fluids

MDPI

Article

Measuring Turbulent Flows: Analyzing a Stochastic Process with Stochastic Tools

Evangelos Rozos [1,*], Jörg Wieland [2] and Jorge Leandro [2]

[1] Institute for Environmental Research & Sustainable Development, National Observatory of Athens, 15236 Athens, Greece
[2] Research Institute for Water and Environment, Faculty IV School of Science and Technology, University of Siegen, Paul-Bonatz-Str. 9–11, 57068 Siegen, Germany; jorge.leandro@uni-siegen.de (J.L.)
* Correspondence: erozos@noa.gr; Tel.: +30-210-3490-849

Abstract: Assessing drag force and Reynolds stresses in turbulent flows is crucial for evaluating the stability and longevity of hydraulic structures. Yet, this task is challenging due to the complex nature of turbulent flows. To address this, physical models are often employed. Nonetheless, this practice is associated with difficulties, especially in the case of high sampling frequency where the inherent randomness of velocity fluctuations becomes mixed with the measurement noise. This study introduces a stochastic approach, which aims to mitigate bias from measurement errors and provide a probabilistic estimate of extreme stress values. To accomplish this, a simple experimental setup with a hydraulic jump was employed to acquire long-duration velocity measurements. Subsequently, a modified first-order autoregressive model was applied through ensemble simulations, demonstrating the benefits of the stochastic approach. The analysis highlights its effectiveness in estimating the uncertainty of extreme events frequency and minimizing the bias induced by the noise in the high-magnitude velocity measurements and by the limited length of observations. These findings contribute to advancing our understanding of turbulent flow analysis and have implications for the design and assessment of hydraulic structures.

Keywords: turbulent flows; stochastic modeling; drag force; Reynolds stresses; experimental hydraulics; hydraulic jump; return period

Citation: Rozos, E.; Wieland, J.; Leandro, J. Measuring Turbulent Flows: Analyzing a Stochastic Process with Stochastic Tools. *Fluids* **2024**, *9*, 128. https://doi.org/10.3390/fluids9060128

Academic Editor: Ramesh Agarwal

Received: 28 April 2024
Revised: 21 May 2024
Accepted: 25 May 2024
Published: 30 May 2024

1. Introduction

Turbulent flow, a prominent phenomenon in many natural water systems, is characterized by vigorous mixing of adjacent layers and continuous random fluctuations of the velocity at any given point. The typical metric used to measure the intensity of these fluctuations at a specific point of a flow is the turbulence intensity (TI) factor, which, in the case of 1D flow, is defined as the ratio $\overline{U'^2}^{1/2}/\overline{U}$, where $U' = U - \overline{U}$ (i.e., the velocity fluctuation) and $\overline{U}$ is the mean flow velocity at the specific point [1,2]. Notably, TI simplifies the coefficient of variation, the ratio of the standard deviation to the mean of U. This connection highlights the inherent link between turbulent flow dynamics and probability theory. Since velocity (U) acts as a random variable, theoretically it has no upper and lower limits. Instead, the probability that U exceeds a specific threshold in a given period is estimated.

The importance of this perspective becomes more apparent when considering the implications the TI value has on the frequency of extreme values of velocity. While a typical TI value of unobstructed flows is 0.1 [3], TI values up to 0.4 are met in hydraulic jumps [4], whereas in the case of axisymmetric wakes behind various bodies, TI ranges from 0.3 to 1.1 [1]. To obtain an idea of how TI influences the frequency of extreme values, suppose an intense turbulent flow where TI equals 0.4 and U follows the normal distribution with $\overline{U} = 2$ m/s. Then, the standard deviation of U will be 0.8 m/s, and it can be easily

verified (e.g., in MATLAB 2021a numerical environment with: `norminv(0.999)*0.8 + 2`, which gives 4.47) the probability U to exceed 4.47 m/s is 0.1%. This means that with a probability of 50% (binomial distribution), 0.1% of the total duration of the turbulent flow, the flow velocity will be greater than 4.47 m/s, i.e., more than twice the average value. This example emphasizes the importance of understanding the statistical properties of the velocity in turbulent flows for thoroughly assessing their short- and long-term effects on nearby structures.

For this reason, various researchers have introduced statistical and stochastic models in the study of turbulent flows. For example, Flores-Vidriales et al. [5] have employed an advanced statistical analysis of measurements of the discharge of a river with a joint probability distribution function of the annual maximum and the number of events in a one-year block, in order to calculate the scour depth around bridge piers following the approach suggested by FHWA as it is described in the HEC-18 report [6]. They also examined a stochastic approach, where the scour is estimated by the SRICOS-EFA methodology [7] that takes as input a synthetic time series of discharge generated with an autoregressive moving integrated average stochastic model. Brandimarte et al. [8] have also suggested the use of a stochastic model, an autoregressive fractionally integrated moving average model, as a probabilistic approach to perform the analysis for contraction scour. In both these studies the stochastic and statistical models operated with daily time steps.

The previously mentioned studies have highlighted the effectiveness of statistical and stochastic methodologies in analyzing turbulent flows using coarse time steps. Note that, from here on, we will use the term "statistical" to refer to approaches based on fitting theoretical distributions to the data, and "stochastic" for methodologies based on numerical schemes that generate synthetic time series with statistical properties similar to those observed. The application of stochastic models with fine time steps has also been suggested by researchers, though in the sense of stochastic filtering, i.e., using "*partial observations and a stochastic model to make inferences about an evolving system*" [9]. For example, He and Liu employed a stochastic model in a data assimilation process that blends low-sampling-frequency data from image velocimetry with high-sampling-frequency data from an array of microphones [10]. Lo et al. used alternative stochastic models to evaluate the effect of instantaneous fluctuations on entrained particles when studying numerically the particle dispersion and deposition in turbulent pipe flows [11]. Olson has employed a similar approach where a stochastic model of a turbulent fluid feeds the differential equation of convection-dispersion to study the effect of fiber length on particle dispersion [12]. Tissot and Cavalieri proposed a stochastic formulation of the Navier–Stokes equations to study the propagation of coherent structures within a turbulent channel flow [13]. Vianna and Nichele proposed a stochastic model to represent the annular flow in a tubular reactor followed by a numerical simulation to generate sample paths fitting the residence time distributions [14]. Luhur et al. suggested a method based on stochastic differential equations to replace traditional look-up table methods used in wind turbine studies by simulating effective response dynamics of lift and drag forces [15].

Despite the availability of sophisticated numerical models, physical models remain essential for studying complex hydraulic processes such as turbulent flows and their impact on hydraulic structures like piers. Modern measuring equipment capable of high-sampling frequency velocity measurements is crucial for turbulent flow studies, providing detailed insights into turbulent behavior and enabling accurate analysis of turbulence statistics and dynamics. However, measurements at higher sampling frequencies are susceptible to increased noise influence. This influence is defined as the variance of the measurement one would obtain if a constant velocity were measured, which is proportional to the noise level and sampling frequency [16]. This noise-originating variance is compounded with the genuine variance of the measured velocity, potentially biasing the statistical properties of the measured time series. Therefore, in this study, we advocate for the use of a stochastic model to effectively reduce the negative influence of this noise, particularly in coping with the increased bias at high values, allowing for more reliable answers to

critical questions, such as determining the velocity that corresponds to a given probability of exceedance. To achieve this, an innovative stochastic model was employed in ensemble simulations following the k-fold concept. A relatively new statistical analysis tool, the climacogram, has been employed, which is ideal for studying the scale invariance of turbulence processes across different temporal scales [17]. Climacogram is the function of the variance versus time scale, wherein the time series is repeatedly aggregated to higher scales, and for each new time series obtained from each aggregation, the corresponding variance is calculated [18].

To validate this approach, we conducted an experiment in the Laboratory of Hydromechanics and Hydraulic Engineering at the University of Siegen. This experiment captured long-duration (36 min) flow measurements. We then employed the novel stochastic model specifically designed to handle data with significant skewness, which is a challenge in the analysis of turbulent flow stresses. The stochastic model is applied in ensemble simulations, leveraging k-fold cross-validation, a commonly used technique in statistics and machine learning for evaluating predictive models [19].

2. Materials and Methods

2.1. The Experimental Setup

The experimental setup, displayed in Figure 1a, consists of a flow in an open channel of the rectangular cross-section with a sluice gate. The experiment had the following characteristics: depth upstream of the sluice gate and downstream the jump 16 and 9.5 cm respectively, height of sluice gate opening 3.5 cm, slope −0.7 cm/1 m, probe location 34.5 cm from the sluice gate, discharge 3.43 l/s, channel width 8.5 cm, and total channel length 5 m. These characteristics correspond to Froude numbers of 0.20, 0.44, and 1.35 before the sluice gate, downstream the jump, and immediately after the sluice gate respectively, whereas the corresponding Reynolds numbers are 8885, 13,085, and 20,562, respectively. The TI at the location of the measurements was 0.20.

Figure 1. The experimental channel: (**a**) general view of the apparatus; (**b**) the velocimeter NORTEK Vectrino Plus.

The flow velocity was monitored with the acoustic velocimeter NORTEK Vectrino Plus, with a sampling frequency of 50 Hz, nominal maximum velocity of 4 m/s, and sampling volume height of 4 mm. The NORTEK Vectrino Plus employs the pulse-to-pulse coherent method to obtain velocity measurements. This technique relies on detecting the phase difference between two reflected pulses. While offering enhanced accuracy compared to standard Doppler-based systems, it does have limitations such as velocity ambiguity and phase wrap. Consequently, careful selection of the nominal maximum velocity is required [16]. The NORTEK Vectrino Plus consists of two components: the logger and the probe. While the unit can only accommodate one probe at a time, it supports two types of probes: the down-looking and the side-looking probe. Each probe is designed for specific

flow measurement applications. In this study, the side-looking probe was used (Figure 1b) following the suggestion of the user guide "*If the flow is dominantly 1D, the side-mounted configuration, where the transmitter is orientated perpendicular to flow direction, is preferable*".

2.2. Data Collection And Cleaning

The Vectrino recorded velocities in an XYZ coordinate system, where the X-axis aligns with the flow direction (providing measurements for U_1), and the Z-axis represents the vertical direction (providing measurements for U_3). The average signal-to-noise ratio (SNR) was 23.0 dB, while the average normalized correlation value was 74%. According to the Vectrino user guide [16], SNR should ideally be greater than 15 dB, and normalized correlations above 70% are considered sufficient for generating good-quality data. For 2160 s, a total of 108,000 measurements were obtained, of which 57,879 exhibited normalized correlation values below 70%.

Figure 2 displays the histograms of the raw data collected with the Vectrino velocimeter. The range of the horizontal axes in these histograms indicates the minimum and maximum values observed. This range is plausible only for the shape of the histogram of U_2.

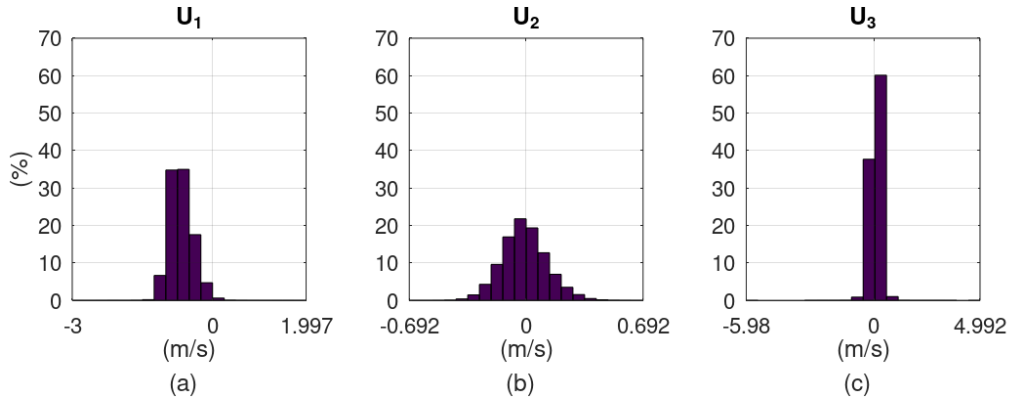

Figure 2. Histogram of velocity components along XYZ axes of the raw data: (**a**) U_1 bin size 0.25 m s^{-1}, (**b**) U_2 bin size 0.069 m s^{-1}, (**c**) U_3 bin size 0.549 m s^{-1}.

According to the Vectrino user guide [16], velocity limits and measurement uncertainty are fundamentally tied to the geometry of the deployment of the central transducer and the receiver arms. The velocity component in the direction of the central transducer, U_2 in this case, yields a lower measurement uncertainty. Additionally, the user guide states: "*Due to the design, the side-looking probe has higher instrument noise in the vertical than in the horizontal*".

For these reasons, we considered the U_2 measurements reliable as is, but we discarded the measurements of U_1 and U_3 having a normalized correlation below 70%. Linear interpolation was found to be the most efficient imputation method (see Appendix A). Random sample imputation and replacement with the mean value were found to introduce bias in the case of U_1. This occurred because velocities with higher magnitudes, tending to have measurements of lower normalized correlation, were mostly on one side of the histogram, the left side (Figure 2a), due to the longitudinal flow in the channel.

Figure 3 displays the spectrum of U_1 before and after filtering. The moving average is also presented in this figure with a yellow line. A moving average is a statistical method used to analyze time series data by creating a series of averages of different subsets of the full data set. The purpose of a moving average is to smooth out short-term fluctuations and highlight longer-term trends or cycles. The size of the window was set equal to 1/200 of the length of the spectrum time series.

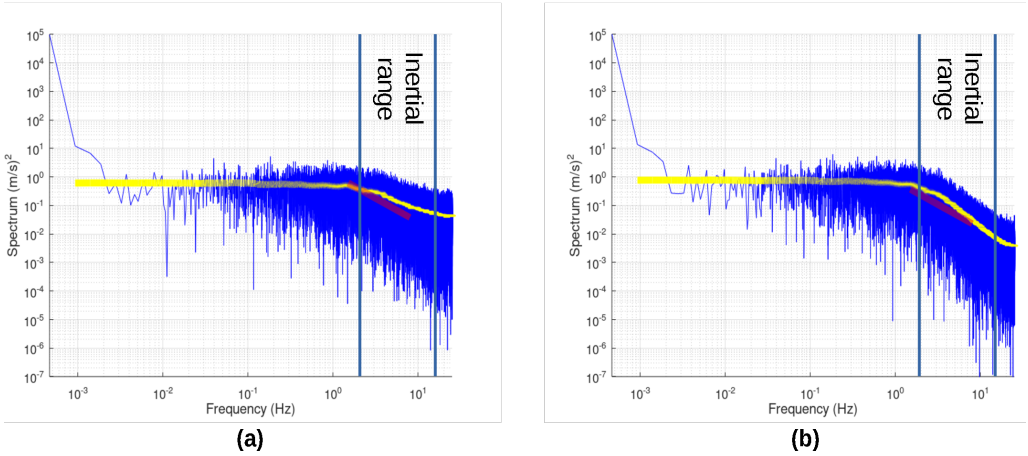

Figure 3. Spectrum of U_1 measurements: (**a**) raw data, (**b**) filtered data; includes a yellow line representing the smoothed spectrum using the moving average method; with a red line indicating the theoretical $-5/3$ slope.

Figure 4 displays the climacogram [18] of U_1 before and after filtering. The value of the Hurst coefficient that corresponds to the large aggregation scales is also displayed in this figure ($H = 0.85$ and $H = 0.84$ for the raw and filtered data respectively).

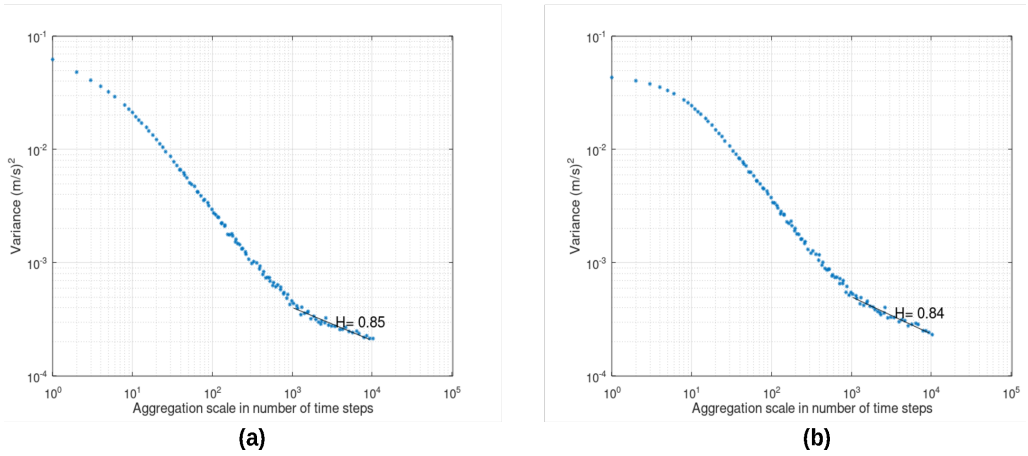

Figure 4. Climacogram of U_1 measurements: (**a**) raw data, (**b**) filtered data.

Figure 5 displays the histogram of the square of the velocity components measurements along the XYZ axes. The histograms suggest the existence of a significant skewness.

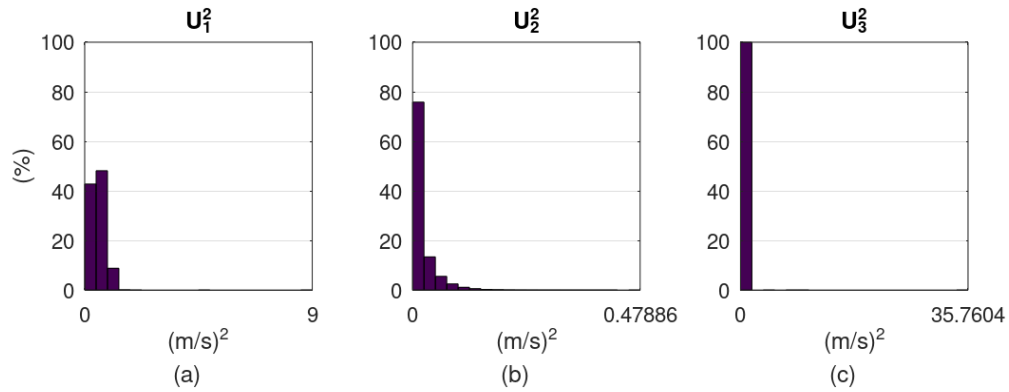

Figure 5. Histograms of U_i^2 obtained from the filtered measurements: (**a**) U_1^2 bin size 0.45 $m^2\,s^{-2}$, (**b**) U_2^2 bin size 0.024 $m^2\,s^{-2}$, (**c**) U_3^2 bin size 1.788 $m^2\,s^{-2}$.

Figure 6 displays the histogram of the absolute value of the product of the fluctuation of the velocity components. The absolute value is employed in order to set the focus on the magnitude, and not on the direction of the shear force. All histograms suggest the existence of a significant skewness.

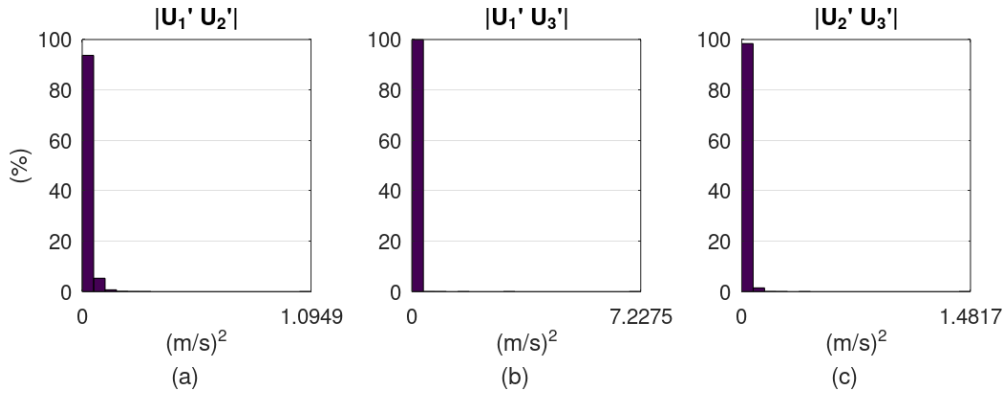

Figure 6. Histograms of $|U_i'U_j'|$ obtained from the filtered measurements: (**a**) $|U_1'U_2'|$ bin size 0.074 $m^2\,s^{-2}$, (**b**) $|U_1'U_3'|$ bin size 0.363 $m^2\,s^{-2}$, (**c**) $|U_2'U_3'|$ bin size 0.111 $m^2\,s^{-2}$.

From Figures 5 and 6 becomes evident that if a stochastic model is going to be applied directly on the U_i^2 or $|U_i'U_j'|$, then this model should be capable of reproducing the significant skewness suggested by these histograms.

2.3. The Stochastic Model

Rozos et al. [20] used a stochastic model to analyze the measurements of a turbulent flow obtained with an image velocimetry algorithm [21,22] applied to a video displaying a hydraulic jump. More specifically, a first-order autoregressive model (AR1) was used to analyze the velocity data obtained simultaneously at 100 locations of the hydraulic jump. In that study, Rozos et al. highlighted some limitations regarding this approach, most notably the inability of the classical AR1 model [23,24] to handle time series with significant skewness. However, in that study, the majority of the flow velocities that were analyzed by the stochastic model exhibited mostly a symmetric histogram, therefore the classical AR1 model handled the data successfully.

In this study, we propose a new methodology based on Rozos et al. [20]. We employed an acoustic velocimeter capable of high-frequency velocity measurements at a single point using the pulse-to-pulse coherent method. This measurement option was chosen for its superior accuracy compared to PIV and its capacity to handle extended time series, making it ideally suited for analyzing flow properties at a single point. After, we analyze the newly collected data with a stochastic model. Furthermore, unlike Rozos et al. [20], we focus on quantities derived from the velocities that are directly related to turbulent flow stresses, namely: U_1^2, U_2^2 and U_3^2 for drag forces, and $|U_1'U_2'|$, $|U_1'U_3'|$ and $|U_2'U_3'|$ for Reynolds stresses. Indices $_{1,2}$ and $_3$ represent directions along the flow, transverse, and vertical axes respectively. This approach aims to improve the model's accuracy by directly simulating the relevant stress quantities. As mentioned previously, some of these derived quantities exhibit significant skewness, posing a challenge for the classical AR1 model.

The basic equation of the univariate AR1 model is [23,24]:

$$x_t = \mu + a(x_{t-1} - \mu) + b\epsilon_t \tag{1}$$

where x_t is the time series value at time step t, ϵ_t is the innovation or error term, $a = r_1$ and $b^2 = s^2(1 - a^2)$; μ, s^2 and r_1 are the sample mean, variance and lag-1 autocorrelation respectively [23].

The previously described model generates a synthetic time series that effectively preserves the mean, variance, and autocorrelation characteristics. However, to ensure the

preservation of skewness, a specific approach was adopted in this study. For ϵ_t, we utilized a generalized Pareto [25] random number generator, of which the parameters (shape, scale, and location) were calibrated to have a mean value equal to 0, variance equal to 1, and a third moment equal to $\mu_3[\epsilon]$, as given by the following equation (see Equation (8) in [26]).

$$\mu_3[\epsilon] = b^{-3}(\mu_3[x] - \mu_3[ax]) \tag{2}$$

It is evident that the inclusion of the random variable ϵ introduces stochasticity into the synthetic data. This allows using Equation (1) to generate multiple equiprobable time series, i.e., ensemble simulations, enabling the assessment of uncertainty in the stochastic analysis. Taking advantage of this, the stochastic model was independently executed to produce numerous synthetic time series with an approach akin to k-fold cross-validation [19]. Employing a k-fold enables the model to enhance its generalization capabilities while improving insight into the overall modeling uncertainty. To accomplish this, the available data were partitioned into 10 equal subsets, and the stochastic model was applied twice to each subset, resulting in a total of 20 runs.

3. Results

Figure 7 displays the return period plots for U_1^2, U_2^2, and U_3^2. The horizontal red line indicates the average value of the observations. The black line represents the return period plot created using the filtered measurements, while the blue marks represent the return period plots of the 20 synthetic time series generated using the stochastic model outlined in Equation (1). For both black line and blue marks, the (empirical) Hazen plotting position method was employed (see Table 5.4 of [18]). The blue line represents the median of the 20 return period plots.

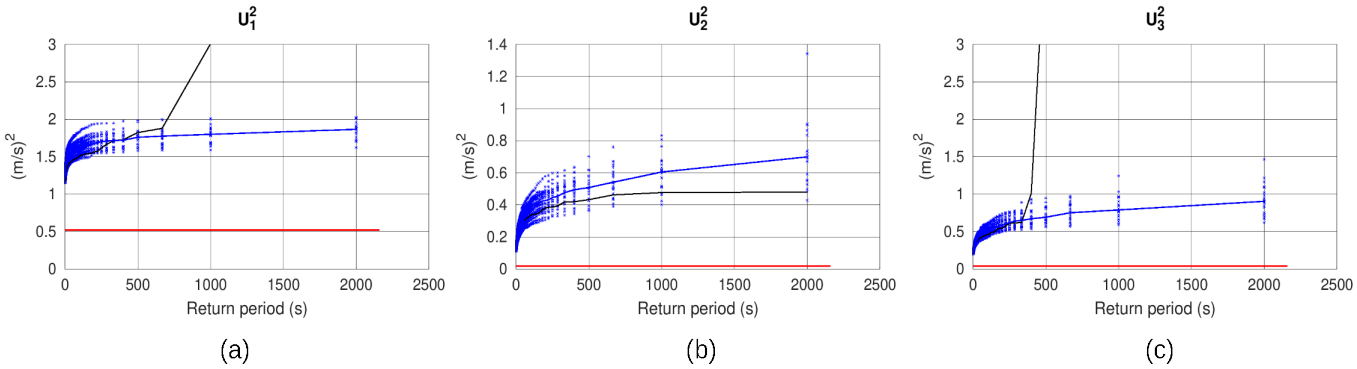

Figure 7. Return period plots of the square of the velocity components along XYZ axes: (**a**) U_1^2, (**b**) U_2^2, (**c**) U_3^2; with black line the return period of measurements; with red line the mean value of U_i^2 of the measurements; with marks in blue the 20 return periods of synthetic time series. The blue line represents the median of the 20 return period plots.

The value of the median of the return period plots from the synthetic U_1^2 values at 2000 s is 1.88 $m^2\,s^{-2}$. There are two values obtained from the filtered observations that are greater than the maximum value of the vertical axis of Figure 7a, namely, 4.90, and 9.00 $m^2\,s^{-2}$. The value of the median of the return period plots from the synthetic U_2^2 values at 2000 s is 0.71 $m^2\,s^{-2}$. The value of the median of the return period plots from the synthetic U_3^2 at 2000 s is 0.96 $m^2\,s^{-2}$. There are five values obtained from the filtered observations that are greater than the maximum value of the vertical axis of Figure 7c, namely, 3.73, 8.38, 8.96, 9.11, and 35.76 $m^2\,s^{-2}$.

Figure 8 displays the return period plots of $|U_1'U_2'|$, $|U_1'U_3'|$, and $|U_2'U_3'|$. The median of the return period plots of synthetic $|U_1'U_2'|$ at 2000 s is 0.84 $m^2\,s^{-2}$. There is one value obtained from the filtered observations that is greater than the maximum value of the vertical axis of Figure 8a, which is equal to 1.09 $m^2\,s^{-2}$. The median of the return period plots of synthetic $|U_1'U_3'|$ at 2000 s is 0.93 $m^2\,s^{-2}$. There are two values obtained from

the filtered observations that are greater than the maximum value of the vertical axis of Figure 8b, namely, 3.11, and 7.22 $m^2\,s^{-2}$. The median of the return period plots of synthetic $|U_2'U_3'|$ at 2000 s is 0.43 $m^2\,s^{-2}$. There is one value obtained from the filtered observations that is greater than the maximum value of the vertical axis of Figure 8c, which is equal to 1.48 $m^2\,s^{-2}$.

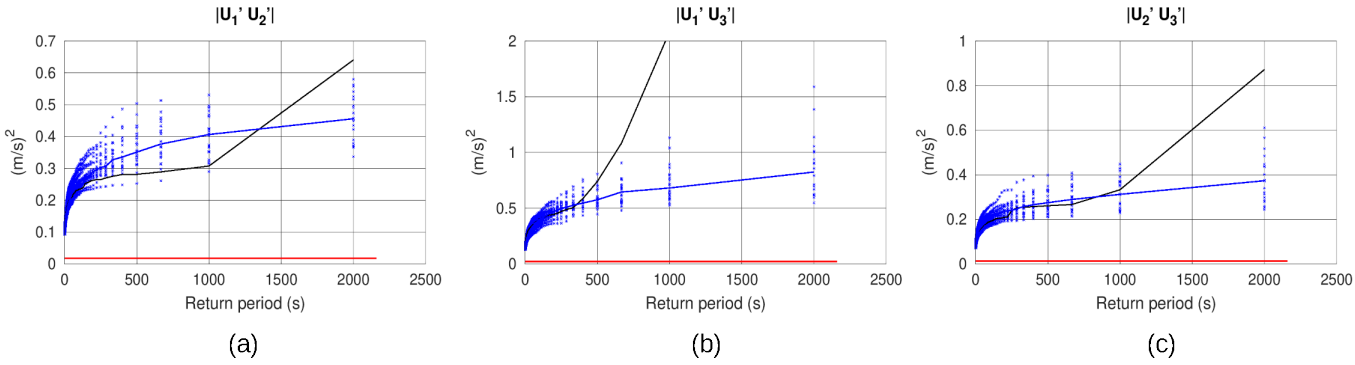

Figure 8. Return period plots of the product of the fluctuation of the velocity components: (**a**) $|U_1'U_2'|$, (**b**) $|U_1'U_3'|$, (**c**) $|U_2'U_3'|$; with black line the return period of measurements; with red line the mean value of the measurements; with marks in blue the 20 return periods of synthetic time series. The blue line represents the median of the 20 return period plots.

4. Discussion

Figures 2–6 highlight the significance of statistically analyzing the available data. Figure 2 initially suggests potential measurement errors in the raw data. Consequently, the raw data underwent filtration following the guidelines outlined in the Vectrino user manual.

Comparing Figure 3a with Figure 3b (as well as Figure 4a with Figure 4b) makes it evident that this filtration process effectively reduces data noise. Particularly noteworthy is the steeper slope within the inertial range in Figure 3b, indicating a noise reduction (similar to the findings in the study by Chowdhury et al. [27]). It is important to note that the filtration process should be executed cautiously, following experimentally derived guidelines (such as those employed in this study, provided by the user guide of the Vectrino velocimeter) or based on evidence-based decisions. For instance, Romagnoli et al. [28] implemented a more aggressive filtering approach by removing all spikes from the data, which were attributed to the presence of numerous bubbles in the flow at the measurement location. In our case study, we meticulously positioned the velocimeter probe in a location where the presence of bubbles in the flow was minimal.

Regarding the deviation of the slope in Figure 3b within the inertial range from the theoretical value of $-5/3$, this could be attributed either to the distance from the hydraulic jump (or the source of disturbance), as observed by Mukha et al. [4] and Buchhave and Velte [29], or to the rapidly rotating type of flow [30,31].

The climacogram shape of the filtered data in Figure 4 appears more plausible compared to that of the raw data (see Figure 9.5 of [18]). Specifically, it is expected that the climacogram curve will be almost horizontal on the left side of the plot, indicating minimal variance reduction with aggregation at small time scales. The scale at which the curve starts bending significantly downwards marks the minimum sampling frequency required to capture the stochastic structure of the studied process. In Figure 4b, this frequency is indicated to be approximately three times lower than the 50 Hz frequency used in this study, approximately 17 Hz. Subsequently, the slope becomes steeper before transitioning to a segment with a gentler slope. The increased slope in the middle segment of the curve is likely due to periodicity in the data, possibly caused by a stagnant wave upstream of the sluice gate (compare with the middle segment of Figure 9.5 of [18], which corresponds to the annual periodicity of river flow). The gentle slope of the final segment indicates significant persistence in the studied process. The Hurst coefficient H, which is dimensionless

and ranges from 0 to 1, is directly related to this slope s via the formula $s = 2H - 2$. The Hurst coefficient in Figure 4 is approximately 0.84, which is within the range of values reported in the literature. For example, for turbulent flow around a jet, Hurst coefficient values between 0.7 and 0.9 have been reported [32].

Lastly, Figures 5 and 6 offer valuable insights into the fundamental characteristics of the stress-related time series. Preserving these characteristics, including skewness, is crucial for the reliable analysis of the turbulent flow impacts.

The drag force and Reynolds stresses play a pivotal role in characterizing the stresses exerted on hydraulic structures. These are determined by the squared velocity U_i^2 and the product of velocity fluctuation components $|U_i'U_j'|$. In the design of hydraulic structures, obtaining accurate estimations of these quantities for a given probability of failure is crucial. This involves first computing the intensity and duration of the design discharge based on the probability of failure. Subsequently, the values of U_i^2 and $|U_i'U_j'|$—corresponding to the design discharge—need to be determined. Caution must be taken when deriving these values from a physical model to avoid potential inaccuracies resulting from simplistic methods. For instance, in the case of Figures 7 and 8, while the average value is significantly lower than the higher values, some extreme values are unrealistic, most likely due to measurement errors. To address these challenges adequately, it is advisable to report, for the given probability of failure, the U_i^2 or $|U_i'U_j'|$ value based on a return period plot, as demonstrated in Figures 7 and 8.

While one might argue against the necessity of a stochastic model, return period plots derived directly from data can be prone to errors, as demonstrated in Figures 7 and 8. Measurement errors, particularly in the range of high values, can distort the shapes of these plots, undermining their reliability. Moreover, even with minimal noise in measurements, a statistical or stochastic approach becomes essential when the return period corresponding to the probability of failure exceeds the duration of observations. To illustrate this point, let us consider an example. Suppose the data collected from the physical model corresponds to an extreme event with an expected duration of 2000 s, and we need to find the value of U_1^2 corresponding to a 4% probability of failure. The return period of observations in Figure 7a exhibits an unrealistic sharp rise in the high-value range, likely due to errors in the measurement. Therefore, to obtain a more reliable estimate, we should use the median value of the return period plots of the synthetic time series at 2000 s, which is 1.88 $m^2\,s^{-2}$. To calculate the probability of failure for this value for 2000 s (note the intentional coincidence in this example of the maximum obtainable return period from the available data and the duration of the design event), we can use the following formula:

$$P = 1 - (1 - 1/T)^n \tag{3}$$

where T is the return period and n is the number of time steps (in the same time unit as T) corresponding to the duration of the design event.

Using Equation (3) with $T = 2000$ s and $n = 2000$ yields $P = 63.2\%$, which is much higher than the target of 4%. To determine the value of U_1^2 corresponding to $P = 4\%$, it is necessary to generate a synthetic time series with a sufficiently large length. This length must be adequate for calculating the U_1^2 value corresponding to a return period of 50,000 s (i.e., setting $T = 50,000$ s and $n = 2000$ in Equation (3) results in a $P = 0.04$). For the specific time step in this study (1/50 s), this suggests that a synthetic time series of a minimum length of 2,500,000 values is required (if ensemble simulations, then multiple time series of this length). This implies that even if the return period plots were not distorted in the range of high values of return periods (see black lines from Figure 7a,c), observations would still need to be 25 times longer than what is currently available (i.e., instead of running the experiment for 36 min, to run it for 15 h). Running the stochastic model to generate a time series of sufficient length (more than 4,000,000 values), we estimated the value of U_1^2 corresponding to a return period of 50,000 s to be 2.1 $m^2\,s^{-2}$.

Figure 7a,c illustrates how measurement errors can significantly impact return period plots derived from the observed data, leading to unreliable estimates. To address this issue, ensemble simulations of a stochastic model with the k-fold method offer a solution. By generating multiple synthetic datasets (in this case 20), we can mitigate the impact of outliers and obtain a set of plots for the return period instead of a single one. This set provides valuable insights into the uncertainty associated with the measurements. The median values from this set offer robust estimates that are less affected by measurement errors, ensuring safer or economical design considerations.

Regarding the choice of observation and simulation time step (sampling frequency), while larger time steps tend to filter out observational noise, they may also smooth out the flow bursts recorded in observations. Bursts are a significant characteristic of turbulent flows, particularly in TI values of medium range [33,34]. Various probabilistic approaches, such as Gessler's study on entrainment, are based on the concept of bursts [35]. Gessler proposed a probabilistic approach to partially offset the smoothing effect caused by temporal aggregation resulting from the finite frequency of sampling in any physical process. This is achieved by increasing the "confidence" for higher values when the standard deviation of the time series is lower. Similar approaches could be explored to address the smoothing effect on other types of stresses exerted by turbulent flows on hydraulic structures.

As noted by other researchers (e.g., Luhur et al. [15]), gaining deeper insight into the underlying process requires proper representation of the conditional probability density function (PDF) of a stochastic process $\underline{X}$, $P(\underline{X}(t+\tau) < x|\underline{X}(t))$, as it contains comprehensive information regarding the process evolution dynamics. However, it has been observed that the stochastic model represented by Equation (1) may not adequately capture the decay of the autocorrelation function in stochastic processes that exhibit significant persistence (Koutsoyiannis et al. [18]). In such cases, a generalized moving average scheme may be a more appropriate option for stochastic simulations.

5. Conclusions

This study utilized an experimental setup featuring a hydraulic jump downstream of a sluice gate. Velocity vector measurements were collected using an acoustic velocimeter, positioned 34.5 cm downstream the sluice gate. Due to the probe geometry and the nature of the flow, the measurement accuracy was non-isotropic. Consequently, raw data of velocity components along the three axes (longitudinal, transverse, and vertical) underwent different pre-processing methods. Despite pre-processing, a few large values persisted, prompting further investigation into their nature (extreme values or measurement errors) through stochastic analysis. Ensemble simulations employing a stochastic model generated a series of synthetic time series related to typical turbulent flow stresses (drag force and Reynolds stresses). Empirical return period plots derived from these synthetic time series yielded crucial insights into turbulent flow measurement analysis.

The suggested methodology can be applied independently to an arbitrary number of different locations within the hydraulic flow. While we cannot obtain measurements simultaneously at multiple locations with the acoustic velocimeter, analyzing data from multiple locations could allow us to capture variations in flow parameters across different points. The findings of this study hold broad applicability and are summarized below:

- When monitoring hydraulic phenomena, careful consideration of measurement errors is crucial, especially in regions of high values critical for design. Employing statistical recognition and meticulous filtering is essential for ensuring accurate and reliable results.
- Maximum values of such measurements may be significantly inflated due to measurement errors. Therefore, probabilistic approaches, such as return period plots, are essential for a comprehensive analysis.
- Return period plots derived from observations may exhibit significant bias at large return periods. Thus, adopting stochastic approaches is necessary to construct more reliable plots and at the same time estimate their uncertainty.

- Stochastic approaches are indispensable for supporting probabilistic assessment of extreme values with return periods significantly larger than the length of observations.

Finally, it should be noted that the proposed methodology and findings are broadly applicable to case studies involving high-frequency measurements of any kind.

Author Contributions: Conceptualization, E.R., J.L., and J.W.; methodology, E.R. and J.W.; software, E.R.; validation, E.R.; formal analysis, E.R.; investigation, E.R. and J.W.; resources, J.L. and J.W.; data curation, E.R.; writing—original draft preparation, E.R.; writing—review and editing, J.L.; visualization, E.R.; supervision, E.R.; project administration, E.R. and J.L.; funding acquisition, E.R. and J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no direct external funding.

Data Availability Statement: The original data presented in this study are openly available on the web page of the first author's DAAD scholarship at https://hydronoa.gr/hydronoa/DAAD-Scholarship.html (accessed on 28 May 2024).

Acknowledgments: The authors wish to thank the Deutscher Akademischer Austauschdienst (DAAD) for covering the travel costs necessary for the completion of this study.

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

Abbreviations

The following abbreviations are used in this manuscript:

AR1	first-order autoregressive model
SNR	signal-to-noise ratio
TI	turbulence intensity

Appendix A

Data filtering was performed with the following MATLAB 2021a code. Columns 15 to 18 of the matrix `Data` contain the normalized correlation values along the X, Y, and Z axes. The velocity measurements are kept in the matrix `Velocities`.

```
missingvalues= sum(Data(:,15:18)<70,2) >0;
for i=[1,3]
  Velocities(:,i)= interp1(find(~missingvalues), ...
  Velocities(~missingvalues,i), 1:length(Velocities), ...
  'linear');
end
```

References

1. Pope, S.B. *Turbulent Flows*; Cambridge University Press: New York, NY, USA, 2000.
2. Ahmad, S.; Takana, H.; Ali, K.; Akhtar, Y.; Hassan, A.M.; Ragab, A.E. Role of localized magnetic field in vortex generation in tri-hybrid nanofluid flow: A numerical approach. *Nanotechnol. Rev.* **2023**, *12*, 20220561. [CrossRef]
3. Basse, N.T. Turbulence Intensity Scaling: A Fugue. *Fluids* **2019**, *4*, 180. [CrossRef]
4. Mukha, T.; Almeland, S.K.; Bensow, R.E. Large-Eddy Simulation of a Classical Hydraulic Jump: Influence of Modelling Parameters on the Predictive Accuracy. *Fluids* **2022**, *7*, 101. [CrossRef]
5. Flores-Vidriales, D.; Gómez, R.; Tolentino, D. Stochastic Assessment of Scour Hazard. *Water* **2022**, *14*, 273. [CrossRef]
6. Arneson, L.; Zevenbergen, L.; Lagasse, P.; Clopper, P. *Evaluating Scour at Bridges*; Technical Report; Federal Highway Administration: Washington, DC, USA, 2013.
7. Briaud, J.L. *Pier and Contraction Scour in Cohesive Soils*; Transportation Research Board: Washington, DC, USA, 2004; Volume 516.
8. Luigia Brandimarte, P.D.; Montanari, A. A probabilistic approach to the analysis of contraction scour. *J. Hydraul. Res.* **2006**, *44*, 654–662. [CrossRef]
9. Bain, A.; Crisan, D. *Fundamentals of Stochastic Filtering*; Springer: New York, NY, USA, 2009.
10. He, C.; Liu, Y. Time-resolved reconstruction of turbulent flows using linear stochastic estimation and sequential data assimilation. *Phys. Fluids* **2020**, *32*, 075106. [CrossRef]
11. Lo, C.; Bons, J.; Yao, Y.; Capecelatro, J. Assessment of stochastic models for predicting particle transport and deposition in turbulent pipe flows. *J. Aerosol. Sci.* **2022**, *162*, 105954. [CrossRef]

12. Olson, J.A. The motion of fibres in turbulent flow, stochastic simulation of isotropic homogeneous turbulence. *Int. J. Multiph. Flow* **2001**, *27*, 2083–2103. [CrossRef]
13. Tissot, G.; Cavalieri, A.V.; Memin, E. Stochastic linear modes in a turbulent channel flow. *J. Fluid Mech.* **2021**, *912*, A51. [CrossRef]
14. Vianna, A.; Nichele, J. Modeling an annular flow tubular reactor. *Chem. Eng. Sci.* **2010**, *65*, 4261–4270. [CrossRef]
15. Luhur, M.R.; Peinke, J.; Schneemann, J.; Wächter, M. Stochastic modeling of lift and drag dynamics under turbulent wind inflow conditions. *Wind Energy* **2015**, *18*, 317–337. [CrossRef]
16. *The Comprehensive Manual for Velocimeters Vector | Vectrino | Vectrino Profiler*; NORTEK AS: Rud, Norway, 2018.
17. Friedrich, R.; Peinke, J.; Sahimi, M.; Reza Rahimi Tabar, M. Approaching complexity by stochastic methods: From biological systems to turbulence. *Phys. Rep.* **2011**, *506*, 87–162. [CrossRef]
18. Koutsoyiannis, D. *Stochastics of Hydroclimatic Extremes—A Cool Look at Risk*, 3rd ed.; KALLIPOS: Athens, Greece, 2023.
19. Norvig, P.R. *Artificial Intelligence, A Modern Approach*, 3rd ed.; Prentice Hall: Upper Saddle River, NJ, USA, 2002.
20. Rozos, E.; Leandro, J.; Koutsoyiannis, D. Stochastic Analysis and Modeling of Velocity Observations in Turbulent Flows. *J. Environ. Earth Sci.* **2024**, *6*, 4–56. [CrossRef]
21. Rozos, E.; Mazi, K.; Koussis, A.D. Probabilistic Evaluation and Filtering of Image Velocimetry Measurements. *Water* **2021**, *13*, 2206. [CrossRef]
22. Rozos, E.; Mazi, K.; Lykoudis, S. On the Accuracy of Particle Image Velocimetry with Citizen Videos—Five Typical Case Studies. *Hydrology* **2022**, *9*, 72. [CrossRef]
23. Salas, J.D. Chapter 19: Analysis and modelling of hydrological time series. In *Handbook of Hydrology*; Maidment, D., Ed.; McGraw-Hill: New York, NY, USA, 1993; Volume 19.
24. Matalas, N.; Wallis, J. Generation of synthetic flow sequences. In *Systems Approach to Water Management*; Biswas, A., Ed.; McGraw-Hill: New York, NY, USA, 1976.
25. Kotz, S.; Nadarajah, S. *Extreme Value Distributions: Theory and Applications*; Imperial College Press: London, UK, 2000.
26. Koutsoyiannis, D. Optimal decomposition of covariance matrices for multivariate stochastic models in hydrology. *Water Resour. Res.* **1999**, *35*, 1219–1229. [CrossRef]
27. Chowdhury, S.; Sen, D.; Dey, S. Submerged wall jet on a macro-rough boundary: Turbulent flow characteristics and their scaling laws. *Environ. Fluid Mech.* **2023**, 1–24. [CrossRef]
28. Romagnoli, M.; Carvalho, R.; Leandro, J. Turbulence characterization in a gully with reverse flow. *J. Hydraul. Eng.* **2013**, *139*, 736–744. [CrossRef]
29. Buchhave, P.; Velte, C.M. Measurement of turbulent spatial structure and kinetic energy spectrum by exact temporal-to-spatial mapping. *Phys. Fluids* **2017**, *29*, 085109. [CrossRef]
30. Heller, V. Self-similarity and Reynolds number invariance in Froude modelling. *J. Hydraul. Res.* **2017**, *55*, 293–309. [CrossRef]
31. Baroud, C.N.; Plapp, B.B.; She, Z.S.; Swinney, H.L. Anomalous self-similarity in a turbulent rapidly rotating fluid. *Phys. Rev. Lett.* **2002**, *88*, 114501. [CrossRef] [PubMed]
32. Dimitriadis, P.; Papanicolaou, P. *Statistical Analysis of Positively Buoyant Turbulent Jets*; EGU General Assembly: Vienna, Austria, 2012.
33. Vanoni, V.A. *Measurements of Critical Shear Stress for Entraining Fine Sediments in a Boundary Layer*; U. S. Public Health Service: North Bethesda, MD, USA, 1964.
34. Hydraulics, I. Scour at Bridge Model Pier. 2010. Available online: https://www.youtube.com/watch?v=48S_k6qAmsY (accessed on 6 April 2024).
35. Gessler, J. Self-stabilizing tendencies of alluvial channels. *J. Waterw. Harb. Coast. Eng. Div.* **1970**, *96*, 235–249. [CrossRef]