

# Assessing the Limitations of Effective Number of Samples for Finding the Uncertainty of the Mean of Correlated Data

## **Barton L. Smith**

Mechanical and Aerospace Engineering  
Utah State University  
Logan, Utah 84322  
Email: bsmith@engineering.usu.edu

## **Douglas R. Neal**

LaVision Inc.  
211 W. Michigan Ave. STE 100  
Ypsilanti, MI 48197  
Email: dneal@gmx.com

## **Mark Feero**

Institute for Aerospace Studies  
University of Toronto  
Toronto, ON M3H 5T6  
Email: m.feero@mail.utoronto.ca

## **Geordie Richards**

Mechanical and Aerospace Engineering  
Utah State University  
Logan, Utah 84322  
Email: geordie.richards@usu.edu

## **Abstract.**

The efficacy of recent and classical theories on the uncertainty of the mean of correlated data have been investigated. A variety of very large data sets make it possible to show that, under circumstances that are often too expensive to achieve, the integral time scale can be used to determine the effective number of independent samples, and therefore the uncertainty of the mean. To do so, the data set must be sufficiently large that it may be divided into many records, each of which is many integral time scales long. In this circumstance, all lags of the autocorrelation should be integrated to determine the integral scale. Some secondary findings include that the classical definition of the integral time scale goes identically to zero if a single record of any length is used and demonstration that measuring the integral scale requires ensemble averaging. Estimation of the integral time scale for a single record requires that the integration of the autocorrelation be truncated. This works well for signals where anti-correlation is not present. Additionally, for anti-correlated samples, the effective number of samples exceeds the number of acquired samples.

*Keywords:*

## 1. Introduction

Measurements often seek to determine statistical quantities of a process, such as the mean or standard deviation. The uncertainty of these quantities is a function of the number of samples acquired,  $N$ , among other variables. The manner in which the estimate of the statistics converges with  $N$  is well-known for independent samples. For samples that are positively correlated to one another, convergence happens more slowly. Since one needs to control the random uncertainty of the mean by choosing the sample number, it is important to be able to estimate this convergence rate, even for dependent samples.

The mean of an independently sampled random variable is itself a random variable with its own statistics. For data that are independent, the standard deviation of the mean is the standard deviation of the samples divided by the square root of the number of samples acquired.

$$s_{\bar{x}} = \frac{s_X}{\sqrt{N}} \quad (1)$$

In their seminal textbook on turbulence, Tennekes and Lumley [1] showed that ensuring independent samples requires that the sampling interval is greater than twice the integral time scale,  $T_u$ . This time scale is commonly defined as the integral over all time-lags of the autocorrelation of the signal,

$$T_u = \int_0^\infty \rho(\tau) d\tau. \quad (2)$$

Noting that the autocorrelation is symmetric and for a discrete signal,

$$2T_u \approx \left( \sum_{i=-\infty}^{\infty} \rho_i \right) \Delta t, \quad (3)$$

where  $\Delta t$  is the sampling interval.

More recently, it has been shown by Sciacchitano and Wieneke [2] that the effective number of samples of a general signal (not limited to turbulent flows) can be found using knowledge of the integral time scale. They took the unique approach of propagating random uncertainty through the definition of the time-mean using a Taylor Series Method [3] including the correlated terms that involve the autocorrelation of the samples. They concluded that the number of effective samples is generally less than the number of actual samples when the data are not independently sampled, and that the scaling depends on integral time scale,

$$N_{eff} = \frac{N}{\sum_{i=-\infty}^{\infty} \rho_i} = \frac{N}{2T_u/\Delta t}, \quad (4)$$

where  $N$  is the number of samples acquired. Clearly infinite lags are not available and these authors suggest, as is common, to terminate the summation when the autocorrelation first crosses zero.

Some of the issues with attempting to use information about the integral time scale to estimate the number of effective samples were discussed in [4]. These authors pointed out that no consensus exists on the lags over which the autocorrelation should be summed, and that any estimate of the integral time scale depends on this choice as well as the length of the record used and the number of lags computed. They concluded that the integral time scale is not useful for determining the number of independent samples and recommended that bootstrapping methods be used instead. Some of these same issues are explored in the present paper based on a larger dataset, and a more extensive comparison of statistical methodologies are the subject of a companion paper. This work, and our companion work comparing broader statistical methodologies [5], are focused on uncertainty analysis techniques that are model-free, meaning that the relevant equations for uncertainty estimation, such as Eq. 4, are easily applied without prior knowledge of the underlying stochastic process. Our focus was not selected to suggest that alternative Bayesian methods (see e.g. [6]) are inferior, but rather to elucidate some of the challenges and considerations required when implementing frequentist techniques that are common in the fluids literature.

In the present work, we note that the definition of the integral time scale in Eq. 3 is problematic since that quantity, if computed from a single record, is identically zero. According to the Wiener-Khinchin theorem, for a wide sense stationary random process, the power spectrum is the Fourier Transform of the autocorrelation [7]:

$$S(f) = \mathcal{F} \{ \rho(\tau) \}, \quad (5)$$

or

$$S(f) = \int_{-\infty}^{\infty} \rho(\tau) e^{-2\pi i f \tau} d\tau. \quad (6)$$

A stochastic process is wide sense stationary if its mean and autocovariance do not vary with respect to time. Any strictly stationary process which has a mean and a covariance is also WSS. We note the zero frequency component of the power spectrum of a signal with the mean removed is zero, so

$$0 = S(f=0) = \int_{-\infty}^{\infty} \rho(\tau) d\tau. \quad (7)$$

That is, for a mean zero WSS random process, it holds that

$$\int_0^{\infty} \rho(\tau) d\tau = \frac{1}{2} \int_{-\infty}^{\infty} \rho(\tau) d\tau = 0. \quad (8)$$

Comparing to Eq. 2, for a mean zero WSS random process, the integral time (or length, via Taylor's hypothesis) scale,  $T_u = 0$ . This surprising result is confirmed experimentally herein (cf. Figs 8-9). In practice, most experimentalists truncate the integration at a zero crossing of the autocorrelation to compute  $T_u$  and/or ensemble-average the data.

We analyze five signals to investigate the uncertainty of the mean. One is the velocity in a turbulent jet measured using a single calibrated hot wire. Two are pressure signals generated by flow through a confined array of cylinders. The final two are hot-wire velocity measurements

performed in the boundary layer of an airfoil. In each case, when the integral time scale is computed based on sufficiently long records (a length many times the integral time scale), the effective number of samples can be used to compute an accurate random uncertainty of the mean for cases where data were acquired too fast for independent samples.

## 2. Theory

We begin with an elementary derivation that will be useful for comparison with practical methods. The sample mean, or estimate of the mean, computed from  $N$  samples  $X_1, \dots, X_N$  is

$$\bar{X}_N = \frac{1}{N} \sum_{i=1}^N X_i. \quad (9)$$

Let us assume that the samples are identically distributed with the unknown mean  $\mu$ , drawn from a WSS and ergodic random process. We aim to estimate the uncertainty of the sample mean  $\bar{X} = \bar{X}_N$  in terms of the autocorrelation of the samples, and to this purpose we compute

$$\begin{aligned} (\bar{X}_N - \mu)^2 &= \left( \frac{1}{N} (X_1 + \dots + X_N) - \frac{1}{N} (\mu + \dots + \mu) \right)^2 \\ &= \left( \frac{1}{N} \sum_{i=1}^N (X_i - \mu) \right) \left( \frac{1}{N} \sum_{k=1}^N (X_k - \mu) \right) \\ &= \frac{1}{N^2} \sum_{i,k=1}^N (X_i - \mu)(X_k - \mu) \\ &= \frac{1}{N^2} \left( \sum_{i=1}^N (X_i - \mu)^2 + 2 \sum_{i=1}^{N-1} \sum_{k=i+1}^N (X_i - \mu)(X_k - \mu) \right). \end{aligned} \quad (10)$$

where the last equality is obtained by invoking symmetry in  $i$  and  $k$ . For each pair  $i, k$ , with  $i < k$ , we consider the index corresponding to the lag  $j = k - i$ , and rewrite the sum as

$$\begin{aligned} (\bar{X}_N - \mu)^2 &= \frac{1}{N^2} \left( \sum_{i=1}^N (X_i - \mu)^2 + 2 \sum_{i=1}^{N-1} \sum_{j=1}^{N-i} (X_i - \mu)(X_{i+j} - \mu) \right) \\ &= \frac{1}{N^2} \left( \sum_{i=1}^N (X_i - \mu)^2 + 2 \sum_{j=1}^{N-1} \sum_{i=1}^{N-j} (X_i - \mu)(X_{i+j} - \mu) \right). \end{aligned} \quad (11)$$

In practice the experimentalist does not have access to the true mean  $\mu$ , but may be able to perform repeated experiments to produce a “parent mean”  $\bar{X}_M$  from a larger collection of  $M \gg N$  samples, where  $M$  is divisible by  $N$ , and substitute this approximation for  $\mu$  in Eq. 11. The experimentalist could further perform an ensemble average of this estimate over the  $M/N$  ensembles of the “record mean”  $\bar{X}_N$ , which is approximately equivalent to computing the sample variance of the mean computed from these ensembles. If we denote ensemble averaging over the  $M/N$  ensembles with  $\overline{(\cdot)}$ , this estimate becomes

$$s_{\bar{X}}^2 = \overline{(\bar{X}_N - \bar{X}_M)^2}. \quad (12)$$

By invoking Eq. 11 for each ensemble, with  $\bar{X}_M$  replacing  $\mu$ , Eq. 12 becomes

$$s_{\bar{X}}^2 = \frac{s_{\bar{X}}^2}{N} \left( 1 + 2 \sum_{j=1}^{N-1} \hat{\rho}_j \right), \quad (13)$$

where  $s_X^2$  is the sample variance computed from all  $M$  samples, and  $\hat{\rho}_j$  is the biased sample autocorrelation coefficient

$$\hat{\rho}_j = \hat{\rho}_j(\bar{X}_M) = \frac{\frac{1}{N} \sum_{i=1}^{N-|j|} (X_i - \bar{X}_M)(X_{i+|j|} - \bar{X}_M)}{s_X^2}, \quad (14)$$

computed separately for each ensemble using the parent mean  $\bar{X}_M$ . Note that this is equivalent to taking

$$s_X^2 = \frac{s_X^2}{N_{eff}} \quad (15)$$

where the effective number of samples is

$$N_{eff} = \frac{N}{2T_u/\Delta t}, \quad (16)$$

provided we have approximated the integral time scale as

$$2T_u = \left( 1 + 2 \sum_{j=1}^{N-1} \bar{\rho}_j \right) \Delta t. \quad (17)$$

Eq. 13 gives a useful estimate for the uncertainty of the sample mean, provided  $M$  is selected appropriately depending on  $N$ . This estimator appears in the literature under a variety of guises. For example, it is the estimate of uncertainty produced by the method of non-overlapping batch means [8], as evidenced by Eq. 12. However, for expensive data, where the experimentalist cannot afford large  $M$ , this approach becomes less effective.

Given access to a single record only, so that  $M = N$ , the approach advocated in the fluids literature [2] is to truncate the sum of autocorrelation coefficients appearing in Eq. 13. That is, the approach is to take

$$s_X^2 = \frac{s_X^2}{N} \left( 1 + 2 \sum_{j=1}^k \hat{\rho}_j \right), \quad (18)$$

where the autocorrelation coefficient  $\hat{\rho}_j$  is computed using the record mean, meaning that

$$\hat{\rho}_j = \hat{\rho}_j(\bar{X}_N) = \frac{\frac{1}{N} \sum_{i=1}^{N-|j|} (X_i - \bar{X}_N)(X_{i+|j|} - \bar{X}_N)}{s_X^2}. \quad (19)$$

Importantly, the parameter  $k$  appearing in Eq. 18 is selected by the user. For example,  $k$  may be taken to be the positive index  $j$  where the sample autocorrelation  $\hat{\rho}_j$  first becomes negative. We remark that Eq. 18 is also equivalent to the method of overlapping batch means applied on batches of size  $k$  selected from the single record [9]. Furthermore, note this approach is equivalent to using Eq. 15 and Eq. 16 provided we have redefined the integral time scale as

$$2T_u = \left( 1 + 2 \sum_{j=1}^k \hat{\rho}_j \right) \Delta t. \quad (20)$$

The latter approach degenerates completely if we select  $k = N - 1$ . That is, if we have a single record only, so that  $M = N$ , we observe from Eq. 12 that this results in  $s_X^2 = 0$ . This

potentially surprising result is demonstrated in Figs. 8 and 9 below. Note that this situation is not improved by taking an ensemble average if record statistics are used in each ensemble. As discussed in Sec. 1, this conclusion can also be reached from the Wiener-Khinchin Theorem: the integral of the autocorrelation is the power spectral density evaluated at zero, which is zero because the mean is removed from the signal prior to computation of the autocorrelation (see also [10]).

In what follows, we will explore issues pertaining to estimating the uncertainty of the time-average of dependent data. While the aim is to provide guidance for cases where the sample size is limited, our investigation is based on very large datasets that will serve as a benchmark. These are described next. This is followed by a discussion of issues in computing the autocorrelation. Results are then shown for ensemble-averaged cases. If the data set is too small to subdivide without resulting in a record of sufficient length relative to  $T_u$ , bootstrapping and other statistical methods may be a more accurate approach for determination of the uncertainty of the mean [11]. This is the topic of our companion paper [5].

### 3. Experimental Data Sets

Data were acquired for four setups in three facilities. The first two of these are located at Utah State University. The first set of data are hot-wire measurements of velocity in the turbulent jet used in previous studies [12, 13]. Details of the facility can be found in these references, but are not important to the present work. Data were acquired  $30h$  downstream on the jet centerline. A short sample of the velocity data is shown in Fig. 1. Clearly, this sample is time-resolved and the samples are not independent.

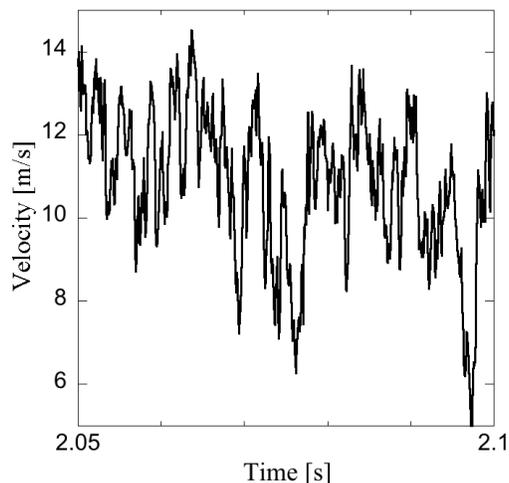


Figure 1: A portion of the jet velocity sample.

A facility for studying flow in a confined array of cylinders [14, 15] was used for the next two sets of data. For the first of these sets, a full array of cylinders was present in the facility, as shown in Fig. 2. In the second case, a single cylinder was used as shown in Fig. 3. The

signal comes from a flush-mounted high-frequency-response pressure sensor. The location of the pressure sensor in both cases is shown by an arrow.

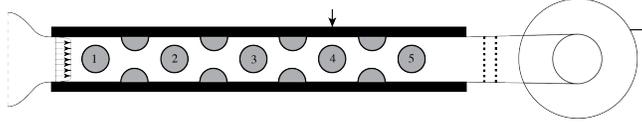


Figure 2: The facility and setup used for the second set of data. Flow is from left to right and the location of the pressure sensor is indicated with an arrow. This case is referred to as “array”.

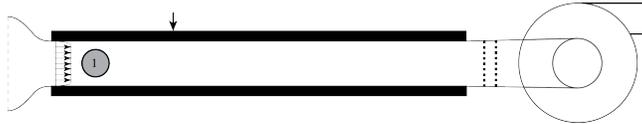


Figure 3: The facility and setup used for the third set of data. Flow is from left to right and the location of the pressure sensor is indicated with an arrow. This case is referred to as “one cylinder”.

For each of these three cases,  $N_{\text{sets}} = 10$  sets of 100,000 data points were acquired at range of data acquisition rates ranging from time-resolved to independent. The sampling rates,  $sr$ , used for each flow are shown in Table 1. We note that these sets were each acquired at a different time.

Table 1: Parameter space for first three data sets.

Case	Sampling Rates, $sr$ (kHz)	$\Delta t/T_u$ range
Jet Velocity	0.1, 0.2, 0.25, 0.4, 0.5, 0.75, 0.8, 1, 1.6, 3.2, 5, 15	0.026 – 3.9
One Cylinder	0.05, 0.1, 0.15, 0.2, 0.5, 1, 2, 4, 10, 50	0.012 – 12.2
Array	0.05, 0.1, 0.15, 0.2, 0.5, 1, 2, 4, 10, 50	0.010 – 9.6

For the fourth and fifth dataset, hot-wire measurements of velocity in the boundary layer of a NACA 0025 airfoil were acquired from a wind tunnel facility at University of Toronto as shown in Fig. 4. At the Reynolds number and angle-of-attack of operation, the boundary layer experiences laminar separation near the leading edge. Velocity was measured at two locations: in the attached laminar boundary layer at 10% chord, and in the turbulent separated shear layer at 40% chord. These will be referred to as the “airfoil laminar” and “airfoil turbulent” cases, respectively. The wall-normal location at each chord location corresponded to the displacement thickness. Data were sampled at a single sampling rate of 5 kHz and at least  $N_{\text{sets}} = 100$  sets of 150,000 data points were acquired. These long, high sampling rate data sets can be used to simulate lower sampling rates through down sampling.

The purpose of the five sets was to provide a range of behavior in the autocorrelation of the signals. The jet and airfoil turbulent velocity data produce autocorrelations with classical

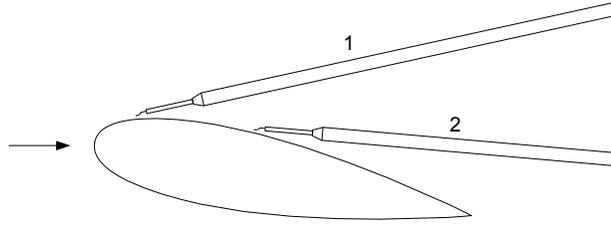


Figure 4: The facility and setup used for the fourth and fifth set of data. Position 1 shows the hot-wire in the laminar, attached boundary layer. Position 2 is for the turbulent, separated shear layer. The data from these cases are referred to as “airfoil laminar” and “airfoil turbulent”, respectively.

turbulent flow behavior. The array data, on the other hand, contains a strong coherent component, resulting in an oscillatory autocorrelation. The behavior of the one cylinder and airfoil laminar cases fall between these two behaviors.

The best estimate of the sum of  $\rho$  is determined by computing  $\rho$  from the highest sampling rate. Prior to computation of the autocorrelation, the time mean must be subtracted from the signal. An autocorrelation is computed for each of the  $N_{ens}$  records, and the resultant autocorrelations are ensemble averaged. These averaged autocorrelations are shown in Figs. 5a-e.

In this case, the mean that was subtracted from the signal was the mean of the same  $N_{rec}$  points that are used to compute each autocorrelation. The validity of this choice will be discussed further in Sec. 5.

The datasets are better understood by examining their respective histograms, shown in Fig. 6 a-e. Each set has a central tendency, and some appear Gaussian. An exception to this is the airfoil turbulent data due to the fact that negative velocity is rectified by the hot-wire probe. The pressure data are negative since the facility operates below atmospheric pressure, but we note that the time average value has no impact on any of the results of this study. Of the five, the one cylinder set has a more narrow distribution as well as a long lower tail. The laminar airfoil case has the same number of samples, but since these data remain correlated over a longer period than all other cases, the histogram appears much less smooth than the others. In spite of these differences, all of these sets will be shown to behave similarly with respect to uncertainty of the mean.

We note that for the smallest sampling rates, the experiments were run over a very long period. One million samples at 50 Hz corresponds to over 5.5 hours. Over this long time, some amount of drift in the mean is to be expected due to zero drift of the sensor or changes in the air properties due to temperature or atmospheric pressure. In fact, the mean of the pressure data drifts by about 4% in the worst case. This issue is easily handled by determining a “local” mean

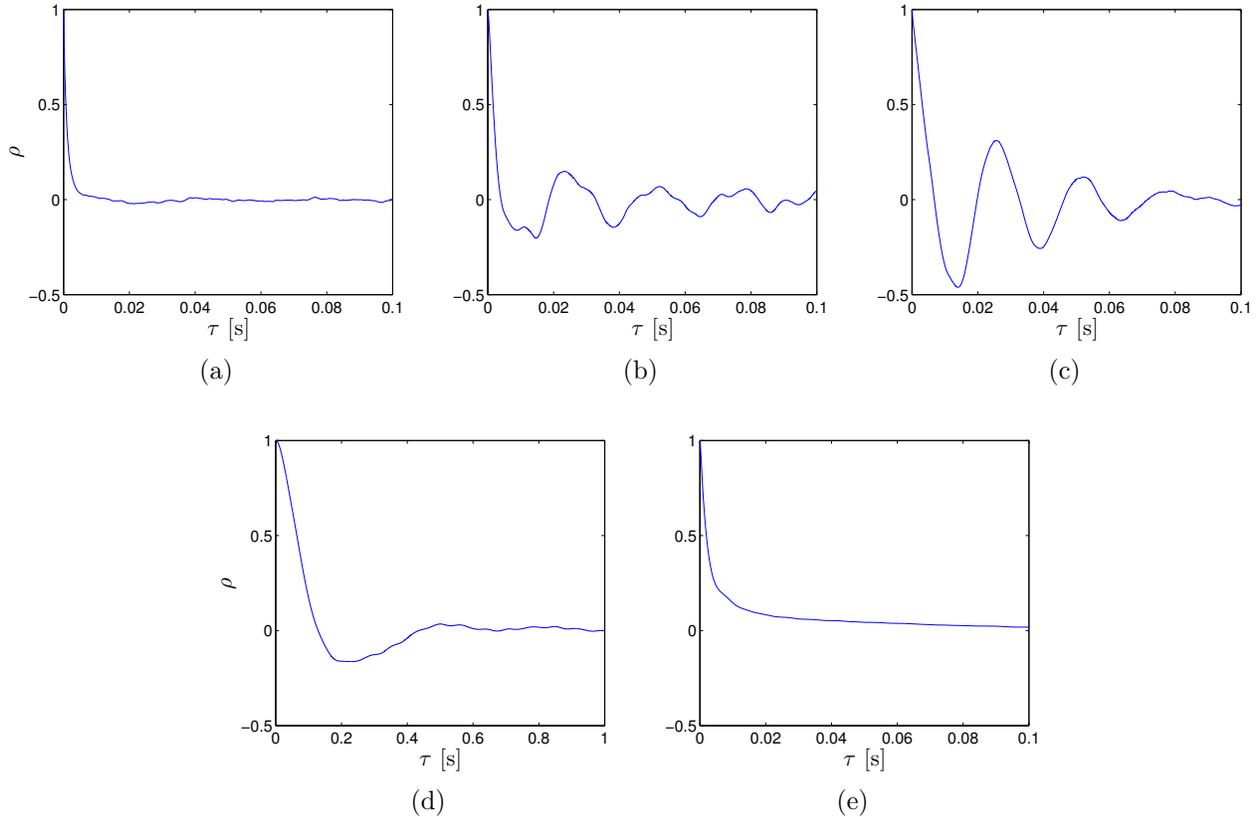


Figure 5: Ensemble averaged autocorrelations for (a) jet velocity, (b) one cylinder, (c) array, (d) airfoil laminar and (e) airfoil turbulent.

at all times based upon the Matlab command **smooth** with local regression using weighted linear least squares and a 2nd degree polynomial model. The smoothing kernel scales so it is long for the slow sampling and is the full dataset for the slowest sampling.

A code was written that could divide these full records into smaller records of arbitrary size  $N_{rec}$ , allowing for ensemble averaging over arbitrary and coupled record length and number of ensembles  $N_{ens}$ . Given the samples acquired, the relationship between these is  $N_{ens} = \text{ROUND}(M/N)$

#### 4. Computation of autocorrelation and integral time scale

Unfortunately, there are many methods to compute an autocorrelation, and subtle differences in the results can have a major impact on the sum of the autocorrelation and thus the integral time scale. Specifically, the three issues are: the sample estimates of the mean and variance used in forming the autocorrelation, the record length used, and the summation limit used when finding  $T_u$ . It is sometimes difficult to investigate these issues separately as they may depend on the same parameter (e.g. longer records result in better estimates of the mean). In this section, the large data sets described in Sec. 3 are used to determine which of these methods provides

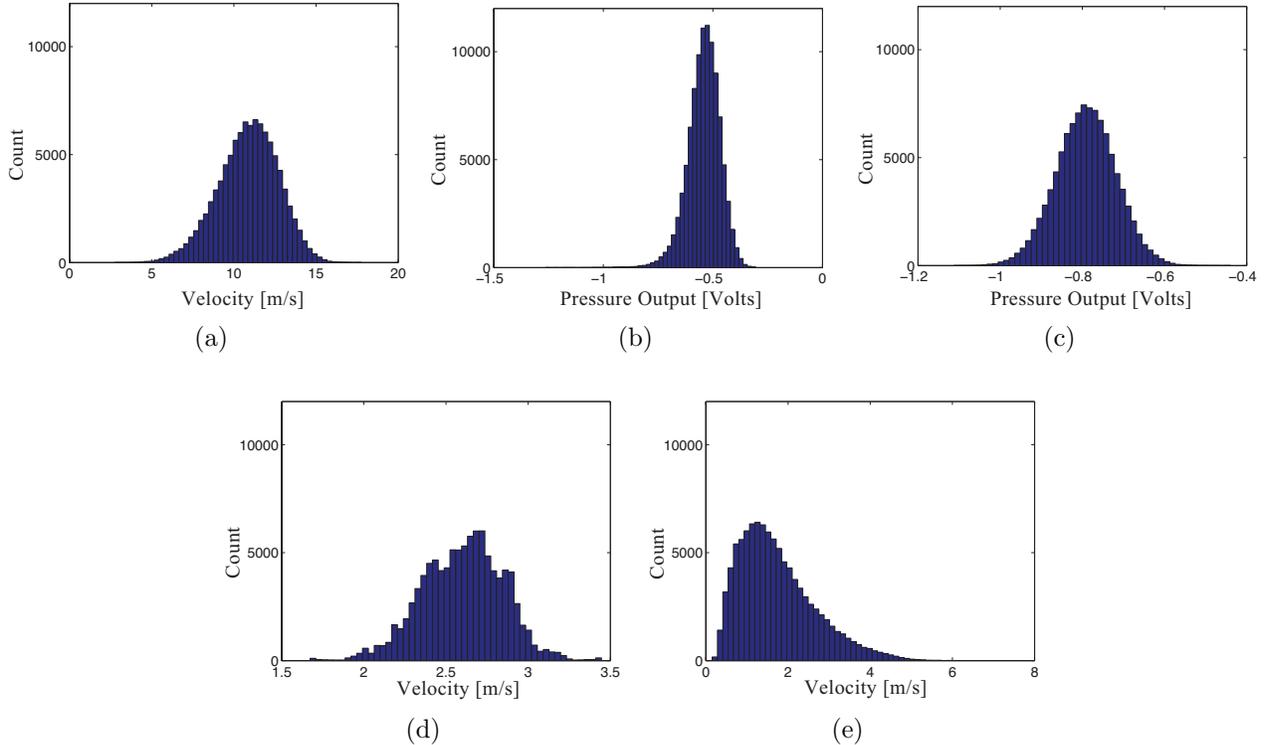


Figure 6: Histograms of 100,000 points for (a) jet velocity, (b) one cylinder, (c) array, (d) airfoil laminar and (e) airfoil turbulent.

results that match the theory in Sec. 2. In all that follows, the so-called “biased” estimate of the autocorrelation is used in keeping with Eq. 14 and as recommended by [10]. While these results are not shown, use of the unbiased estimate of the autocorrelation generally resulted in non-physical behavior at large lag values.

#### 4.1. Impact of the record length on the autocorrelation

In order to separate this issue from the impact of errors in the mean and variance, the results shown here are based on “parent population” estimates of the mean and variance. These are found based on the entire data set; i.e., as in Eq. 14,  $\bar{X}$  and  $s_X$  are computed from the parent population.

In Fig. 7, the ensemble-averaged autocorrelation for the jet velocity data sampled at 15 kHz is shown as a function of the length of the record used to compute the autocorrelation relative to the integral time scale. The integral time scale is taken from the converged value obtained from long record lengths using parent statistics. We note that the autocorrelation drops toward zero more rapidly for the shorter records than for longer ones, and that the result seems to converge for records longer than  $20T_u$ . In fact, if these autocorrelations are integrated over all lags using parent statistics to find  $T_u$ , the value converges for records of length  $10T_u$ , which is consistent with the results of [16] who reported that domains of size  $5T_u - 10T_u$  were required to accurately

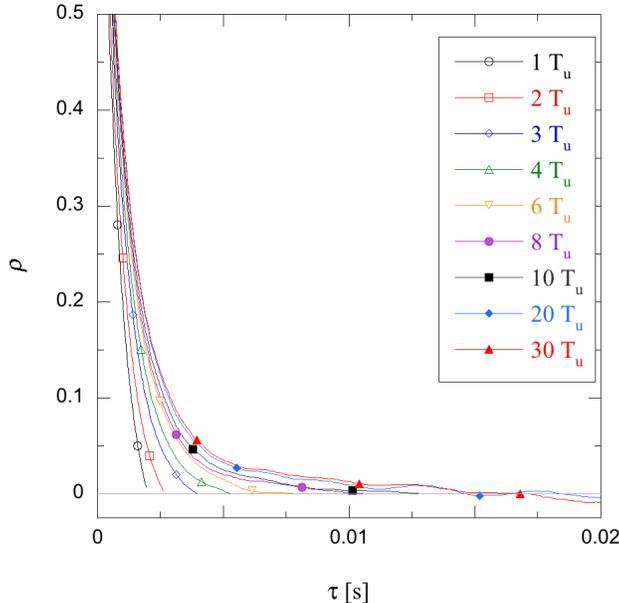


Figure 7: The autocorrelation computed using various record lengths relative to  $T_u$  for the jet velocity data.

compute the integral length scale.

#### 4.2. Choice of integration limit of autocorrelation

It will be shown below that the best estimate of the integral scale (and thus  $N_{eff}$ ) is found by ensemble-averaging autocorrelations that are computed using parent statistics. However, in many cases for small, single-record samples, only a single estimate of the autocorrelation is available. Examining the results for single records, the autocorrelation of the jet velocity data for several randomly chosen records of length  $80T_u$  are shown in Fig. 8. Note that in each case, when  $k = N - 1$ ,  $2T_u$  is estimated (incorrectly) to be zero. While not shown, the other data sets (and, indeed, every signal we have tried, no matter its length) behaved similarly in this respect, in keeping with Eq. 8 and the Wiener-Khinchin Theorem.

Since any use of record statistics, including single record or an ensemble-average with record statistics used to estimate the autocorrelation, results in an estimate of  $T_u$  equal to zero when all lags are summed, some limit  $k < N - 1$  must be chosen in Eq. 3. Several other schemes have been investigated [4], including the commonly employed limit of the first zero crossing of the autocorrelation. While this works somewhat well for the autocorrelation of the classical homogeneous turbulent flow (similar to Fig. 5a or e), it is not appropriate for a signal with an oscillatory autocorrelation (such as Fig. 5b or c). The signal remains correlated (or anti-correlated) well beyond the first zero crossing. We note that anti-correlated samples are not independent, and as shown in Sec. 5, have a smaller uncertainty of the mean for the same number of independent samples. While these are not velocity samples, we note that autocorrelations similar to these would result from measurements in regions of coherent motion, such as vortex

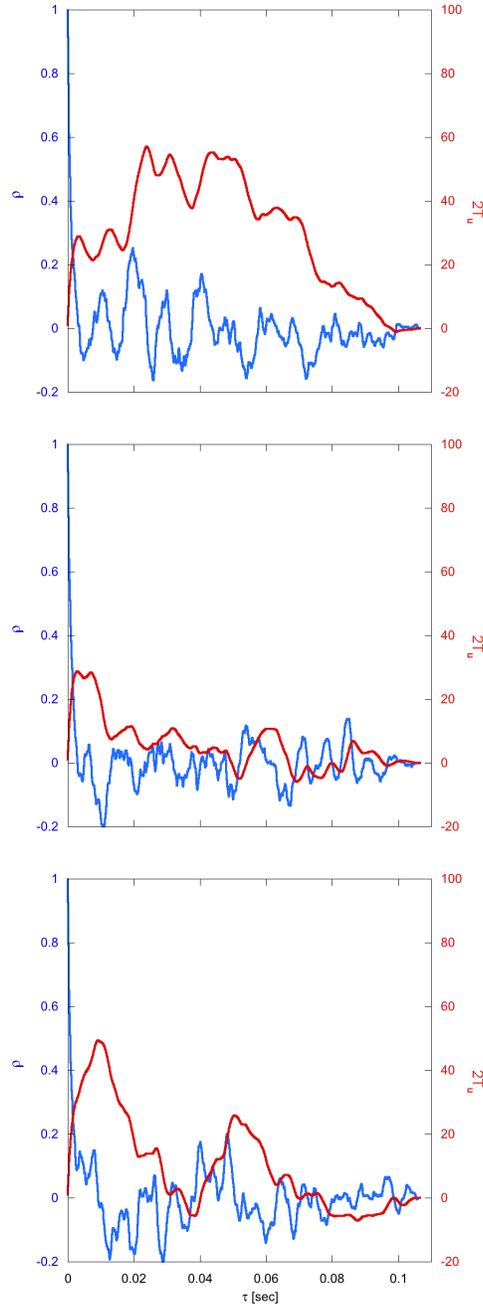


Figure 8: Autocorrelation (blue) and the sum of the autocorrelation, or  $2T_u$  (Eq. 3 based on integration up to  $\tau = k\Delta t$ , where  $\Delta t$  is the sample interval [s], red) for a record length of  $80T_u$ , for three arbitrarily chosen records of the jet velocity data.

shedding.

Several of these points are made in Fig. 9 which shows results for the jet velocity data (Fig. 9a) and the cylinder array data (Fig. 9b). The autocorrelation is plotted along with  $2T_u$  computed with Eq. 3 as a function of  $\tau = k\Delta t$  where  $\Delta t$  is the sampling interval. The heavy lines correspond to use of parent statistics to compute the autocorrelation while the thin lines correspond to record statistics. In other words, while both cases are ensemble averaged, for the heavy lines, the sample mean and variance of the entire data set is used, as in Eq. 14, while for the thin lines, the sample mean and variance of each record is used to compute each autocorrelation, as in Eq. 19. As mentioned above, the latter practice leads to the result that  $2T_u = 0$  when  $k = N - 1$  by the Wiener-Khinchin Theorem independent of the number of records used. For both cases (as well as others not shown), the sum of the autocorrelation does go to zero.

Using parent statistics in Eq. 14 provides the best estimate for  $2T_u$  for  $k = N - 1$  (i.e., heavy line at the right side of Figs. 9a-b). If the record statistics are to be used (for instance, when only 1 record is acquired), one must truncate the summation of the autocorrelation at a value of  $k < N - 1$ . The horizontal dashed lines in Fig. 9 show the value of  $2T_u$  that would result from choosing a  $k$  value corresponding to the first zero crossing of the autocorrelation while the dotted line corresponds to a  $k$  value chosen as the index corresponding to the minimum of the autocorrelation.

For the jet velocity data in Fig. 9a, the zero crossing choice results in a good estimate of  $2T_u$  as found from using parent statistics. This is not true of the cylinder array data in Fig. 9b, where the use of the minimum provides a very good estimate. Without a priori knowledge of the behavior of the autocorrelation, it is not clear which choice is generally superior. Recalling that these are ensemble average results, they do not suffer from random fluctuations in the autocorrelation that are common and clearly visible in Fig. 8. Such fluctuations make it difficult to accurately estimate  $T_u$  or  $s_{\bar{x}}$  based upon a single record no matter the choice in  $k$ . Bootstrapping methods, which are no more difficult to implement, are found to be more robust than using an estimate of  $T_u$  to find an  $N_{eff}$ . This is the topic of a companion paper [5].

## 5. Results

The integral time scale is computed and used to determine  $N_{eff}$  for each case. The parent mean is subtracted from the signal and the autocorrelation is integrated over all lags (i.e.  $k = N - 1$ ) to find  $T_u$ . The validity of  $N_{eff}$  is assessed by observing how often the uncertainty band based on  $s_{\bar{x}}$  covers the local mean. For Gaussian statistics, the uncertainty bands should include the mean 68% of the time. We use this  $1 - \sigma$  value based upon the recommendation of [3] that uncertainties be computed at  $1 - \sigma$  and combined before expansion to  $2 - \sigma$ . Fig. 10 show means formed using  $N_{rec} = 10,000$  samples for the jet velocity data and sampling rate  $sr = 50$  kHz. For this time-resolved case, the uncertainty bands based on  $N$  rarely include the actual mean. However, the uncertainty bands based on  $N_{eff}$  (Fig. 10b) do nearly 68% of the time as shown in Fig. 11.

The precise number of times that the uncertainty bands include the mean can be tracked

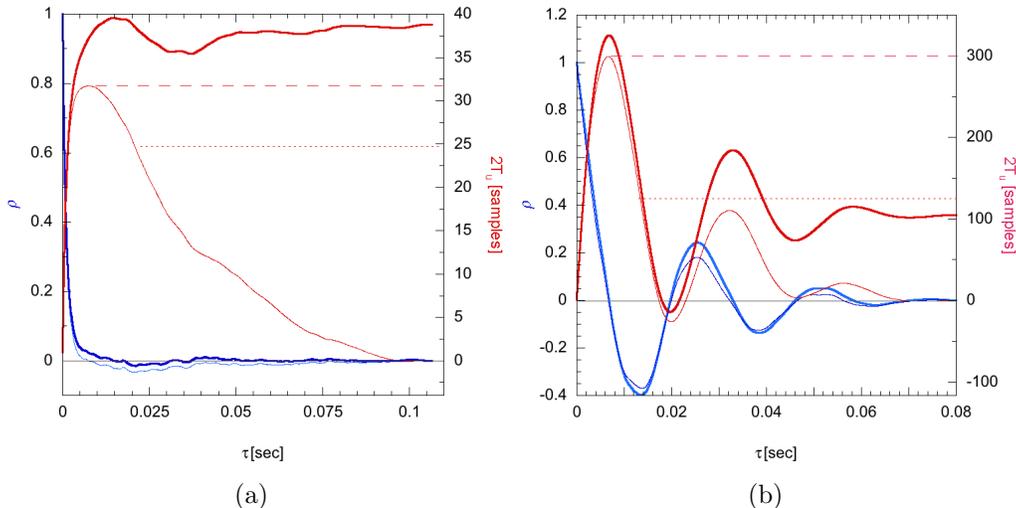


Figure 9: Ensemble average autocorrelation (blue) and the sum of the autocorrelation, or  $2T_u$  (Eq. 3 based on integration up to  $\tau = b\Delta t$ , where  $\Delta t$  is the sample interval [s], red) for a record length of  $80T_u$ . Heavy lines are based on parent statistics while thin lines are based on record statistics. The dashed line is  $2T_u$  based on record statistics and summing to the zero crossing of the autocorrelation while the dotted line is summed to the minimum of the autocorrelation. (a) Jet velocity data and (b) Cylinder array pressure data.

for all cases. Records of a large and somewhat arbitrary number of samples ( $80T_u$ ) are used to ensure in a sufficiently large  $M$ . The coverage is shown in Fig. 11. For independent sampling based on the claims of [1] ( $\Delta t/T_u > 2$ , where  $\Delta t = 1/sr$ , indicated by a vertical line) the jet velocity data behave as expected with near 68% of the uncertainty bands covering the mean. None of the turbulent or laminar separation results reach the independent regime. The two pressure data sets, however, have greater than 68% coverage. While both cases peak and return to 68% for longer  $\Delta t$  values, the pressure array case has a larger peak and maintains larger coverage than the single case. We note that this behavior is in line with the autocorrelations of the two signals in that the pressure array data generates a longer and larger magnitude negative correlation period than the single cylinder. We conclude from this behavior that negatively correlated samples result in faster mean convergence than uncorrelated data. As the samples become closer together in time (moving to the left on the plot), uncertainty bands based on  $N$  no longer cover the mean well, but uncertainty bands based on  $N_{eff}$  result in very close to 68% coverage, including the negatively correlated results near  $2T_u$ . We conclude, based upon this result, that Eq. 13 provides a good estimate, where the autocorrelation is computed based upon ensemble averages and using parent statistics.

To emphasize that the method described above arrives at the best possible estimate of  $T_u$ , the same data are used to compute ensemble-averaged autocorrelations based upon record statistics. These autocorrelations are then integrated to the first zero crossing to find  $T_u$ . Recall that integrating all lags will result  $T_u = 0$ . In Fig. 12, statistics based on records of length  $80T_u$  are used to form the autocorrelation. The jet velocity data result is somewhat similar

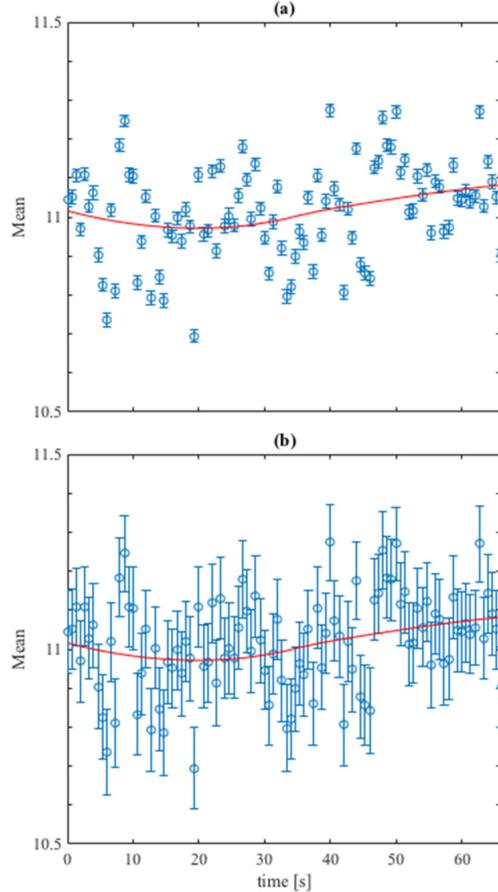


Figure 10: Time averages of jet velocity data formed with  $N_{rec} = 10,000$  samples and uncertainty bands based on (a)  $s/\sqrt{N}$  and (b)  $s/\sqrt{N_{eff}}$ . The red line indicates the local mean.

to the parent statistics result, but the cases with oscillatory autocorrelations now have very over-predicted uncertainty values and thus over-predicted coverage. Furthermore, since the abscissa of this plot also depends on  $T_u$ , these errors result in left-to-right scatter in the result.

## 6. Conclusions

Five datasets from various experimental setups were analyzed at different sampling rates. The sampling rates were carefully chosen to provide several sets of independent data and several sets of dependent data. For the cases where the data are sampled independently, the random uncertainty of the mean (or standard error) covers the actual mean about 68% of the time, as expected for Gaussian distributions. The cases with dependent samples had random uncertainties of the mean that covered the true mean less often than 68% (i.e. under-predicting the true uncertainty bands). The level of correlation between samples, as measured by the autocorrelation, predicts the extent to which the uncertainty of the mean covers the true mean.

As previously reported, an effective number of independent samples  $N_{eff}$  can be found based upon the autocorrelation. We show that  $N_{eff}$  is an effective method of assessing the

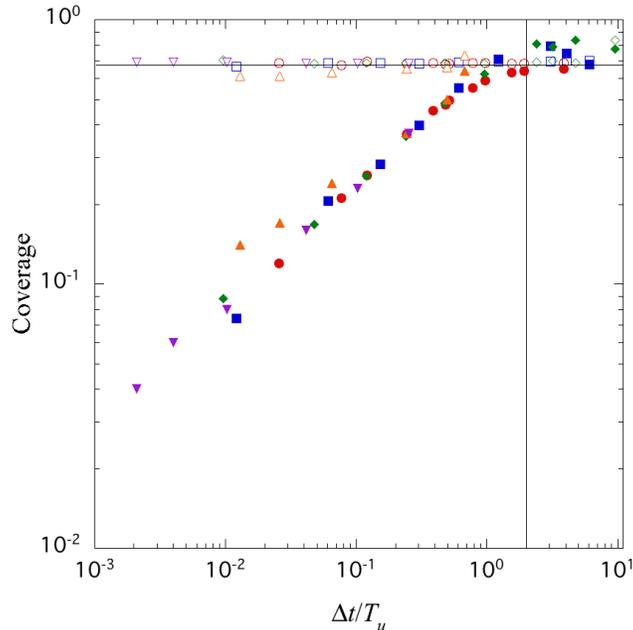


Figure 11: Using parent statistics and integrating the full autocorrelation, the percentage of cases where the uncertainty bands cover the mean for all three flows based both on  $N$  and  $N_{eff}$  as a function of the ratio of the sampling interval  $\Delta t$  to the integral scale. The horizontal line indicates 68% coverage, while the vertical line indicates 2. Values larger than 2 indicate independent sampling. Closed symbols are based on  $N$  and open symbols are based on  $N_{eff}$ . (●) jet velocity, (■) one cylinder, (◆) array, (▲) airfoil turbulent, (▼) airfoil laminar

actual uncertainty of dependent samples. However, one must be careful to ensure that the autocorrelation is computed accurately which requires a sufficiently long record (in terms of the integral time scale  $T_u$ ) to compute the autocorrelation accurately. Previous studies had suggested this need to be  $6T_u$ , but the present data suggest this is at least  $20T_u$ . The required length may depend on some other aspect of the data, such as the behavior of the autocorrelation itself.

Unfortunately, short records are more typically the norm for many current measurement technologies, such as particle image velocimetry (PIV) where the data are expensive (both in terms of storage capacity and processing time). This situation is further exacerbated by the trend towards high-speed or time-resolved acquisition (leading to dependent samples) that are also limited by camera RAM (leading to shorter records). In spite of this difficulty, it is straightforward to show that a data set is independently sampled by computing the autocorrelation and ensuring that it rapidly decays to zero.

When one has access to a long record, it is advised that this be divided into a number of subsets (records) and ensemble averaging be performed (using the parent statistics). However, instead of using the zero crossing, it is advised to integrate the normalized autocorrelation function to all lags (consistent with the accepted definition of the autocorrelation).

If one only has access to a short dataset, ensemble averaging is not practical and thus one

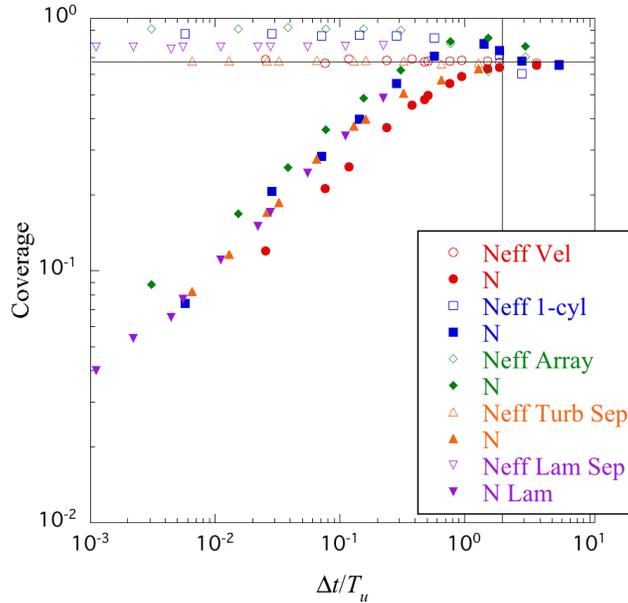


Figure 12: Using parent statistics and integrating the autocorrelation to the first zero crossing, the percentage of cases where the uncertainty bands cover the mean for all three flows based both on  $N$  and  $N_{eff}$  as a function of the ratio of the sampling interval  $\Delta t$  to the integral scale. The horizontal line indicates 68% coverage, while the vertical line indicates 2. Values larger than 2 indicate independent sampling. Closed symbols are based on  $N$  and open symbols are based on  $N_{eff}$ . ( $\bullet$ ) jet velocity, ( $\blacksquare$ ) one cylinder, ( $\blacklozenge$ ) array, ( $\blacktriangle$ ) airfoil turbulent, ( $\blacktriangledown$ ) airfoil laminar.

cannot integrate all lags to compute the integral scale. In this case, the integration must be truncated at some arbitrary lag. In some cases, the first zero crossing of the autocorrelation works well while in others the minimum of the autocorrelation may be better. The best performance depends on the nature of the autocorrelation (how negative it becomes and for how many lags), which cannot be known based on a small data set. We demonstrate that use of a zero crossing works well for a signal with small, short, negative autocorrelation, but tends to overestimate the integral scale (and thus the uncertainty of the mean) for more negative autocorrelations.

We demonstrate that the common definition of the integral time scale (the integral over all lags of the autocorrelation) is problematic in that it goes identically to zero if the autocorrelation is computed from a single record. If, on the other hand, many autocorrelations based on subsets of the data (with the mean of the entire dataset removed) are ensemble averaged, the integral scale can be accurately computed.

Finally we show that the surprising result that uncertainty of the mean of anti-correlated data is smaller for the same number of samples when compared to uncorrelated (independent) data. This means that  $N_{eff} > N$  for data that are anti-correlated.

## Acknowledgements

BLS would like to thank Professor Jacob Gunther for many productive conversations about the properties of the autocorrelation function and Austin Parker for acquiring the pressure data.

## References

- [1] Tennekes, H., and Lumley, J. L., 1992. *A first course in turbulence*. The MIT Press.
- [2] Sciacchitano, A., and Wieneke, B., 2016. “PIV uncertainty propagation”. *Measurement Science and Technology*, **27**(8), p. 084006.
- [3] Coleman, H. W., and Steele, W. G., 2009. *Experimentation, Validation, and Uncertainty Analysis for Engineers*, 3rd ed. John Wiley and Sons, Hoboken, NJ.
- [4] Theunissen, R., Di Sante, A., Riethmuller, M., and Van den Braembussche, R., 2008. “Confidence estimation using dependent circular block bootstrapping: application to the statistical analysis of PIV measurements”. *Experiments in Fluids*, **44**(4), pp. 591–596.
- [5] Richards, G., Smith, B. L., and Neal, D. R., 2018. “Finding the uncertainty of the mean for correlated data from PIV”. In 19th International Symposium on the Application of Laser and Imaging Techniques to Fluid Mechanics, Lisbon, Portugal.
- [6] Barber, D., Cemgil, A. T., and Chiappa, S., 2011. *Bayesian time series models*. Cambridge University Press.
- [7] Yates, R. D., and Goodman, D. J., 1998. *Probability and Stochastic Processes*. John Wiley & Sons.
- [8] Law, A. M., and Kelton, W. D., 2007. *Simulation modeling and analysis*, Vol. 3. McGraw-Hill New York.
- [9] Politis, D. N., and White, H., 2004. “Automatic block-length selection for the dependent bootstrap”. *Econometric Reviews*, **23**(1), pp. 53–70.
- [10] Percival, D. B., 1993. “Three curious properties of the sample variance and autocovariance for stationary processes with unknown mean”. *The American Statistician*, **47**(4), pp. 274–276.
- [11] Lahiri, S. N., 2013. *Resampling methods for dependent data*. Springer Science & Business Media.
- [12] Wilson, B. M., and Smith, B. L., 2013. “Uncertainty on PIV Mean and Fluctuating Velocity Due to Bias and Random Errors”. *Meas Sci Technol*, **24**(3), p. 035302.
- [13] Neal, D. R., Sciacchitano, A., Smith, B. L., and Scarano, F., 2015. “Collaborative framework for PIV uncertainty quantification: the experimental database”. *Meas Sci Technol*, **26**.
- [14] Smith, B. L., Stepan, J. J., and McEligot, D. M., 2007. “Velocity and pressure measurements along a row of confined cylinders”. *J. Fluids Eng.*, **129**(10), OCT, pp. 1314–1327.
- [15] Wilson, B. M., Smith, B. L., and Spall, R. E., 2012. “Examples of unsteady CFD validation system response quantities in a cylinder array”. *Nuclear Engineering And Design*, **243**, FEB, pp. 153–167.
- [16] O'Neill, P., Nicolaides, D., Honnery, D., and Soria, J., 2004. “Autocorrelation functions and the determination of integral length with reference to experimental and numerical data”. In 15th Australasian Fluid Mechanics Conference The University of Sydney, Sydney, Australia, pp. 13–17.