal of Hydraulics. 2024 Mar 20;19(1):67-86. https://doi.org/10.30482/jhyd.2023.395247.1643
9. Camnasio E, Orsi E, Schleiss AJ. Experimental study of velocity fields in rectangular shallow reservoirs. Journal of Hydraulic Research. 2011 Jun 1;49(3):352-8.

https://doi.org/10.1080/00221686.2011.574387
10. Sarkardeh H, Jabbari E, Zarrati AR, Tavakkol S. Velocity field in a reservoir in the presence of an air-core vortex. InProceedings of the Institution of Civil Engineers-Water Management 2014 Jun;167(6):356-64. Thomas Telford Ltd. https://doi.org/10.1680/wama.13.00046
11. Mulligan S, Casserly J, Sherlock R. Effects of geometry on strong free-surface vortices in subcritical approach flows. Journal of Hydraulic Engineering. 2016 Nov 1;142(11):04016051. https://doi.org/10.1061/(ASCE)HY.1943-7900.0001194
12. Naderi V, Gaskin S. A 3D study of an intake air-core vortex structure using PIV & flow visualization. In: Proc of 7th International Symposium on Hydraulic Structure. 2018: Aachen, Germany.
13. Naderi V, Farsadizadeh D, Lin C, Gaskin S. A 3D study of an air-core vortex using HSPIV and flow visualization. Arabian Journal for Science and Engineering. 2019 Oct;44(10):8573-84. https://doi.org/10.1007/s13369-019-03764-3
14. Duinmeijer A, Oldenziel G, Clemens F. Experimental study on the 3D-flow field of a free-surface vortex using stereo PIV. Journal of Hydraulic research. 2020 Jan 2;58(1):105-19. https://doi.org/10.1080/00221686.2018.1555558
15. Hamidi, S., Sarkardeh, H., Bashtani, A. Spiral Flow Measurement at Vertical Intakes Using PTV Method. Journal of Hydraulic Structures, 2024; 11(1): 41-50. doi: 10.22055/jhs.2024.47215.1304
16. Azarpira M, Sarkardeh H, Tavakkol S, Roshan R, Bakhshi H. Vortices in dam reservoir: A case study of Karun III dam. Sadhana. 2014 Oct;39(5):1201-9. https://doi.org/10.1007/s12046-014-0252-7
17. Amiri SM, Zarrati AR, Roshan R, Sarkardeh H. Surface vortex prevention at power intakes by horizontal plates. InProceedings of the institution of civil engineers-water management 2011 Apr;164(4):193-200. Thomas Telford Ltd. https://doi.org/10.1680/wama.1000009
18. Khanarmuei M, Rahimzadeh H, Sarkardeh H. Investigating the effect of intake withdrawal direction on critical submergence and strength of vortices. Modares Mechanical Engineering. 2015;14(10):35-42.
19. Khanarmuei MR, Rahimzadeh H, Kakuei AR, Sarkardeh H. Effect of vortex formation on sediment transport at dual pipe intakes. Sādhanā. 2016 Sep;41(9):1055-61. https://doi.org/10.1007/s12046-016-0531-6
20. Khanarmuei M, Rahimzadeh H, Sarkardeh H. Effect of dual intake direction on critical submergence and vortex strength. Journal of Hydraulic Research. 2019 Mar 4;57(2):272-9. https://doi.org/10.1080/00221686.2018.1459896
21. Monshizadeh M, Tahershamsi A, Rahimzadeh H, Sarkardeh H. Comparison between hydraulic and structural based anti-vortex methods at intakes. The European Physical Journal Plus. 2017 Aug 3;132(8):329. https://doi.org/10.1140/epjp/i2017-11608-4
22. Monshizadeh M, Tahershamsi A, Rahimzadeh H, Sarkardeh H. Vortex dissipation using a hydraulic-based anti-vortex device at intakes. International Journal of Civil Engineering. 2018 Sep;16(9):1137-44. https://doi.org/10.1007/s40999-017-0266-8
23. Taghizadeh A, Sarkardeh H, Azamathulla HM. Mechanisms of free surface vortex formation at water intakes (a review). Dams and Reservoirs. 2025 Sep 11. https://doi.org/10.1680/jdare.25.01007
24. Amini AA, Sarkardeh H, Jabbari E, Eidi A. Understanding vortex characteristics in hydraulic systems: a temperature-driven analysis. Modeling Earth Systems and Environment. 2024 Jun;10(3):3559-71. https://doi.org/10.1007/s40808-024-01969-6
25. Zhang B, Ma L, Cheng L, Zhu B, Qin Y. Characterization of free-surface vortex evolution and its impact on hydraulic stability of the pump sump. Experimental Thermal and Fluid Science. 2025 Jul 1;166:111466. https://doi.org/10.1016/j.expthermflusci.2025.111466
26. Jiao W, Fan H, Xi X, Shen J, Zhou L, Jia X, Wang Z, Cheng L. Study on the key influencing factors and vortex characteristics of the water jet propulsion pump underwater suction vortex. Physics of Fluids. 2025 Sep 1;37(9). https://doi.org/10.1063/5.0287735
27. Mondal RK, Rohilla L, Kumar P. Free surface vortex and associated air entrainment during liquid drainage using two outlets. Experimental Thermal and Fluid Science. 2025 Jan 1;160:111314.

https://doi.org/10.1016/j.expthermflusci.2024.111314
28. Roushangar K, Akhgar S, Salmasi F, Shiri J. Modeling energy dissipation over stepped spillways using machine learning approaches. Journal of Hydrology. 2014 Jan 16;508:254-65. https://doi.org/10.1016/j.jhydrol.2013.10.053
29. Ghorbani MA, Salmasi F, Saggi MK, Bhatia AS, Kahya E, Norouzi R. Deep learning under H2O framework: A novel approach for quantitative analysis of discharge coefficient in sluice gates. Journal of Hydroinformatics. 2020 Nov 1;22(6):1603-19. https://doi.org/10.2166/hydro.2020.003
30. Ghorbani MA, Khatibi R, Singh VP, Kahya E, Ruskeepää H, Saggi MK, Sivakumar B, Kim S, Salmasi F, Hasanpour Kashani M, Samadianfard S. Continuous monitoring of suspended sediment concentrations using image analytics and deriving inherent correlations by machine learning. Scientific reports. 2020 May 22;10(1):8589. https://doi.org/10.1038/s41598-020-64707-9
31. Salmasi F, Abraham J. Expert system for determining discharge coefficients for inclined slide gates using genetic programming. Journal of Irrigation and Drainage Engineering. 2020 Dec 1;146(12):06020013. https://doi.org/10.1061/(ASCE)IR.1943-4774.0001520
32. Salmasi F, Sihag P, Abraham J, Nouri M. Experimental investigation and prediction of free fall jet scouring using machine learning models. International Journal of Sediment Research. 2023 Jun 1;38(3):405-20. https://doi.org/10.1016/j.ijsrc.2022.11.004
33. Anwar HO, Weller JA, Amphlett MB. Similarity of free-vortex at horizontal intake. Journal of Hydraulic Research. 1978 Apr 1;16(2):95-105. https://doi.org/10.1080/00221687809499623
34. Liu FT, Ting KM, Zhou ZH. Isolation forest. In2008 eighth ieee international conference on data mining 2008 Dec 15 (pp. 413-422). IEEE. https://doi.org/10.1109/ICDM.2008.17
35. Manevitz LM, Yousef M. One-class SVMs for document classification. Journal of machine Learning research. 2001;2(Dec):139-54.
36. Bushra AA, Yi G. Comparative analysis review of pioneering DBSCAN and successive density-based clustering algorithms. IEEE Access. 2021 Jun 14;9:87918-35. https://doi.org/10.1109/ACCESS.2021.3089036
37. Cea L, Puertas J, Pena L. Velocity measurements on highly turbulent free surface flow using ADV. Experiments in fluids. 2007 Mar;42(3):333-48. https://doi.org/10.1007/s00348-006-0237-3

© 2026 by the authors. Licensee SCU, Ahvaz, Iran. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 International (CC BY 4.0 license) (http://creativecommons.org/licenses/by/4.0/).