Journal of Hydraulic Research

ISSN: 0022-1686 (Print) 1814-2079 (Online) Journal homepage: http://www.tandfonline.com/loi/tjhr20

Statistical analysis methods for transient flows – the dam-break case

Rui Aleixo, Sandra Soares-Frazão & Yves Zech

To cite this article: Rui Aleixo, Sandra Soares-Frazão & Yves Zech (2018): Statistical analysis methods for transient flows – the dam-break case, Journal of Hydraulic Research

To link to this article: https://doi.org/10.1080/00221686.2018.1516700

Published online: 30 Oct 2018.

Journal of Hydraulic Research, 2018
https://doi.org/10.1080/00221686.2018.1516700

Research paper

Statistical analysis methods for transient flows – the dam-break case

RUI ALEIXO (IAHR Member), Researcher, *GHT Photonics, Italy*
Email: ferreira_aleixo@yahoo.co.uk (author for correspondence)

SANDRA SOARES-FRAZÃO (IAHR Member), Professor, *Institute of Mechanics Materials and Civil Engineering, Université catholique de Louvain, Louvain-la-Neuve, Belgium*
Email: sandra.soares-frazao@uclouvain.be

YVES ZECH (IAHR Member), Emeritus Professor, *Institute of Mechanics Materials and Civil Engineering, Université catholique de Louvain, Louvain-la-Neuve, Belgium*
Email: yves.zech@uclouvain.be

ABSTRACT
Transient flows differ from steady flows in the sense that their dynamic variables, e.g. velocity, change with time. This imposes several constraints both on the measurement techniques to use as well on the data processing. The dam-break flow, as an example of severe transient flow, is used in this paper as test case for the application of different statistical methods identified in the literature: the common statistical tools (e.g. time average, standard deviation), the application of a Fourier-series-based method and the ensemble statistics approach. Data were collected using a fast acquisition camera and particle image velocimetry was used to obtain quantitative information about the velocity field of a set of seven dam-break flows. From the analysis made in this paper it was possible to identify in which conditions each statistical method can be applied to the dam-break flow.

Keywords: Dam-break; data analysis; particle image velocimetry; statistical analysis; transient flows

1 Introduction

Transient flows are ubiquitous in nature, occurring naturally in many circumstances: from valve openings to flash floods. A dam-break flow is an extreme example, in which a large amount of water is released in a short time interval. This kind of flow can have disastrous consequences in the natural landscape as it was seen in many past events (ICOLD, 1995). It is known that turbulence plays an important role in many phenomena of fluid mechanics, in particular in sediment transport; however, due to its severe transient character not many studies are found regarding the turbulence of dam-break flows.

On the general subject of turbulent analysis in time-dependent flows different studies have been presented in the past. For example, in the case of duct oscillatory flows Jensen, Sumer, and Fredsøe (1989) presented laser Doppler velocimetry (LDV) measurements of turbulent variables; Tardu, Binder, and Blackwelder (1994) performed a similar analysis of the turbulence in the flow subjected to forced oscillations. In the open-channel flows domain, Song and Graf (1996) made measurements of both velocity and turbulence profiles using acoustic Doppler velocity profiler (ADVP), and Nezu, Kadota, and Nakagawa (1997) analysed the turbulent structure of unsteady open channel flows with LDV. Later, Dey, and Lambert (2005) studied the Reynolds and bed shear stress in non-uniform unsteady channel flows and proposed a theoretical model that was well fitted to the Song (1994) dataset.

In the particular case of dam-break flows, despite the many studies found in the literature, not many of them have focused on flow statistics. In the past, the focus of many dam-break studies was the wave celerity and water depth evolution (e.g. Bell, Elliot, & Hanif Chaudry, 1992; Dressler, 1952; Fraccarollo & Toro, 1995; Ozmen-Cagatay & Kocaman, 2012; Ozmen-Cagatay, Kocaman, & Guzel, 2014; Reichstetter & Chanson, 2013). Indeed, from a practical point of view, the more significant variables of this problem are wave-celerity and water-depth evolution. However, the lack of suitable experimental techniques did not allow for detailed statistical analysis of the flow. With the development of imaging techniques and quantitative image processing techniques like particle image

Received 2 December 2016; accepted 23 August 2018/Currently open for discussion.

ISSN 0022-1686 print/ISSN 1814-2079 online
http://www.tandfonline.com

velocimetry (PIV), it is now possible from flow images not only to determine the water level evolution but also to obtain the velocity field in the dam-break flow. Different authors have already proposed measurements of the dam-break flow velocity field, e.g. Aleixo, Soares-Frazão, and Zech (2011), who used particle tracking velocimetry (PTV) and Oertel and Bung (2012) who used PIV. These works considered a pure hydrodynamic dam-break flow, that is, a dam-break flow over a smooth fixed bed. Regarding dam-break flow over a bed made of mobile sediments, measurements of sediment wave-celerity were presented by Leal, Ferreira, and Cardoso (2006, 2009); velocity measurement in the sediment layer were made by Capart and Young (1998) and by Spinewine and Zech (2007) and simultaneous measurements of water and sediment layer velocity fields have been presented by Aleixo, Spinewine, Soares-Frazão, and Zech (2010) considering a bed made of sand, and by Aleixo (2011) considering a bed made of PVC pellets. Aleixo, Zech, and Soares-Frazão (2012) presented a study on the statistical analysis of the dam-break flow, using a fast acquisition camera and repeating the experiment several times.

In this paper, different methods of transient flow analysis from the literature are applied to the dam-break case database of Aleixo et al. (2012). Using data from this database, a statistical analysis of the dam-break flow velocity is made, and the different methods are analysed. Applying statistical methods for the analysis of a dam-break flow can be a way to investigate its turbulence. It is therefore important to know the limitations of such statistical methods when applied to the particular case of the dam-break flow.

2 Statistical analysis of dam-break flow variables

The dam-break flow is a gravity driven flow. The idealized dam-break flow over a smooth fixed bed, described by Ritter (1892), is schematized in Fig. 1. This is a flow that has strong time dependence especially in the first instants. From Ritter's analysis, it is possible to determine both a time scale, t_0, and velocity scale, c_0, scales for this flow. One has respectively:

$$t_0 = \sqrt{\frac{h_0}{g}} \tag{1}$$

and

$$c_0 = \sqrt{g h_0} \tag{2}$$

where g is the gravity acceleration and h_0 is the initial water height in the upstream reservoir. The natural time scale, t_0, is important to use as reference for the sampling time needed to acquire information about the flow variables.

To analyse this flow from a statistical point of view, different methods, often applied to transient phenomena, are described in the following sections: single realization statistics (Section 2.1), Fourier analysis (Section 2.2) and the ensemble statistics (Section 2.3).

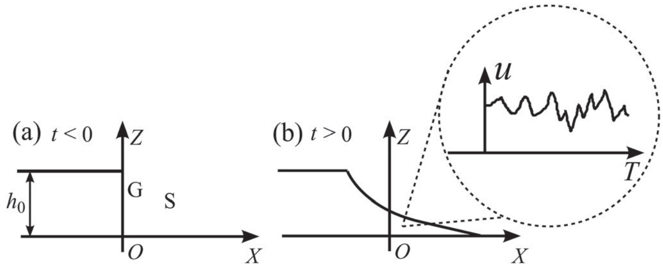

Figure 1 (a) Initial conditions prior to the dam-break. G refers to the downward moving gate location and S indicates the test section where the measurements are carried out (b). After the dam-break the focus is placed on the point wise velocity measurements and their time dependence

2.1 *Single realization statistics*

In order to describe a stationary random process, it is possible to define for one realization of a generic flow variable, $\phi(t)$, the time average $\mu_{\phi\propto}$ as:

$$\mu_{\phi\infty} = \lim_{T_s\to\infty} \frac{1}{T_s}\int_0^{T_s} \phi(t)\mathrm{d}t \tag{3}$$

where T_s is the sample duration of ϕ and $\mu_{\phi\infty}$ will be here referred, in the discrete case, as the mean of the population of variable ϕ. If ϕ is a stationary random process, $\mu_{\phi\infty}$ does not change over different sampling periods. If, in addition, the average value of different realizations (i.e. different runs of the experiments) is the same as the previously defined time average, the process is said to be ergodic. The main advantage of an ergodic process is that one realization of a given process could be sufficient, thus reducing significantly the data acquisition process. In practice, the sample duration, T_s is not infinite and thus, instead of a complete realization, ϕ, one will have a sample of it, usually in a form of time series. The mean of a sample can be written as:

$$\mu_\phi(t_i) = \frac{1}{T_s}\int_{t_i}^{t_i+T_s} \phi(t)\mathrm{d}t \tag{4}$$

where T_s is the sample duration of ϕ and it must be chosen to obtain a converged value of the mean. If $\mu_\phi(t_i)$ does not depend on t_i and T_s the sample is said to be stationary. Based on the previous definitions, the expressions for higher order moments are easily derived. For example, the variance of a sample, ${\sigma_\phi}^2$, is given by:

$$\sigma_\phi^2(t_i) = \frac{1}{T_s}\int_{t_i}^{t_i+T_s} (\phi(t) - \mu_\phi)^2\mathrm{d}t \tag{5}$$

It becomes also possible to calculate the cross-momentums of two statistical variables, ϕ and θ as:

$$\psi_{\phi\theta}(t_i) = \frac{1}{T_s}\int_{t_i}^{t_i+T_s} (\phi(t) - \mu_\phi)(\phi(t) - \mu_\theta)\mathrm{d}t \qquad (6)$$

Equations (3)–(6) are easily adapted to the discrete case by replacing the integrals by discrete sums over the measurement time interval. The number of terms in this sum will depend on the data acquisition frequency that will result in different possible numbers of elements in the considered sample. When dealing with samples it is important to know what the ideal size should be to obtain meaningful statistics. An answer to this question can be found, for example, in the turbulence analysis of stationary flows. For the particular case of velocity measurements in turbulent flows, Yanta and Smith (1973) made an investigation about the sample lengths required for turbulence measurements. According to those authors the sample length was found to be a function of turbulence intensity. Following the study of Yanta and Smith (1973), Hitching and Lewis (1999) proposed the following expression for the sample size, N:

$$N = \beta\left(\frac{\mu_{\phi\infty}}{\mu_{\phi\infty} - \mu_\phi}\right)^2 \frac{\sigma_\phi}{\mu_\phi} \qquad (7)$$

where $\mu_{\phi\infty}$ is the mean of the population, μ_ϕ is the mean of the sample, σ_ϕ is the standard deviation of the sample and β is a constant that is a function of the confidence level. According to Hitching and Lewis (1999) $\beta = 4$ for a confidence level of 95%. Bates and Hughes (1976, 1977) for turbulent flow in a pipe proposed $N = 3500$. The number of elements in a sample is also important to ensure a small statistical error. In a similar way, following Lumley (1979), Durst, Kikura, Lekakis, Jovanović, and Ye (1996) used the following expression to determine the required number of elements in a sample related to variable ϕ to achieve a given statistical error ϵ_ϕ:

$$\frac{\varepsilon_\phi}{\mu_\phi} = \sqrt{\frac{2}{N}}\frac{\sigma_\phi}{\mu_\phi} \qquad (8)$$

Durst et al. (1996) used $N = 3.5 \times 10^4$ for $\epsilon_\phi/\mu_\phi = 0.3\%$ and $\sigma_\phi/\mu_\phi = 40\%$.

For the case of the Reynolds stresses measurements Hitching and Lewis (1999), following Yanta and Smith (1973) and Bates and Hughes (1976), proposed a value $N \geq 3000$.

Considering an acquisition frequency of 1 kHz, a sampling duration of 3.5 s would be needed to obtain 3500 data points as proposed by Bates and Hughes (1976), or 35 s to obtain the 35,000 data points following Durst et al. (1996). This highlights the fact that, in order to obtain such large numbers of elements in a short period of time, very large acquisition frequencies are needed. However, the acquisition frequency is often limited by the measurement equipment characteristics.

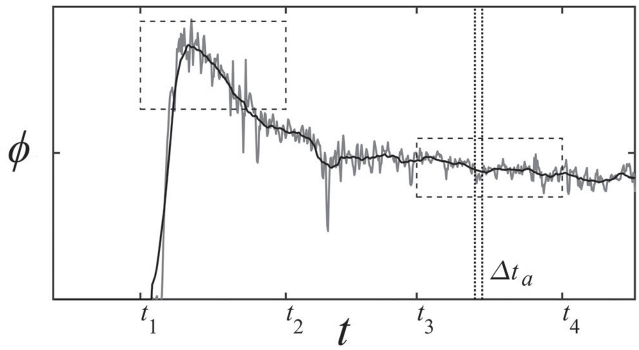

Figure 2 Exemplifying the transient behaviour of a generic variable ϕ. Between t_1 and t_2 the mean value changes with time, whereas between t_3 and t_4 the mean value is quasi-independent of the sampling time

In the case of a transient flow, additional care must be taken with the sampling period because of the time variation of the variables. A too-large sampling time may smooth out the transient characteristics of the phenomenon; a too-small interval may not be significant for statistical purposes. It is however sometimes possible to identify, within a transient time-series, a time interval where the flow can be assumed locally stationary. This is illustrated in Fig. 2. If in the time interval $\Delta t_{43} = t_4 - t_3$ the time series does not vary too much, and if the sampling time interval Δt_a is much smaller than Δt_{43}, it is then possible to determine the statistics in the time interval ΔT_{43} by means of Eqs (3)–(6). This method is thus not the best appropriate for fast varying variables but can be applied if locally stationary time intervals can be identified in the flow.

2.2 *Fourier component method*

The Fourier-component method allows to determine the mean value from just one realization of a random transient variable in particular when the mean value changes in time as depicted in Fig. 2 between t_1 and t_2. As reported in Nezu et al. (1997) this method is quite suitable for flood flows. Let ϕ_i with $i = 0, 1, 2, \ldots, (N-1)$ be the instantaneous values of a generic variable. Using a discrete Fourier series, it is possible to give an approximation for the mean value by using only the frequency components lower than a threshold frequency f_c:

$$f_c = \frac{(m-1)}{2T_h} \qquad (9)$$

where T_h is a convenient time scale and m is the number of Fourier terms that are used for the calculation of the mean value. As an example, Nezu et al. (1997) used the time period of the measurements of a hydrograph as time scale T_h. The mean velocity component can then be calculated as follows:

$$\mu_\phi = \frac{1}{2}a_0 + \sum_{k=1}^{(m-1)/2} a_k\cos(\omega_{ik}) + b_k\sin(\omega_{ik}) \qquad (10)$$

with the following definitions:

$$a_k = \frac{2}{N}\sum_{i=0}^{N-1}\phi_i \cos(\omega_{ik}) \tag{11}$$

$$b_k = \frac{2}{N}\sum_{i=0}^{N-1}\phi_i \sin(\omega_{ik}) \tag{12}$$

and

$$\omega_{ik} = 2\pi\frac{i}{N}k \quad i = 1, 2, \ldots, m \tag{13}$$

The cut-off frequency f_c has to be reasonably chosen. According to Nezu et al. (1997), f_c has to be much smaller than the turbulent burst frequency. Also, those same authors used $m = 7$ in their work. This value will be considered in the present work, and the sensitivity of the results to the value of m will be discussed.

2.3 *Ensemble statistics*

Ensemble statistics can be defined as statistics obtained over different realizations of a same phenomenon, i.e. different runs of a given experiment. Let $\phi_1, \phi_2, \ldots \phi_n$ be an example of an ensemble of realizations of the generic variable ϕ. It is possible to define an ensemble average in a specific instant, t_1, as (Bendat & Piersol, 2011):

$$< \mu_\phi(t_1) >= \lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}\phi_k(t_1) \tag{14}$$

where N is the number of realizations and ϕ_k is each realization of variable ϕ at time t_1. This procedure is depicted in Fig. 3. In the case of a stationary flow, if the process can be considered as ergodic, the average value obtained at a given time from different runs (the ensemble average) and the time average calculated on a single run should be equal.

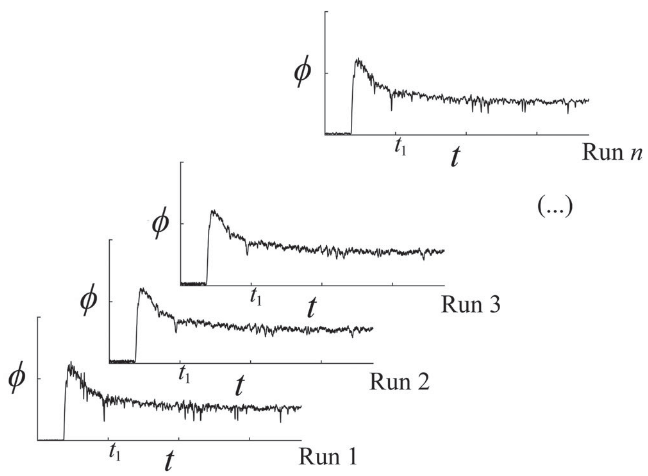

Figure 3 Exemplifying the ensemble average procedure. Several runs of the same experiment are performed. The statistical analysis is made in the ensemble of runs

The procedure of ensemble averages was used by Song and Graf (1996) in the analysis of open channel transient flows. Equations (4)–(6) can be applied to determine higher order statistical moments, considering the number of runs instead of the number of element in a sample over a given time interval. However, to obtain meaningful results, the experiments need to be repeated many times, which constitutes a serious limitation of this procedure. If the flow variables can be considered as stationary over a short time interval around the considered time t_1, then a combination of the single realization statistical method and the ensemble average method can be applied, in order to increase the number of data points in the calculation of the flow statistics.

3 Experimental set-up and measurement techniques

In this section the experimental set-up, the tested flow conditions and the measurement techniques are described.

3.1 *The dam-break channel and experimental conditions*

To perform dam-break simulations at a laboratorial scale the dam-break flume of the Civil Engineering Department of the Institute of Materials, Mechanics and Civil Engineering, Université catholique de Louvain, Belgium was used. This channel was 6 m long and had a 0.5 m high by 0.25 m wide rectangular cross section. The channel had glass windows to allow for optical access to the flow and a smooth bed made of Perspex. At mid-length, a gate divided the channel in two equal reaches, one to be used as upstream reservoir, the other as downstream test section for the experiments. The dam-break flow was simulated by the sudden opening of the mentioned gate. The gate that represents the dam was made of a 7 mm thick aluminium plate and it was pulled downwards by a pneumatic jack installed under the channel and remotely controlled by a switch. With this system, the maximum removal time of the gate was measured to be 120 ms, which complies with the Lauber and Hager (1998) criterion for the dam break to be considered as instantaneous. The channel bed was made of smooth Perspex plates with exception of a small region close to the gate that was made of aluminium. A detailed description of this channel can be found in Spinewine and Zech (2007). To minimize the leakage due to discontinuity in the channel to accommodate the downward moving gate, a small layer of plumber sealing was used to seal the area between the gate and the bed. No sediments were used during the experiments. Figure 4a shows a side view of the channel where the imaged area, located immediately after the gate, is depicted as a rectangle chosen within the trapezoid shape of the laser light sheet. Figure 4b shows a top view of the channel, where the location of the laser sheet is indicated.

The gate effect on the flow was studied by Aleixo, Soares-Frazão, Altinakar, and Zech (2013) and Aleixo, Soares-Frazão, and Zech (2016) who concluded that, for the tested condition

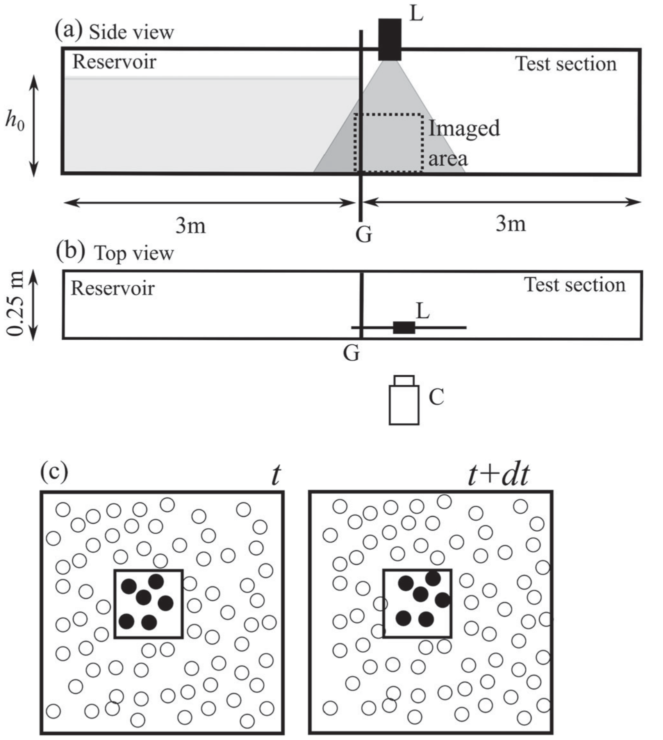

Figure 4 (a) Scheme of a side-view of the dam-break channel with the downward moving gate, G, and the laser light sheet issued from laser L. (b) Scheme of the top view of the channel and camera C. (c) PIV working principle based on the cross-correlation of interrogation windows in consecutive flow images

($h_0 = 0.325$ m) the measured removal time, $t_r \sim 120$ ms was sufficiently short to consider the gate removal as instantaneous.

To simulate the dam-break the upstream reservoir was filled with water up to a height of 0.325 m. It was observed that, after the complete opening of the gate (time $t = 0$ s), the reservoir would be completely empty in about 15 s. Seven runs of this experiment were performed and are used in the present analysis. It was observed, through the many experiments made in the dam-break channel, that the dam-break flow, if performed in the same conditions, is quite repeatable (Aleixo et al., 2011). This property allows isolation of one of the different runs made to test and illustrate the statistical methods described above in Sections 2.1 and 2.2.

3.2 *Measurement techniques*

PIV was used to measure flow data. To capture flow images a Photron Fast Camera (Photron, Tokyo, Japan) was used and placed perpendicular to the flow (Fig. 4b). This camera had a resolution of 1 megapixel for a maximum acquisition frequency of 2 kHz. For the experiments here described the acquisition frequency was set at 1 kHz, corresponding to an acquisition time interval of $\Delta t_a = 0.001$ s. Due to the limited memory buffer of the camera the total acquisition time was 1.361 s. This time, that can appear as very short, corresponds to a non-dimensional time $T = t/t_0 = 7.48$ that is much larger than the limit $T = 2.5$ identified by Aleixo et al. (2011) after which the flow can be considered as parallel. Also, the considered duration of the experiments covers both the first very transient stage of the dam-break flow and the slow varying stage where the channel progressively empties.

To isolate a flow section, a laser light diode with power $P = 40$ mW and emission on red was used. At the laser diode exit an existing diverging optics element generated a laser light sheet allowing lighting of a larger area of the flow (Fig. 4b). The laser sheet was placed at a distance of 0.025 m from the channel wall as, due to the limited laser power available, it was not possible to obtain clear images with the laser sheet placed on the channel centreline. Ideally the measurements should have been carried at the channel's centreline to minimize the side wall effect in the flow.

Despite the time evolution of the free surface, it was verified through different experiments that for the analysed time instants the laser light sheet was not significantly affected since the free surface was quite regular. On the other hand, it was verified that floating seeding particles would sometimes aggregate and cast a shadow in the flow. This was deemed not to be so important since it was usually of short duration.

To obtain quantitative information regarding the flow velocity field, the water in the reservoir was seeded with white pliolite particles. The pliolite particles used are roughly of spherical shape with a mean diameter of about 0.5 mm and had a relative density of 1.02, making them roughly neutrally buoyant.

The PIV technique relies on cross-correlation between consecutive images. For a detailed presentation of the method, the reader should refer to the extensive literature existing on the subject, namely the works of Raffel, Willert, Werely, and Kompenhans (2007), Tropea, Yarin, and Foss (2007), Adrian and Westwerweel (2010). The basic principles are recalled here. In PIV, two consecutive images taken at time t and $t + dt$ are divided in smaller areas, referred as interrogation windows (Fig. 4c). These interrogation windows are cross-correlated, resulting in a function whose peak is a measure of the most probable displacement of the ensemble of tracers inside the interrogation window. Since the time interval between two images is known, the velocity vector, $\boldsymbol{v}$, associated with each interrogation window, can be calculated as:

$$\mathbf{v} = \frac{\mathrm{d}\mathbf{s}}{\mathrm{d}t} \tag{15}$$

where d$\boldsymbol{s}$ is the infinitesimal displacement vector associated to the cross-correlation peak. This procedure is extended to all the interrogation windows, thus resulting in the whole velocity field.

The PIV acquisition and processing system used in the present work was the Davis® 7.2 by Lavision. The initial interrogation window was set to 256 × 256 pixels and reduced to 32 × 32 pixels. A window overlap of 75% was imposed allowing a spatial resolution of about 2 mm. PIV post-processing filtering was kept to a minimum to avoid blurring eventual features of the flow.

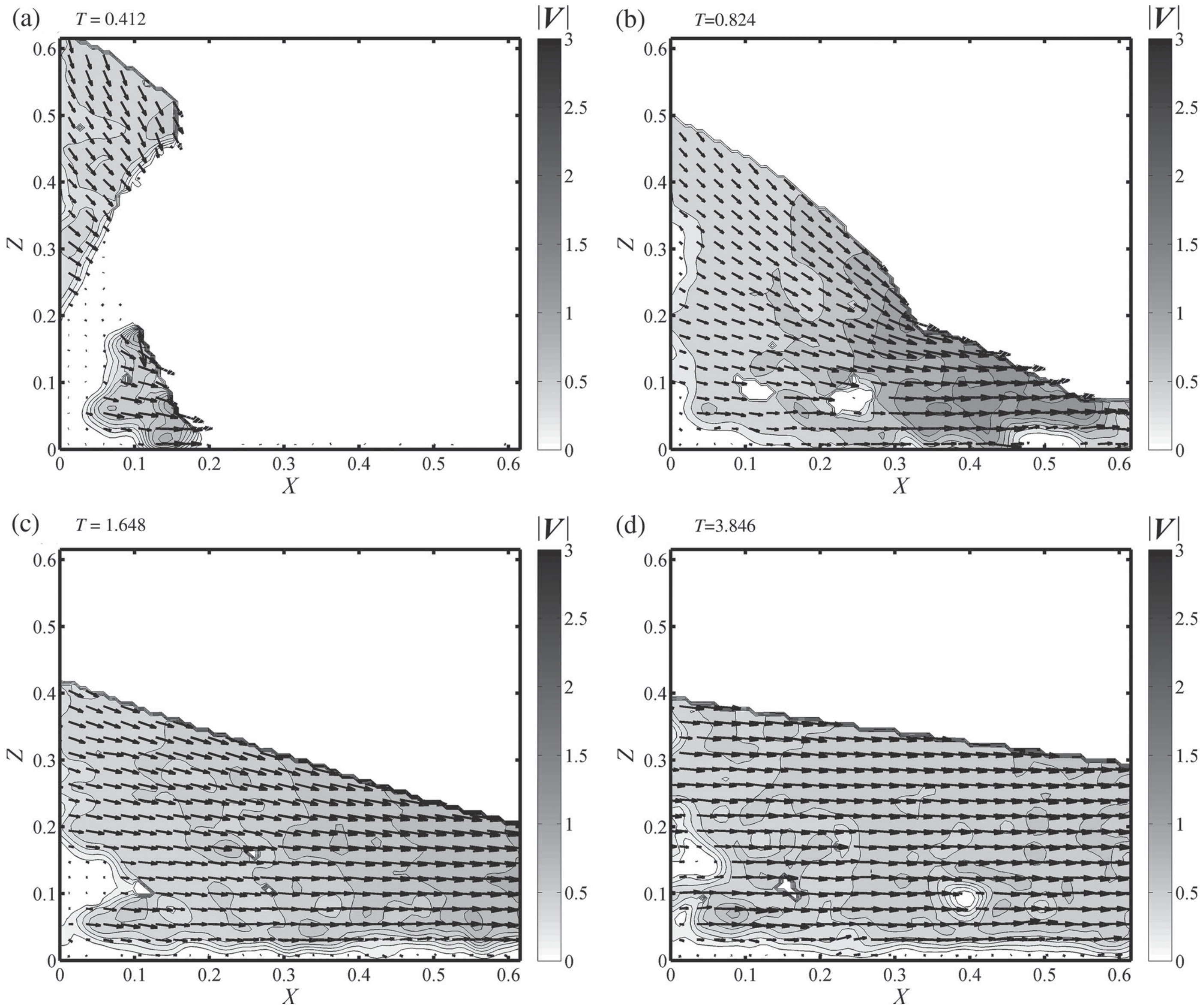

Figure 5 Dam-break flow run #4 at different instants (a) $T = 0.412$; (b) $T = 0.824$; (c) $T = 1.648$; (d) $T = 3.846$. (For visualization purposes, only one out of four vectors are depicted in each direction)

One of the challenges of the dam-break flow with respect to PIV analysis is the moving free surface. To handle the time-dependent free surface the PIV software should be robust to track which zones of the field of view are under the free surface. Alternatively, some manual treatment may be required to isolate the velocity data below the free surface. In the present case, and only for PIV time-series processing purposes, a rectangular region was considered inside the imaged area defined by the coordinates $0 < X < 0.6$ and $0 < Z < 0.25$, where $X = x/h_0$ and $Z = z/h_0$ are the non-dimensional spatial coordinates.

3.3 *Flow description*

Figure 5 shows the flow profile and the velocity vectors obtained after PIV analysis for four different instants of the flow for one of the runs at different times. In Fig. 5a an instant before the breaking of the wave is illustrated ($T = 0.412$). It is possible to see a clear vertical descending motion in the upper layer of the flow, whereas the lower layer already shows a more parallel flow. This movement is characteristic for the downward moving gate. In Fig. 5b ($T = 0.824$), the flow immediately after the collapse of the dam-break wave is shown. Close to the gate, a significant vertical component of the velocity is still present, but at the tip, the flow is parallel to the channel bed. For the later times of the flow (Fig. 5c and 5d) the vertical component of the flow is progressively reduced, and the flow becomes parallel to the channel bed. A detailed analysis of this evolution can be found in Aleixo et al. (2011).

4 Experimental results and data analysis

In this section the obtained results, the data processing and the comparison of the different methods for the analysis of the flow statistics are presented.

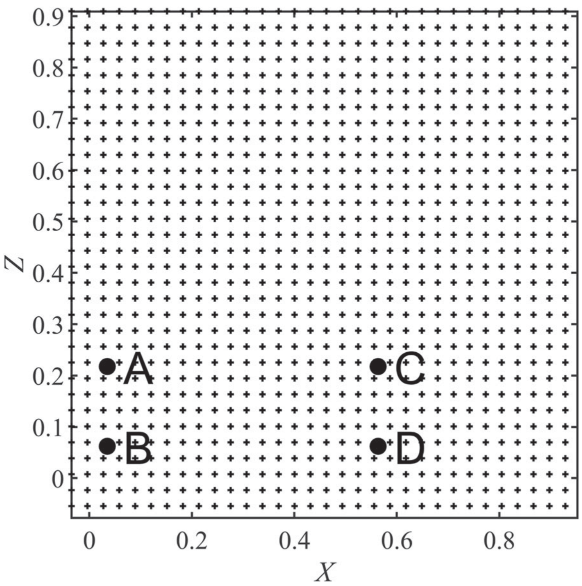

Figure 6 Measurement grid and four selected points for data series analysis

4.1 *Single realization*

In this case only run #4 of the seven available runs in the dataset is considered as example. Figure 6 shows the location of the four points where the time evolution of the velocity will be analysed. The non-dimensional coordinates of these points are: A = (0.0350, 0.2176), B = (0.0350, 0.0622), C = (0.5635, 0.2176) and D = (0.56335, 0.0622). The time-series of velocity at these four points are presented in Fig. 7. Contrasting with points A, and B, located very close to the gate, the transient character is particularly clear for points C and D. At point D the transient behaviour can be observed for the time period $0.75 \leq T \leq 2.75$, with the sharp transition between no-flow and flow occurring at $T = 0.75$. Afterwards, and for the duration of the measurement, the mean appears to become time independent.

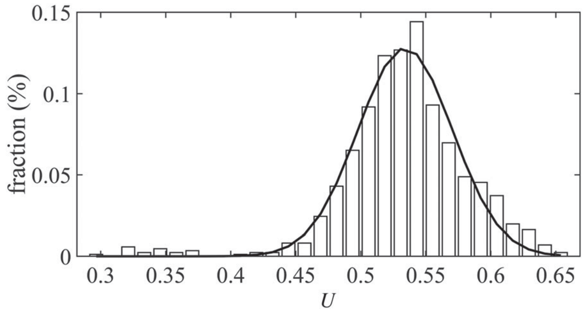

Figure 8 Histogram of the time-series of U at point D for the non-dimensional time interval $T = [2.75, 7.48]$

In the time interval ($0.75 \leq T \leq 2.75$) the application of the usual statistical tools will smoothen the transient character of the flow. This method will thus not be applied in this interval. However, for $2.75 < T \leq 7.48$, where the mean value does not seem to change with time, the methodology of Eqs (2)–(6) can be used. This is confirmed by the histogram of the horizontal velocity component that is depicted in Fig. 8: this histogram is symmetric around the mean value and can be fitted by a Gaussian curve.

For the case of point D, and considering $T \geq 2.75$, one has $N = 860$ data points. For this case, and considering $\phi = U$, the following values are obtained: $\mu_\phi = 0.954\,\text{m s}^{-1}$ and $\sigma_\phi = 0.0864\,\text{m s}^{-1}$. Replacing the known values in Eq. (7) yields for the horizontal component: $\mu_\phi/\mu_\infty = 0.979$, which means that the sample mean is only 2.1% different from the

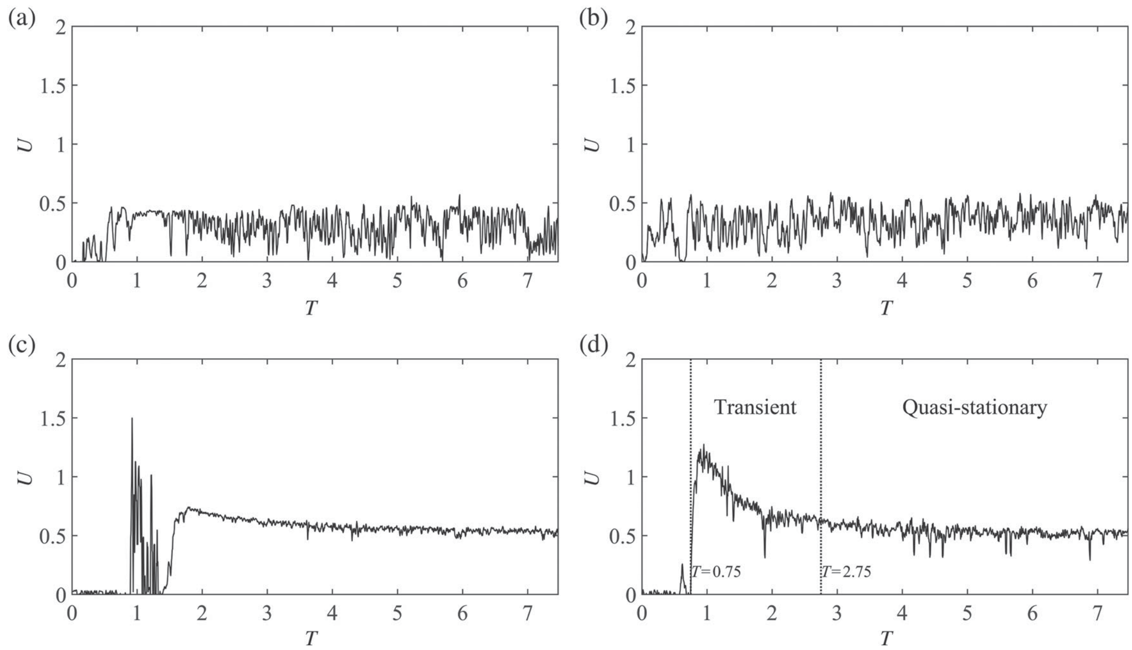

Figure 7 Dam-break run#4 time series for the different selected points (a) point A, (b) point B, (c) point C and (d) point D

Table 1 Mean values and respective standard deviations for each measured run at each point

	Point A		Point B		Point C		Point D	
	μ_ϕ/c_0	σ_ϕ/c_0	μ_ϕ/c_0	σ_ϕ/c_0	μ_ϕ/c_0	σ_ϕ/c_0	μ_ϕ/c_0	σ_ϕ/c_0
Run 1	0.440	0.0541	0.463	0.0454	0.561	0.0319	0.509	0.0468
Run 2	0.428	0.0616	0.446	0.0663	0.561	0.0311	0.538	0.0409
Run 3	0.403	0.0770	0.434	0.0806	0.557	0.0327	0.531	0.0445
Run 4	0.297	0.1178	0.366	0.1026	0.558	0.0325	0.535	0.0484
Run 5	0.331	0.1221	0.383	0.0944	0.560	0.0322	0.523	0.0432
Run 6	0.325	0.1090	0.387	0.0823	0.554	0.0316	0.524	0.0436
Run 7	0.337	0.1179	0.397	0.0933	0.561	0.0304	0.525	0.0588

mean of the population. By means of Eq. (8), the statistical error associated to the mean is 0.44%.

Table 1 shows the results of this statistical analysis for the four considered points and the seven runs. It can be observed that for each measurement point, the statistical values between the different runs are quite close to each other, especially for points C and D. Also, as expected from Fig. 7, the standard deviation is larger for points A and B than for points C and D, located further downstream.

4.2 *Fourier-component method*

As it was shown in the previous section, the strong transient character of the flow in the first instants, $0.75 < T < 2.75$, prevented the use of the common statistical tools as defined by Eqs (3)–(6). To have an estimation of a local mean value, the Fourier method described in Eqs (9)–(13) is here applied to the time interval [0.75; 7.48].

The value m is related with the threshold frequency as indicated in Eq. (9). Given that the PIV images were acquired at a frequency of 1 kHz, according to the Shannon–Nyquist criterion, the maximum frequency that can be described is 500 Hz. Replacing $f_c = 500$ Hz, one obtains $m = 182$. The following procedure was made to assess the influence of m in the results. For each run i measured at a given point, say point C, different values of $m = 7, 50, 100, 150, 250$ were tested. Then the standard-deviation of the obtained Fourier time-series in the quasi-stationary time interval was compared with the standard-deviation calculated as in Section 4.1 for the same time interval, by means of the following expression:

$$\bar{\varepsilon}_C(m) = \frac{\text{mean}(|\sigma_{\text{Fourier}} - \sigma|_{i=1,2,\ldots,7})}{\mu_{u(C)}} \tag{16}$$

where $\bar{\varepsilon}_C$ is the relative mean standard deviation error, σ_{Fourier} is the standard-deviation of the resulting Fourier-component method time-series in the quasi-stationary time interval, σ is the standard-deviation of the experimental realizations in the same time interval and $\mu_{u(C)}$ is the mean velocity at point C for that same time interval. The obtained results are depicted in Fig. 9 for each point A, B, C and D of Fig. 6. It is possible to see that after $m = 150$, the value of $\bar{\varepsilon}_C$ does not change significantly.

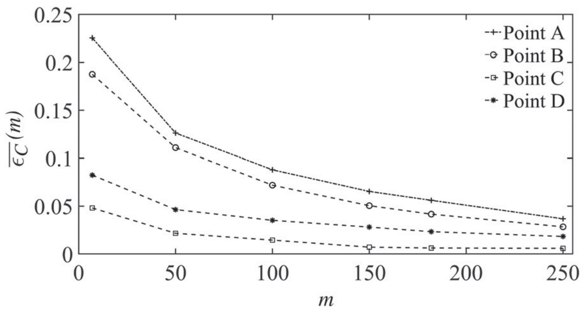

Figure 9 Determination of the relative mean standard deviation error as a function of the m parameter

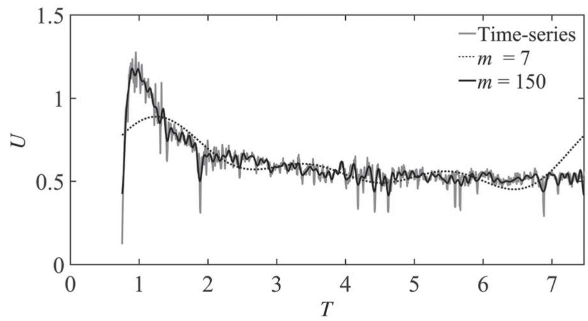

Figure 10 Application of the Fourier method to the interval $T = [0.75, 7.48]$ and evaluation of the Fourier component method for two values of m parameter at point D

For this dam-break condition $m = 150$ was then chosen corresponding to a cut-off frequency of $f_c = 409$ Hz. This value has the advantage of taking a conservative approach to the Nyquist–Shannon criterion. In Fig. 10, the measured time-series for run #4 and the obtained Fourier mean time-series with $m = 7$ and $m = 150$ are depicted. It is possible to see that the value used by Nezu et al. (1997) is not enough for the dam-break flow. A side consequence of this method is to provide a means to smooth the data as it can be seen in Fig. 10.

4.3 *Ensemble average*

To apply the ensemble average procedure, seven runs of the dam-break flow were analysed. Although the number of runs is limited, they will be used here to illustrate the application of Eq. (14). In Fig. 11 the ensemble average velocity fields are depicted. One of the advantages of the ensemble average

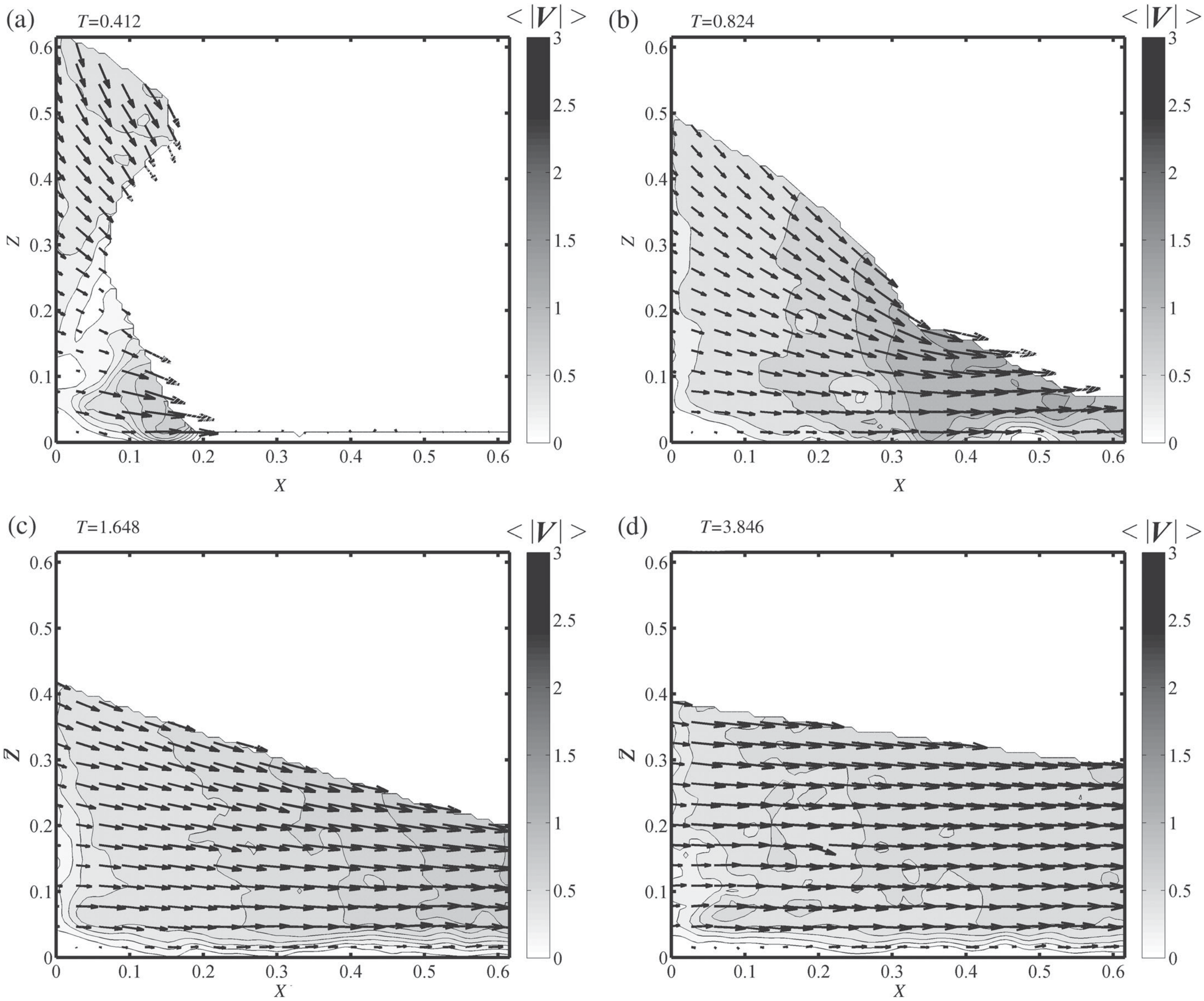

Figure 11 Obtained ensemble average of the dam-break flow at different instants (a) $T = 0.412$; (b) $T = 0.824$; (c) $T = 1.648$; (d) $T = 3.846$. (For visualization purposes, only one out of four vectors are depicted in each direction)

is to provide a natural way of filtering noise. Since each run is independent, the noise will not be correlated, and by means of ensemble averaging it is possible to smooth the noise in the measured variables. Another advantage is to allow completing the dataset. As it can be seen by comparing Fig. 5a and Fig. 11a, the repetition of experiments allowed more data to be obtained in the region $X \leq 0.1$ and $Z \leq 0.3$. In this process, the ensemble-averaged values are computed using only valid data; missing data are not treated as zero and do not contribute to the average value calculation.

To investigate if seven runs are enough to obtain a better data ensemble, different ensembles will be considered by taking ensemble averages of 1, 2, 3, 4, 5 and 6 runs, resulting in 7, 21, 35, 35, 21 and 7 possible combinations, respectively. The different sets obtained considering ensembles of 1, 2, 3, 4, 5 and 6 runs are here referred as ensemble sets. Each one of these combinations is compared with the ensemble average of the seven runs by means of their relative root mean square error, defined as:

$$\mathrm{RMSE}_{ik}^2 = \frac{1}{N}\sum\left(\frac{< u_{ik} > - < u >}{< u >}\right)^2 \tag{17}$$

where $< u_{ik} >$ is the ensemble average of the i-combination of k-runs, ($k = 1, 2, 3, 4, 5, 6$) out of a 7-element set, $< u >$ is the ensemble average with all the seven runs, and N is the number of points in the time series. For each combination Eq. (17) is calculated and then averaged by ensemble k. The results for $k = 6$ are exemplified, for Point D, in Fig. 12 and Table 2. For $k = 6$ one has seven possible combinations ($i = 1, \ldots, 7$) listed in Table 2. Figure 12a depicts the different ensembles (light grey dashed lines), the ensemble average corresponding to the combination $i = 4$ in (dark grey dash line) and the ensemble average with all the runs (continuous black line). In Fig. 12b Eq. (17) is

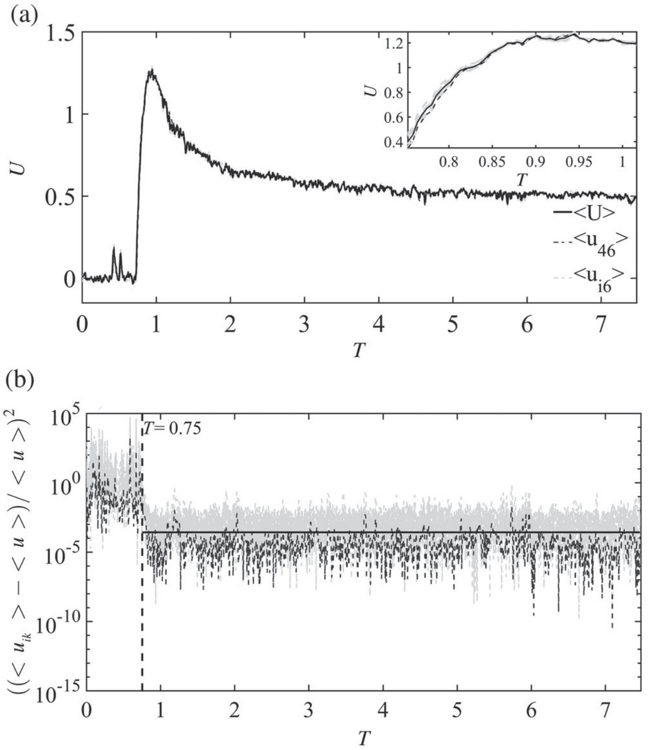

Figure 12 Determination of RMSE for the combination of ensembles using $k = 6$ runs. (a) Ensemble values. (b) Squared difference between each ensemble of six runs and the ensemble with all the runs

Table 2 Example of the RMSE obtained for each combination of six runs out of a set with seven runs

$k = 6$	Point D		
i	Run combination	$\mathrm{RMSE}_{i6}{}^2$	RMSE_{i6}
1	1-2-3-4-5-6	1.779×10^{-4}	0.0133
2	1-2-3-4-5-7	1.323×10^{-4}	0.0115
3	1-2-3-4-6-7	1.735×10^{-4}	0.0132
4	1-2-3-5-6-7	2.685×10^{-4}	0.0164
5	1-2-4-5-6-7	1.514×10^{-4}	0.0123
6	1-3-4-5-6-7	1.195×10^{-4}	0.0109
7	2-3-4-5-6-7	1.664×10^{-4}	0.0129
		Mean	0.0129

calculated for each combination (dashed grey lines) and the case corresponding to $i = 4$ is highlighted (dark grey dashed line), its mean value is then computed in the interest region (right side of the vertical dashed line). This case corresponds to the line $i = 4$ in Table 2. The final results of the mean RMSE for points A, B, C and D are depicted in Fig. 13. It is possible to see that the RMSE decreases as the number of runs in the ensemble increases. Although it is not possible to directly conclude that seven runs are enough, it is possible to see that by increasing the number of runs the RMSE is reduced.

In Figs 14–16 the mean velocity profile, the respective standard deviation, and the velocity cross-correlation ψ_{UW} at vertical AB and CD for each run calculated in the time interval $T = [2.75, 7.48]$, are illustrated and compared with the

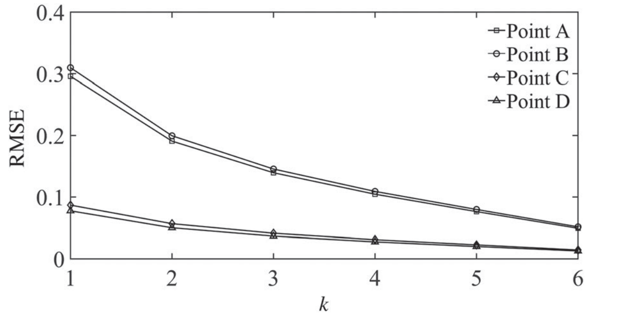

Figure 13 Variation of the RMSE with the number of runs in each ensemble

ensemble average. It is possible to see how the ensemble averaged velocity profile of U evolves between $X = 0.0350$ and $X = 0.56335$. At this latter location, the profile resembles a boundary-layer velocity profile. The standard deviation was also found higher closer to the gate than at $X = 0.56335$. The cross-correlation ψ_{UW} showed positive values at $X = 0.0350$ and $0.05 < Z < 0.12$, whereas for $X = 0.56335$ only in a small region $Z < 0.04$, the cross-correlation was found to be positive. Figures 14–16 show that the variability between runs was higher in a limited region closer to the gate than at the downstream sections. This is most probably related with the gate region conditioning. Figures 14a, 15a and 16a show one of the interesting features of the ensemble average, that is to obtain an average value in a region where there is high variability.

4.4 *Comparison between the Fourier method and the ensemble average*

To evaluate if the Fourier component method applied to a single run can be compared with the ensemble average method,

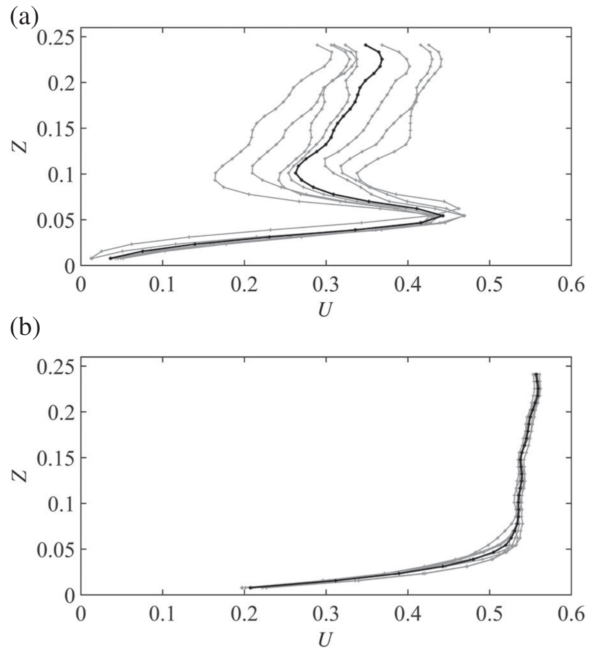

Figure 14 Ensemble average velocity profiles measured at: (a) $X = 0.0350$ and (b) $X = 0.5635$

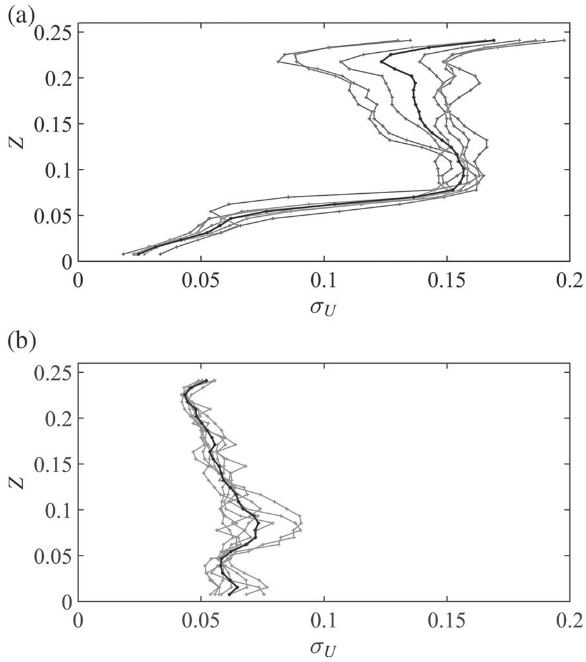

Figure 15 Ensemble average of the standard deviation of the velocity profiles measured at: (a) $X = 0.0350$ and (b) $X = 0.5635$

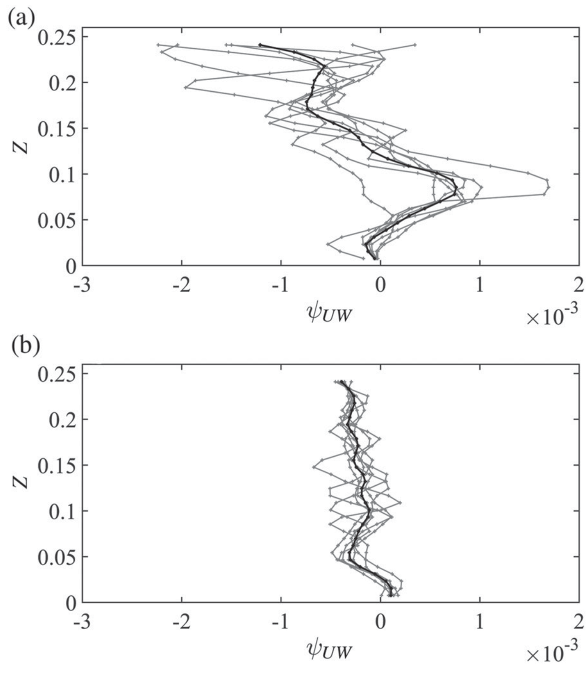

Figure 16 Ensemble average of the velocity correlation of the velocity profiles measured at: (a) $X = 0.0350$ and (b) $X = 0.5635$

a comparison between the results obtained with the Fourier method and the ensemble average results is depicted in Fig. 17. As it can be seen, practically every single run fits within one ensemble-standard deviation of the ensemble-mean value. To measure the discrepancy between the methods the mean root mean squared relative error is calculated using the following

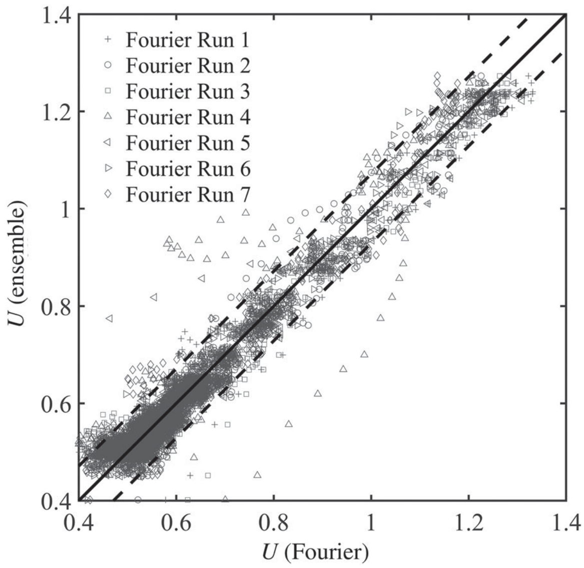

Figure 17 Comparing the non-dimensional velocity obtained with the Fourier component method with the ensemble average with all the runs at point D. Dashed lines: U(ensemble) $\pm\,\sigma$

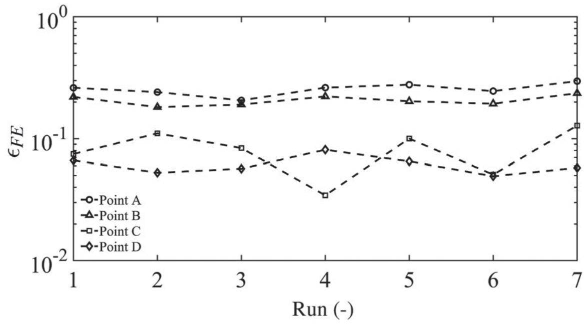

Figure 18 Root mean square error between Fourier component method and ensemble average for each run (1, 2, ... , 7) in points A, B, C and D

expression:

$$\varepsilon_{FEi}^2 = \frac{1}{N}\sum\left(\frac{u_{Fi} - u_E}{u_E}\right)^2 \tag{18}$$

where ϵ_{FE} is the root mean square error, u_{Fi} and u_E are the obtained time-series using the Fourier method with $m = 150$ for each run i, and the ensemble mean respectively, and N is the number of points in the time-series. Figure 18 depicts ϵ_{FE} for each run at each point A, B, C and D. It is possible to see that for points C and D, $0.03 < \epsilon_{FE} < 0.13$ whereas for points A and B $0.18 < \epsilon_{FE} < 0.30$.

4.5 *Ensemble averages of time-averaged profiles*

Finally, to complete this study, the ensemble average is applied for the time-averaged data for $T > 2.75$, at different sections downstream of the gate. The obtained time-ensemble averaged profiles are depicted in Fig. 19. This method combines the time average procedure discussed in Section 4.2 with the ensemble statistical method described in Section 4.4. This allows a

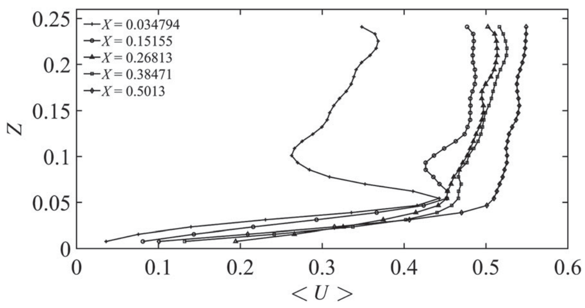

Figure 19 Ensemble average of the velocity profiles measured at different locations

smoother dataset to be obtained. Furthermore, it is possible to see that, in a relative short length ($dX \sim 0.12$) the initial distorted velocity profile converges to the typical boundary layer velocity profile. The difference between velocity profiles measured at $X = 0.034974$ and $X = 0.15155$ respectively emphasizes the gate section conditioning effect on the flow.

5 Summary and conclusions

Three different statistical methods were applied to a case of dam-break flow. The first method consisted in the use of the classical statistical tools, the second was the Fourier component method, and third was the ensemble average method.

The first method application is limited to the instants where the flow can be considered quasi-stationary. This feature constraints its application to the quasi-stationary part of the flow. Some criteria for the convergence of mean values as function of the sample size were also found in the literature and used, as this is important for future experiment planning.

The Fourier component method proved to be applicable to the whole transient time-series, making it more robust. However, this method requires the determination of a cut-off frequency, which is a function of the number of terms considered for the Fourier series. To determine the cut-off frequency, or equivalently, the number of terms in the Fourier series, two different approaches were applied that lead to similar results: one based on the maximum frequency possible to measure with the Shannon–Nyquist criterion, the other based on the standard-deviation of the quasi-stationary part of the time-series. The comparison with the ensemble average value, for a point away from the gate, has shown that in this case the Fourier component method gives results quite close to the ensemble average method. Furthermore, it was observed that the Fourier component method could be used as a way to smooth the data.

Finally, the ensemble average method can be applied to all the transient time-series and was applied to the available runs. The first consequence of repeating the same experiment in the same conditions was the possibility to complete the dataset. Since noise is not correlated between independent runs, the ensemble average method can also be used as a method to smooth the data. It was also seen that, for time-averaged profiles in regions with significant variability, this method provides a way to define a mean value. Although it was not demonstrated that seven runs are sufficient in terms of ensemble averages convergence, it was shown that by repeating the experiment it is possible reduce the root mean square error. From an experimental point of view, this method is the most expensive one since it requires the same experiment to be repeated several times.

From what was exposed it is possible to devise an approach to study the dam-break flows and transient flows in general using statistical methods:

(1) If only a single run of the experiment can be carried out, the Fourier component method should be used for the analysis of the measured time-series. If there are regions where the flow can be considered quasi-stationary, then common statistical tools can be used provided that the number of points is sufficient to ensure the convergence of the statistical moments.
(2) If it is possible to repeat the same experiment several times, the ensemble average method should be applied and supplemented by the application of the Fourier method for each individual run.

By means of the application of the statistical methods it was also possible to observe some features of the dam-break flow in the region close to the gate. One of these features is the possibility of analysing the flow using the boundary-layer theory as seen by the measured velocity profiles. Another future use of the statistical methods is to analyse the turbulence in a dam-break flow, thus leading to a better comprehension of such flows. These measurements should be made preferably along the centreline of the channel to avoid the side wall effect on the flow.

Acknowledgements

The first author thanks Dr Magdalena Mrowkowska for the provided bibliographic references and helpful discussion, and Dr Carlo Someda and Bruno Griffoni, from GHT Photonics, for the computational and physical resources made available.

Funding

The dam-break experiments reported in this paper were made possible thanks to the scholarship SFRH/BD/36023/2007 from the Fundação para a Ciência e Tecnologia, Portugal given to the first author.

Notation

a_0 = Fourier coefficient
a_k = Fourier co-sinus coefficient

b_k = Fourier sinus coefficient
c_0 = dam-break wave celerity (m s^{-1})
$d\boldsymbol{s}$ = infinitesimal displacement vector (m)
dt = time interval between consecutive images (s)
f_c = cut-off frequency (Hz)
g = gravity acceleration (m^{-2})
h_0 = initial water height in the reservoir (m)
i = summation index (–)
k = summation index, number of runs in each combination (–)
m = cut-off parameter (–)
N = number of samples (–)
T = non-dimensional time (–)
t_0 = dam-break natural time scale (s)
t_i = beginning instant of the sample process (s)
t_1 = generic time instant (s)
t_2 = generic time instant (s)
t_3 = generic time instant (s)
t_4 = generic time instant (s)
T_h = phenomenon time scale (s)
T_s = sampling period (s)
U = non-dimensional horizontal velocity component (–)
$< u >$ = ensemble average non-dimensional velocity time-series obtained from the ensemble average with all the runs at a given point (–)
u_E = non-dimensional velocity time-series obtained from the ensemble average with all the runs at a given point (–)
u_{Fi} = non-dimensional velocity time-series obtained from the Fourier component method for run i at a given point (–)
u_{ik} = non-dimensional velocity time-series obtained from the Fourier component method for run i at a given point (–)
$< u_{ik} >$ = ensemble average of the i-combination of k-runs
$\boldsymbol{v}$ = velocity vector (m s^{-1})
$|\boldsymbol{V}|$ = non-dimensional velocity vector modulus (–)
$< |\boldsymbol{V}| >$ = ensemble average of the non-dimensional velocity vector modulus (–)
W = non-dimensional vertical velocity component (–)
X = non-dimensional horizontal coordinate (–)
Z = non-dimensional vertical coordinate (–)
β = constant (–)
Δt_{43} = time interval between two instants (s)
Δt_a = sampling time interval (s)
ε_ϕ = statistical error of variable ϕ
$\bar{\varepsilon}_C$ = relative mean standard deviation error at point C (–)
ε_{FE} = relative root mean square error between the Fourier component of run i and the ensemble average (–)
θ = generic variable
$\mu_{u(C)}$ = mean velocity measured at point C (m s^{-1})
σ_ϕ^2 = variance of a sample of variable ϕ
σ_ϕ = standard deviation of a sample of variable ϕ
ϕ = generic variable
$\psi_{\phi\varphi}$ = cross-correlation of variables ϕ and φ
ψ_{UW} = cross-correlation of variables U and W
μ_ϕ = mean value of a sample of variable ϕ

ORCID

Sandra Soares-Frazão http://orcid.org/0000-0001-5088-0278

References

Adrian, R., & Westwerweel, J. (2010). *Particle image velocimetry*, Cambridge: Cambridge University Press.
Aleixo, R. (2011). Time resolved PIV analysis of the initial stages of the dam break. In E. M. Valentine, C. J. Apelt, J. Ball, H. Chanson, R. Cox, R. Ettema, G. Kuczera, M. Lambert, B. W. Melville, & J. E. Sargison (Eds.), *Proceedings of the 34th IAHR conference* (pp. 3313–3320). Brisbane: International Assn for Hydro-Environment Engineering and Research (IAHR).
Aleixo, R., Soares-Frazão, S., Altinakar, M., & Zech, Y. (2013). The gate removal effect in the initial instants of the dam-break flow. *Proceedings of the 35th IAHR conference*. Chengdu: International Assn for Hydro-Environment Engineering and Research (IAHR).
Aleixo, R., Soares-Frazão, S., & Zech, Y. (2011). Velocity-field measurements in a dam-break flow using a PTV Voronoï imaging technique. *Experiments in Fluids*, *50*, 1633–1649.
Aleixo, R., Soares-Frazão, S., & Zech, Y. (2016). Before the dam breaks: Analysis of the flow behind a downward moving gate. In G.Constantinescu, M.Garcia, & D.Hanes (Eds.), *Proceedings of the Riverflow2016 conference*. St. Louis, MI: CRC Press.
Aleixo, R., Spinewine, B., Soares-Frazão, S., & Zech, Y. (2010). Velocity profiles in dam-break flows: Water and sediment layers. *Proceedings of the Riverflow2010 conference* (pp. 533–540). Braunschweig: Bundesanstalt für Wasserbau.
Aleixo, R., Zech, Y., & Soares-Frazão, S. (2012). Velocity profiles in dam-break flows: Water and sediment layers. In R. M. Munoz (Ed.), *Proceedings of the Riverflow2012 conference* (pp. 311–318). San Jose: CRC Press.
Bates, C., & Hughes, T. (1977). The effect of both sample size and sampling rate on the statistical fluid flow parameters in a high Reynolds number, low turbulence intensity flow. *5th biennial symposium on turbulence* (pp. 125–131). University of Missouri-Rolla, USA.

Bates, C. J., & Hughes, T. D. (1976). Real-time statistical LDV system for the study of a high Reynolds number, low turbulence intensity flow. *Journal of Physics E: Scientific Instruments*, *9*, 955–958

Bell, S. W., Elliot, R. C., & Hanif Chaudry, M. (1992). Experimental results of two-dimensional dam-break flows. *Journal of Hydraulic Research*, *30*(2), 225–252.

Bendat, J. S., & Piersol, A. G. (2011). *Random data: Analysis and measurement procedures*. Wiley Series in Probability and Statistics (4th ed.). Hoboken, NJ: Adison-Wesley.

Capart, H., & Young, D. L. (1998) Formation of a jump by the dam-break wave over a granular bed. *Journal of Fluid Mechanics*, *372*, 165–187.

Dey, S., & Lambert, M. F. (2005). Reynolds stress and bed shear in nonuniform unsteady open-channel flow. *Journal of Hydraulic Engineering*, *131*(7), 610–614.

Dressler, R. F. (1952). Hydraulic resistance effect upon the dam-break functions. *Journal of Research of the National Bureau of Standards*, *49*(3), 217–225.

Durst, F., Kikura, H., Lekakis, I., Jovanović, J., & Ye, Q. (1996). Wall shear stress determination from near-wall mean velocity data in turbulent pipe and channel flows. *Experiments in Fluids*, *20*, 417–428.

Fraccarollo, L., & Toro, E. F. (1995). Experimental and numerical assessment of the shallow water model for two-dimensional dam-break type problems. *Journal of Hydraulic Research*, *33*(6), 843–864.

Hitching, E., & Lewis, A. W. (1999). Sampling techniques for accurate measurement of Reynolds stresses using laser Doppler anemometry. *Flow Measurement and Instrumentation*, *10*, 241–247.

ICOLD, (1995). *Dam failures statistical analysis*, Bulletin 99. Paris: International Commission on Large Dams (ICOLD).

Jensen, B. L., Sumer, B. M., & Fredsøe, J. (1989). Turbulent oscillatory boundary layers at high Reynolds numbers. *Journal of Fluid Mechanics*, *206*, 265–297.

Lauber, G., & Hager, W. H. (1998). Experiments to dambreak wave: Horizontal channel. *Journal of Hydraulic Research 36*, 291–307.

Leal, J. G. A. B., Ferreira, R. M. L., & Cardoso, A. H. (2006). Dambreak wave-front celerity. *Journal Hydraulic Engineering*, *132*(1), 995–999.

Leal, J. G. A. B., Ferreira, R. M. L., & Cardoso, A. H. (2009). Maximum level and time to peak of dam-break waves on mobile horizontal bed. *Journal Hydraulic Engineering*, *132*(1), 69–76.

Lumley, J. (1979). *Stochastic tools in turbulence*. New York, NY: Academic Press.

Nezu, I., Kadota, A., & Nakagawa, H. (1997). Turbulent structure in unsteady depth-varying open-channel flows. *Journal of Hydraulic Engineering*, *123*(9), 752–763.

Oertel, M., & Bung, D. B. (2012). Initial stage of two-dimensional dam-break waves: Laboratory versus VOF. *Journal of Hydraulic Research*, *50*(1), 89–97.

Ozmen-Cagatay, H., & Kocaman, S., (2012). Investigation of dam-break flow over abruptly contracting channel with trapezoidal-shaped lateral obstacles. *ASME Journal of Fluids Engineering*, *134* (081204), 1–7.

Ozmen-Cagatay, H., Kocaman, S., & Guzel, H., (2014). Investigation of dam-break flood waves in a dry channel with a hump. *Journal of Hydroenvironmental Research*, *8*, 304–315.

Raffel, M., Willert, C., Werely, S., & Kompenhans, J. (2007). *Particle image velocimetry – a practical guide* (2nd ed.). Heidelberg: Springer.

Reichstetter, M., & Chanson, H. (2013). Unsteady turbulent properties in negative waves in open channels. *European Journal of Mechanics – B/Fluids*, *37*, 1–9.

Ritter, A. (1892). Die Fortpflanzung der Wasserwellen [The propagation of water waves]. *Zeitschrift Verein Deutscher Ingenieure*, *36*(2, 33), 947–954.

Song, T. (1994). *Velocity and turbulence distribution in non-uniform and unsteady open channel flow* (*Doctoral dissertation*). Ecole Polytechnique Federale de Lausanne, Lausanne.

Song, T., & Graf, W. H. (1996). Velocity and turbulence distribution in unsteady open-channel flows. *Journal of Hydraulic Engineering*, *122*(3), 141–154.

Spinewine, B., & Zech, Y. (2007) Small-scale laboratory dam-break waves on movable beds. *Journal of Hydraulic Research*, *45*(Extra Issue), 73–86.

Tardu, S. F., Binder, G., & Blackwelder, R. F. (1994). Turbulent channel flow with large-amplitude velocity oscillations. *Journal of Fluid Mechanics*, *267*, 109–151.

Tropea, C., Yarin, A., & Foss, J. F. (Eds.). (2007). *Springer handbook of experimental fluid mechanics*. Berlin: Springer-Verlag.

Yanta, W., & Smith, R. (1973). Measurement of turbulence transport properties with a laser Doppler velocimeter. *Proceedings of the 11th Aerospace science meeting*, AIAA. Washington DC, USA.