Assessment of the statistical relevance of TR-PIV datasets

Ludovic Chatellier[1], Young Jin Jeon[1], Patrick Braud[1], Victor Parent[1] and Laurent David[1]
[1] Institut PPRIME, UPR3346, CNRS – Université de Poitiers – ISAE-ENSMA, FRANCE
ludovic.chatellier@cnrs.pprime.fr

ABSTRACT

The increasing interest for high repetition rate global optical measurement techniques such as Time-Resolved Particle Image Velocimetry (TR-PIV) raises a number of questions concerning their ability to provide relevant statistical and spectral quantities. In an effort to address this issue, complementary TR-PIV and Laser Doppler Velocimetry (LDV) measurements have been carried out. An application to the analysis of the detached flow over a NACA 0015 airfoil at a Reynolds number Re=10^5 in a water tunnel is proposed.

INTRODUCTION

The rapid emergence of PIV as a reliable solution for global diagnostics of fluid flows has raised this technique to a high level of accuracy combined with an always increasing resolution and growing repetition rates. For any PIV experiment, the eduction of converged statistical quantities is possible as long as an appropriate, i.e. sufficient, number of mutually independent velocity fields is obtained, as illustrated by the practical analysis proposed in [1]. So-called classical, dual frame, PIV setups capable of maximum repetition rates of the order of 10Hz are generally used for that purpose. In most applications, this type of apparatus cannot natively provide time-resolved data, for which TR-PIV solutions are then needed. However, time-resolved systems comprise high repetition rate lasers which cannot always be efficiently used at both high and low frequencies, thus prohibiting the use of a single TR-PIV system for acquiring both time-resolved and non time-resolved images series. Moreover, and since TR-PIV cameras depend on limited on-board memory, the statistic and temporal analyses are limited to several thousands of samples, which may limit their ability to fully describe unsteady and turbulent flows.
Additionally, considerable effort is put on the development of specific TR-PIV algorithms, which rely on the analysis of more than two consecutive images for enhancing the accuracy of the computed velocity fields [2,3,4]. Hence, TR-PIV experiments shall be designed to ensure that both instantaneous and averaged quantities are obtained with the best possible accuracy.

1-EXPERIMENTAL SETUP

In order to assess how TR-PIV experiments may best combine accuracy, statistical relevance and high temporal resolution, independent series of multi-block TR-PIV experiments have been conducted and completed by separate LDV measurements. An application to the fully detached flow around a NACA 0015 airfoil of chord c=80mm at an angle of attack α=30° and a Reynold number Re=10^5 is described (Figs.1 and 2), for which an analysis of the statistical and spectral relevance of the experimental datasets is proposed.

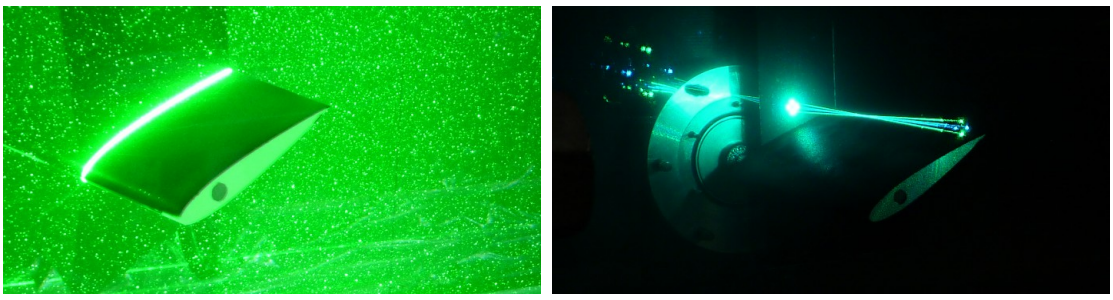

Figure 1 Experimental 2D2C TR-PIV (left) and 2C LDV (right) setups.

The TR-PIV series were acquired using two 1024^2 pixels Photron FASTCAM SA-1.1 cameras operated in tandem at a rate f_{PIV}=1500fps. The Region Of Interest (ROI) covered by the cameras is 4,65 chords long and 2,34 chords high, with a 2.6mm wide overlapping strip in the measurement plane (see Figure 2) . VESTOSINT polyamide particles were used for seeding and a double sided illumination was provided by two Nd:YAG lasers (Quantronix Darwin Duo, New-Wave Research Pegasus).
The 32GB cameras' on-board memory allow recordings of 21841 images. The baseline series of measurements comprises a single block of 21841 images. A first set of multi-block image series containing 78 blocks of 280 images

was recorded following four duty cycles: 50%, 25%, 10% and 1%. A second set of multi-block image series containing 910 blocks of 24 images was recorded following three duty cycles: 8.5%, 3% and 1%. The chosen duty cycles correspond to equivalent time lags between TR-PIV blocks and are respectively based on 1, 3, 9 and 99 (78×281 set only) vortex shedding periods.

A multi-threaded TR-PIV Multi-Frame Pyramid Correlation Algorithm [3] coded in C++ [4, 5] and using blocks of 6 successive images was used to evaluate the velocity fields.

The LDV measurements were carried out within the TR-PIV ROI over a 24×11 cartesian grid of step 11.7 mm, corresponding to 64 pixels in the TR-PIV images (see Figure 2), so that the velocity components can be evaluated coincidentally in both systems. A two-components Dantec (FiberFlow ; BSA-F80 ; BSA Flow-Software) system equipped with Spectra-Physics Beamlok 2060 lasers and operating in coincident mode was used, providing 5mn recordings with mean datarates ranging from 1.6kHz to 6kHz.

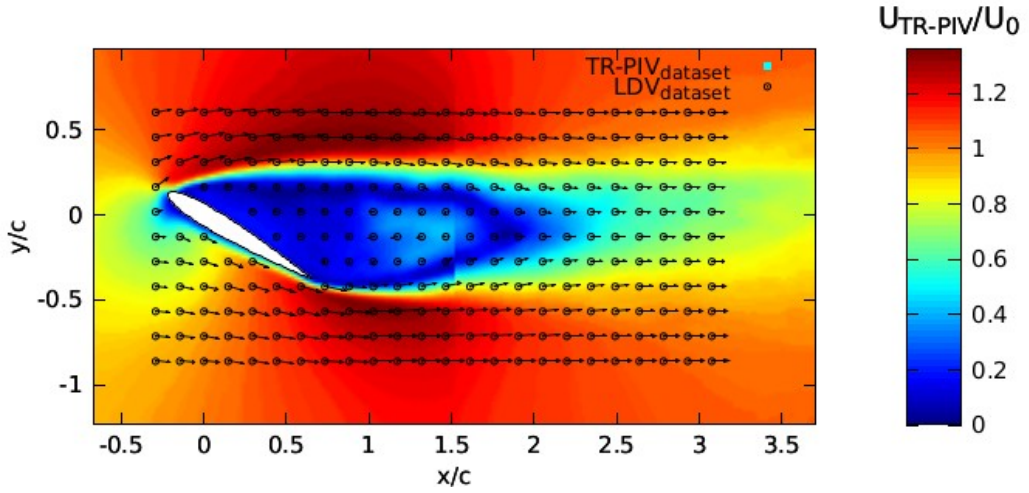

Figure 2 Distribution of averaged velocities on superimposed TR-PIV and LDV measurement domains.

2-THEORY

2.1 Expected value and variance of sample estimates

The influence of the data acquisition procedure (sample size, data rate, ...) on the flow diagnostics via the intricacy of statistical (mean and rms values, probability density functions) and temporal (auto-correlation functions) quantities can be derived from classical signal processing concepts.

Given N observations of a random variable x, any quantity of interest s characterizing x may be evaluated using the available sampled values $(x_i)_{i=0...N-1}$ of x using a chosen estimator $\hat{s}$. The quality of the estimator may be estimated using its expected value $E[\hat{s}]$ and variance $V[\hat{s}]$ from which its bias $B[\hat{s}]$ and mean squared error $MSE[\hat{s}]$ can be derived:

$$MSE[\hat{s}] = E\left[(\hat{s}-s)^2\right] = V[\hat{s}] + B[\hat{s}]^2 \tag{1}$$

with

$$B[\hat{s}] = E[\hat{s}] - s \quad \text{and} \quad V[\hat{s}] = E\left[(\hat{s}-E[\hat{s}])^2\right] = E[\hat{s}^2] - E[\hat{s}]^2 . \tag{2}$$

The quantity s may be classically evaluated from a single series $(x_i)_{i=0...N-1}$ of N sampled values of x or from multiple series, using separated blocks of data.
We shall also consider here that $(x_{Lk+i})_{k=0...M-1;i=0...Bk-1}$ represents M blocks of length B_k separated by intervals L_k. For comparison purposes, N may represent the maximum number of available samples of x, so that $\Sigma B_k \leq N$, or, for blocks of unique size B, $M \times B \leq N$. The case of a single series of data corresponds to M=1; B_0=B=N.

The estimator $\hat{s}$ is then defined from M estimators $\hat{s}_k$ arranged with convenient arithmetics. For most statistical quantities, $\hat{s}$ will be defined as the arithmetic mean of $(\hat{s}_k)_{k=0...M-1}$:

$$\hat{s} = \frac{1}{M}\sum_{k=0}^{M-1} \hat{s}_k \ , \tag{3}$$

yielding:

$$E[\hat{s}]=\frac{1}{M}\sum_{k=0}^{M-1}E[\hat{s}_k] \text{ and } V[\hat{s}]=\frac{1}{M^2}\sum_{k=0}^{M-1}V[\hat{s}_k]+\frac{1}{M^2}\sum_{k=0}^{M-1}\sum_{\substack{l=0\\l\neq k}}^{M-1}E\left[(\hat{s}_k-E[\hat{s}])(\hat{s}_l-E[\hat{s}])\right]. \quad (4)$$

Similarly,

$$B[\hat{s}]=\frac{1}{M}\sum_{k=0}^{M-1}B[\hat{s}_k] \text{ and } MSE[\hat{s}]=\frac{1}{M^2}\sum_{k=0}^{M-1}MSE[\hat{s}_k]+\frac{1}{M^2}\sum_{k=0}^{M-1}\sum_{\substack{l=0\\l\neq k}}^{M-1}E\left[(\hat{s}_k-s)(\hat{s}_l-s)\right]. \quad (5)$$

In other terms, multi-block acquisitions logically average, and may mostly preserve, the expected value and bias of the estimator, whereas its variance and MSE are altered accordingly to the covariance terms within the $\hat{s}_k$ series. Depending on how each individual MSE evolves and on how related are the $\hat{s}_k$ series, multiple data blocks will exhibit different behaviors.

For an application to our recorded data and comparison purposes, we restrict our approach to blocks of equal length B regularly separated by interval L. All the estimators $\hat{s}_k$ then exhibit identical characteristics so that:

$$\forall k, E[\hat{s_k}]=E_B \;;\; V[\hat{s_k}]=V_B \;;\; B[\hat{s_k}]=B_B \;;\; MSE[\hat{s_k}]=MSE_B. \quad (6)$$

The above forms become:

$$E[\hat{s}]=E_B \text{ and } B[\hat{s}]=B_B, \quad (7)$$

whereas

$$V[\hat{s}]=\frac{1}{M}V_B+\frac{1}{M^2}\sum_{k=0}^{M-1}\sum_{\substack{l=0\\l\neq k}}^{M-1}E\left[(\hat{s}_k-E[\hat{s}])(\hat{s}_l-E[\hat{s}])\right] \quad (8)$$

and

$$MSE[\hat{s}]=\frac{1}{M}MSE_B+\frac{1}{M^2}\sum_{k=0}^{M-1}\sum_{\substack{l=0\\l\neq k}}^{M-1}E\left[(\hat{s}_k-s)(\hat{s}_l-s)\right]. \quad (9)$$

The block length B explicitly conditions all these terms, and the interval length essentially appears in cross products. However, depending on the order of the estimated quantity s, both parameters will implicitly appear in all these forms.

2.2 Estimation of first and second order moments

The first quantities of interest are generally the mean, variance and covariance. Given N equally spaced observations of a random variable x of mean value μ and standard deviation σ, the sample arithmetic mean is simply expressed as:

$$m=\frac{1}{N}\sum_{n=0}^{N-1}x(n), \quad (10)$$

the covariance as

$$\forall l, \forall i, C(l)=E\left[(x_i-\mu)(x_{i+|l|}-\mu)\right], \quad (11)$$

and the sample covariance as

$$\forall l, c(l)=\frac{1}{N-|l|}\sum_{i=0}^{N-|l|-1}(x_i-m)(x_{i+|l|}-m). \quad (12)$$

Not that depending on how one needs to handle both bias and MSE, the denominator N-|l| may be replaced by N, N-1 or N-|l|-1.

The variance σ^2=C(0) of x is then expressed as

$$s^2=c(0)=\frac{1}{N}\sum_{i=0}^{N-|l|-1}(x_i-m)^2 \text{ or } s^2=c(0)=\frac{1}{N-1}\sum_{i=0}^{N-|l|-1}(x_i-m)^2. \quad (13,14)$$

It is well established that m is an unbiased estimator of μ that exhibits a variance:

$$V[m]=MSE[m]=\frac{\sigma^2}{N}\left[1+2\sum_{l=1}^{N-1}\left(1-\frac{l}{N}\right)\rho(l)\right], \quad (15)$$

in which ρ(l) is the auto-correlation, the normalized auto-covariance C(l)/σ^2, of x [6] .

This expression indicates that the estimation of μ from discrete samples of x does not follow the well-known behavior V[m]=σ^2/N encountered for identically and independently distributed (iid) data as soon as the samples of x are not mutually independent. The increase of V[m] corresponding to positive values of ρ(l), as it is encountered in time-resolved flow velocity measurements, that contain lags smaller than the integral time scale of the signal, may easily be

compensated by increasing N accordingly. Additionally, oscillations of ρ(l) corresponding to periodical components of the flow can significantly alter V[m] depending on the length of the recording, the time scales and temporal coherence of these components.

Note that Eqs. (15), (4) and (5) are built the exact same way, illustrating how the variance, and MSE, of an averaged multi-block estimator are comparable to those of the sample mean.
Equation (15) also demonstrates that, for sufficiently large N and integrable ρ, the variance of m is altered by a discrete approximation of the integral time scale of x, which, in continuous form, formally writes [7]:

$$\bar{\tau}_x = \int_{l=0}^{\infty} \rho_x(l)\, dl \ . \tag{16}$$

Similarly, the estimation of c(l) exhibits a variance in which fourth order moments are involved. Bartlett [8] has shown that for normally distributed variables:

$$V[c(l)] = \frac{\sigma^4}{N} \sum_{k=-N+|l|+1}^{N-|l|-1} \left(\rho^2(k) + \rho(k+l)\rho(k-l)\right) - 4\rho(l)\rho(k)\rho(k-l) + 2\rho^2(l)\rho^2(k) \ , \tag{17}$$

in which the cross products of ρ illustrate how these estimates become less accurate as mutually correlated samples are used.

2.3 Multi-block estimators

As shown in Eqs. (4) and (8), the variance of a multi-block estimator is governed by two types of terms: an average of individual variances and a sum of covariances. If statistically independent blocks are recorded, the covariance terms vanish, otherwise, they can contribute significantly to the formulation. It is quite straightforward to obtain the variance of the sample multi-block mean defined for equally long and spaced blocks as:

$$m = \frac{1}{MB} \sum_{k=0}^{M-1} \sum_{i=0}^{B-1} x_{kL+i} = \frac{1}{MB} \sum_{k=0}^{M-1} m_k \tag{18}$$

Leading to

$$V[m] = \frac{\sigma^2}{MB} \left[1 + 2 \sum_{l=1}^{B-1} \left(1 - \frac{l}{B}\right) \rho(l) + \frac{1}{M^2} \sum_{i=0}^{M-1} \sum_{\substack{j=0 \\ j \neq k}}^{M-1} Cov\left[m_i(l), m_j(l)\right] \right], \tag{19}$$

which results in summing up correlations along larger blocks lags.
For the auto-covariance estimator c(l):

$$c(l) = \frac{1}{M(B-|l|)} \sum_{k=0}^{M-1} \sum_{i=0}^{B-|l|-1} (x_{kL+i} - m)(x_{kL+i+l} - m) = \frac{1}{M(B-|l|)} \sum_{k=0}^{M-1} c_k(l) \ , \tag{20}$$

obtaining the variance is more tedious without strong assumptions, but a starting formulation is:

$$V[\hat{c}(l)] = \frac{\sigma^4}{MB} \sum_{k=-N+|l|+1}^{N-|l|-1} \left(\rho^2(k) + \rho(k+l)\rho(k-l) - 4\rho(l)\rho(k)\rho(k-l) + 2\rho^2(l)\rho^2(k)\right) + \frac{1}{M^2} \sum_{i=0}^{M-1} \sum_{\substack{j=0 \\ j \neq k}}^{M-1} Cov\left[c_i(l), c_j(l)\right] \ , \tag{21}$$

which can be further developed using Bartlett's [8] formulae.

3-RESULTS

The behaviors of multi-block estimators are benchmarked against conventional estimates from random distributions of points using a bootstrap method [9] applied to the LDV data.

Fig. 3 shows the evolution of V[m] and V[s^2] for increasing sample sizes when estimators m and s^2 are applied to the streamwise component u of the velocity measured 1.5 chords downstream of the leading edge of the airfoil. Bootstrapped random samples, which, if totally iid, should verify $V[m]=\sigma^2/N$ and $V[s^2]=2\sigma^4/N$, are compared to bootstrapped blocks to illustrate the possible deviation of these quantities from the theory.
The theoretical variances of the estimators are computed assuming that all blocks of data are independent, and superimposed to the plot, showing how their quality may be predicted;
The theoretical behavior of V[m] perfectly agrees with the bootstrap simulation up to 10^4 samples per block. Beyond this value, the few millions of available LDV points cannot ensure that all the bootstrapped blocks are sufficiently

independent to validate the comparison. Comparatively, the approximated estimates of $V[s^2]$ poorly compares to the bootstrapped results beyond approximately 100 samples/blocks. This discrepancy may be also due to the quality of the bootstrap method applied to s^2, as the bootstrapped results for random samples also largely deviate from theory.

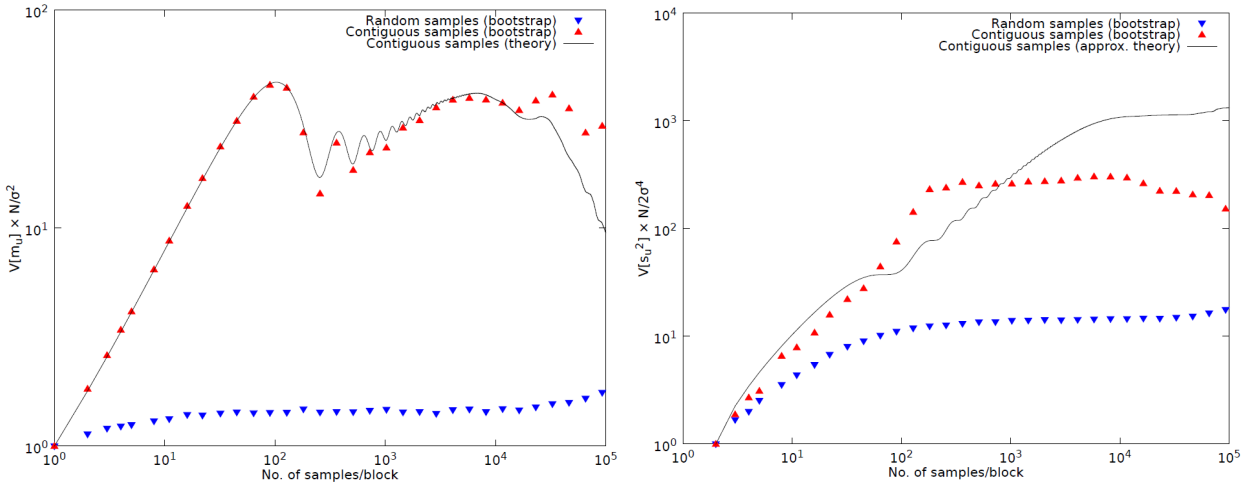

Figure 3 Estimation (bootstrap) and prediction of the variance of estimator m (10) and s^2 (14) applied to the longitudinal velocity measured by LDV 1.5 chords downstream of the leading edge of the NACA0015 airfoil.

Figure 4 presents the ratio of variance $V[m_u]$ on its theoretical value for iid variables, and estimated over the full LDV measurement domain for the three block sizes used in the TR-PIV experiment. These plots demonstrate how the choice of the length of measurement blocks can alter the evaluation of averaged quantities, as this ratio is also a measure of the equivalent number of samples needed to reach the statistical relevance of iid data. Larger time-resolved block sizes, while offering an improved view of the spatio-temporal features of the flow, tend to fully integrate the correlation functions, resulting in a degraded relevance of the statistical quantities. On the other hand, smaller sample sizes integrating a limited part of the correlation functions restrict this effect, but need advanced post-processing tools to allow spatio-temporal analyses of the flow. A rough compromise can be found in limiting the blocks sizes so that they comply with the dominant frequencies encountered in the flow, especially for flows with large integral time scales.

CONCLUSION

A study of the statistical relevance of limited, auto-correlated, datasets is conducted in order to assess how TR-PIV experiments may be designed to improve the estimation of averaged quantities. LDV measurements have been taken for reference to extract the expected values and variances of single and multi-block estimators, demonstrating their sensitivity to the distribution of data samples.
Similar analyses focused on the other statistical and spectral quantities classically estimated from PIV or TR-PIV may be usefully carried out. For this purpose, a full range of time-series analysis tools is available, from which the quality of parametric and non parametric analysis methods may be qualified.

ACKNOWLEDGMENTS

This work is supported by the European FP7 Contract AFDAR (Advanced Flow Diagnostics for Aeronautical Research).

REFERENCES

[1] J. Stafford, E. Walsh, V. Egan, A statistical analysis for time-averaged turbulent and fluctuating flow fields using Particle Image Velocimetry, Flow Measurement and Instrumentation 26 (2012), pp 1–9
[2] R. Hain, C. J. Kähler, Fundamentals of multiframe particle image velocimetry (PIV), Experiments in Fluids 42 (2007), pp 575-587
[3] A. Sciacchitano, F. Scarano, B. Wieneke, Multi-frame pyramid correlation for time-resolved PIV, Experiments in Fluids 53 (2012), pp 1087-1105
[4] Y.J. Jeon, L. Chatellier, P. Braud, L. David, Evaluation of fluid trajectory in time-resolved PIV, 10th International Symposium On Particle Image Velocimetry – PIV13, Delft, The Netherlands, July 2-4, 2013
[5] B. Tremblais, L. David, D. Arrivault, J. Dombre, L. Chatellier, L. Thomas, SLIP: Simple Library for Image Processing (Version 1.0) [Software, CeCILL-C Licence], University of Poitiers, France, http://www.sic.sp2mi.univ-poitiers.fr/slip/, 2010.

[6] D.B. Percival, A.T. Walden, Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques, Cambridge University Press, Cambridge, 1993
[7] S.B. Pope, Turbulent Flows, Cambridge University Press, Cambridge, 2000
[8] M. S. Bartlett , On the Theoretical Specification and Sampling Properties of Autocorrelated Time-Series , Supplement to the Journal of the Royal Statistical Society, Vol. 8, No. 1 (1946), pp. 27-41
[9] B. Efron, Bootstrap methods: Another look at the jackknife. Ann. Statist. 7 1–26, 1979

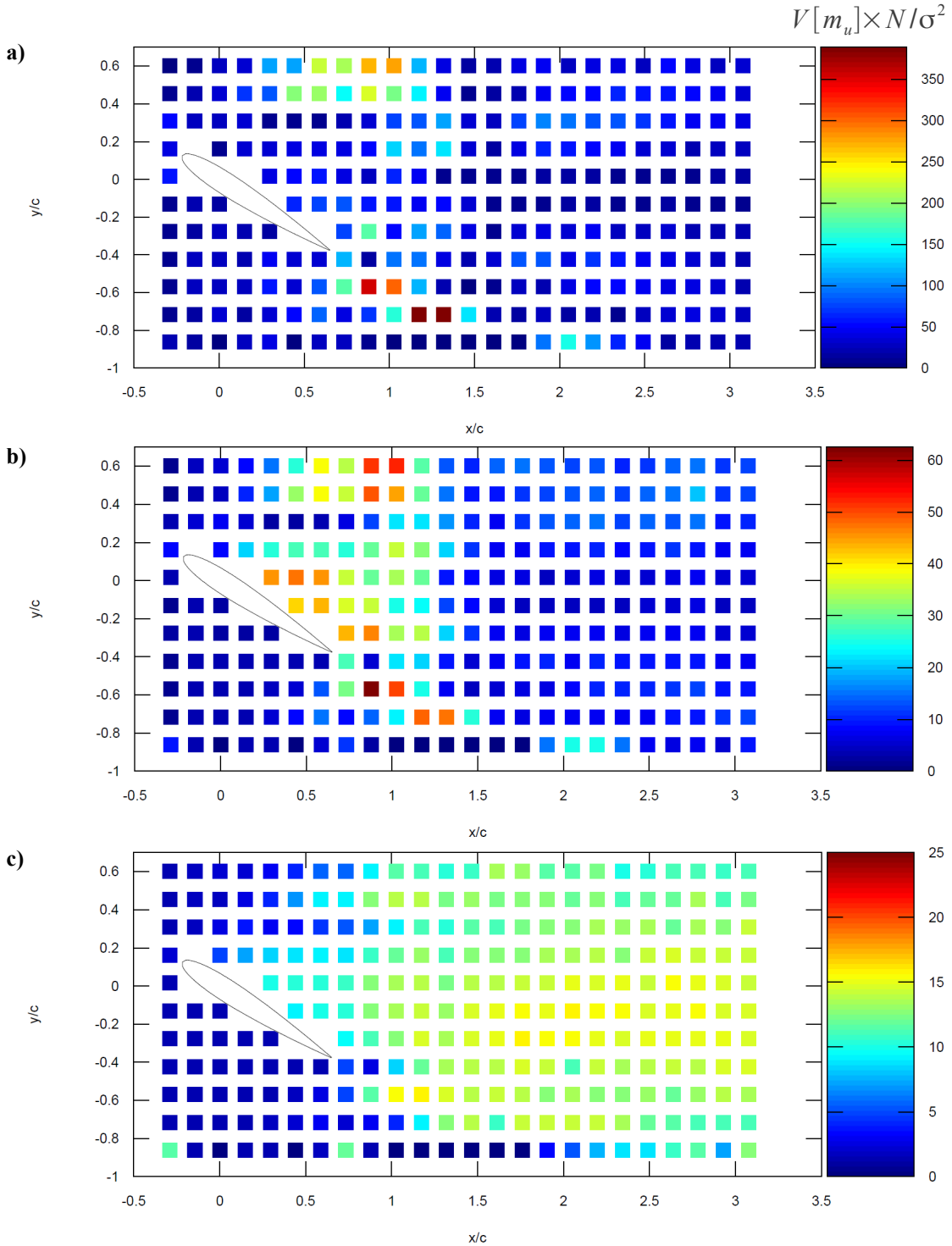

Figure 4 Ratio of variance $V[m_u]$ on its theoretical value for iid variables as a function of the sample size, and estimated over the full LDV measurement domain: **a)** maximum block size (2184 images); **b)** One vortex shedding period equivalent (280 images); **c)** reduced block size (24 images)